

1 **A retrospective investigation of the population structure and geospatial distribution of**
2 ***Salmonella Paratyphi A* in Kathmandu, Nepal**

3

4 Elli Mylona ^{1,2}, Pham Thanh Duy ³, Jacqueline Keane ^{1,2}, Sabina Dongol ⁴, Buddha Basnyat ⁴,

5 Christiane Dolecek ⁵, Phat Voong Vinh ³, Nga Tran Vu Thieu ³, To Nguyen Thi Nguyen ⁶,

6 Abhilasha Karkey ⁵ and Stephen Baker ^{1,2,7}

7

8 ¹ Cambridge Institute of Therapeutic Immunology & Infectious Disease (CITIID), University of Cambridge,
9 Cambridge, UK

10 ² Department of Medicine, University of Cambridge School of Clinical Medicine, Cambridge Biomedical
11 Campus, Cambridge, UK

12 ³ The Hospital for Tropical Diseases, Wellcome Trust Major Overseas Program, Oxford University Clinical
13 Research Unit, Ho Chi Minh City, Vietnam

14 ⁴ Oxford University Clinical Research Unit, Patan Academy of Health Sciences, Kathmandu, Nepal

15 ⁵ The Nuffield Department of Medicine, University of Oxford, Oxford, UK

16 ⁶ Department of Microbiology, Monash Biomedicine Discovery Institute, Monash University, Melbourne,
17 Victoria, Australia

18 ⁷ IAVI Human Immunology Laboratory, Chelsea and Westminster Hospital, London, UK

19

20 **Key words**

21 *Salmonella Paratyphi A*, enteric fever, chronic carriers, paratype, spatiotemporal dynamics

22

23 **Abstract**

24 *Salmonella* Paratyphi A, one of the major etiologic agents of enteric fever, has been on the rise over
25 the last decades in certain endemic regions compared to *S. Typhi*, the most prevalent cause of enteric
26 fever. Despite this, data on the prevalence and molecular epidemiology of *S. Paratyphi* A remain
27 scarce. Here, we analysed the whole genome sequences of a total of 216 *S. Paratyphi* A isolates
28 originating in Nepal between 2005 and 2014, of which 200 were from acute patients and 16 from
29 chronic carriers of enteric fever. By using the recently developed genotyping framework for *S.*
30 *Paratyphi* A (paratype), we identified several genotypes circulating in Kathmandu. Importantly, we
31 observed an unusual clonal expansion of genotype 2.4.3 over a four-year period that spread
32 geographically and replaced other genotypes. This rapid genotype replacement is hypothesised to
33 have been driven by both reduced sensitivity to fluoroquinolones and genetic changes to virulence
34 factors, such as functional and structural genes of type 3 secretion systems. Finally, we show that
35 person to person is likely the most common mode of transmission and chronic carriers play a limited
36 role in maintaining disease circulation.

37

38 **Introduction**

39 Enteric fever is a life-threatening systemic infection caused by *Salmonella enterica* serovars Typhi (*S.*
40 Typhi) and Paratyphi A (*S. Paratyphi A*), B and C. Enteric fever poses a major public health burden in
41 countries with poor sanitation, poor food handling and limited access to clean water. There are an
42 estimated 11–27 million cases annually and >120,000 associated deaths globally [1]. Incidence is
43 higher among children, teenagers, and young adults, and transmission occurs via the faecal-oral route
44 with contaminated food or water [1,2]. Symptoms typically include fever, headache, anorexia, and/or
45 constipation or diarrhoea [2]. In addition, a fraction of enteric fever cases (1–5%) in endemic regions
46 can progress into chronic carriage, which is typically characterized by colonisation of the gallbladder,
47 clinical clearance but shedding of bacteria in faeces, and is thought to be an ecological niche
48 promoting bacterial adaptation and disease transmission [3,4].

49

50 While *S. Typhi* is responsible for most enteric fever cases, *S. Paratyphi A* isolation rate has been on
51 the rise in the recent years causing about a third of global cases, and >40% of cases in certain
52 countries, such as India and Nepal [1,5,6]. However, sparse epidemiological data, inappropriate
53 diagnostic tools and poor surveillance in several endemic countries underestimate the actual incidence
54 of *S. Paratyphi A* and there is limited understanding of the population structure, spatiotemporal
55 spread, and transmission dynamics. To this end and aligning with efforts towards recording and
56 standardizing the genotyping of *S. Typhi*, a single nucleotide polymorphism (SNP)-based tool was
57 recently developed to assign genotypes to *S. Paratyphi A* organisms called ‘Paratype’ [7]. While the
58 symptomatic disease caused by *S. Typhi* and *S. Paratyphi A* is clinically indistinguishable [8], the two
59 organisms are thought to have differential epidemiology, with *S. Paratyphi A* infection more likely to
60 occur outside the household as opposed to *S. Typhi* [9]. The development of tools such as paratype
61 [10], promise to improve understanding of *S. Paratyphi A* epidemiology and genomic surveillance.
62 Notably, and unlike *S. Typhi*, there are currently no licensed vaccines targeting *S. Paratyphi A*, but
63 several are in development [11]. The most likely antigen for a vaccine is the *S. Paratyphi A* O-antigen
64 (O2); however, some isolates have been found to harbour SNPs in the O-antigen encoding region,
65 which now also form part of the Paratype output in order to keep track of such changes [10]. It

66 remains unclear whether nonsynonymous SNPs in this region may lead to changes in the O2-antigen
67 structure that could potentially complicate vaccine-induced antibody binding.

68

69 Unlike *S. Typhi*, reports of multi-drug resistance (MDR; resistance to ampicillin, cotrimoxazole, and
70 chloramphenicol) in *S. Paratyphi A* are rare [12–14]. However, the emergence of resistance to key
71 antimicrobials used for treatment is increasingly being reported, and most *S. Paratyphi A* organisms
72 have reduced susceptibility to nalidixic acid and ciprofloxacin, due to mutations in the quinolone
73 resistance-determining region [15–19]. Additionally, sporadic cases of ceftriaxone resistance, as well
74 as azithromycin resistance have also been reported in the Indian subcontinent [20–22]. Antimicrobial
75 resistance (AMR) becomes a larger problem in areas with poor sanitation and food hygiene, namely
76 areas where *S. Paratyphi A* is endemic and, therefore, incidence reporting and AMR detection
77 becomes crucial for managing enteric fever.

78

79 South Asia is a hotspot for *S. Paratyphi A* infections, with Nepal being a highly endemic area of
80 enteric fever [1]. While there has been an overall decline in the incidence of enteric fever in the last
81 decades, the prevalence of *S. Paratyphi A* isolation in Nepal rose from 23 to 34% of *Salmonella* cases
82 between 1993 and 2003 [14], and remained high between 2005 and 2009 at 31.5% of total enteric
83 fever cases [23]. Furthermore, 1.6% of patients undergoing cholecystectomy were found to be
84 carrying *S. Paratyphi A* organisms in their gallbladder [24]. There is a limited understanding of the
85 population structure of *S. Paratyphi A* and the genetic relatedness between isolates from acute cases
86 and carriers. Here, we retrospectively studied 216 *S. Paratyphi A* organisms, both acute and carrier,
87 isolated in an urban hospital in Kathmandu, Nepal, between 2005 and 2014 by whole genome
88 sequencing (WGS). We determined the structure, genetic diversity, and dynamics of this population,
89 both in a local and a global context. We observed a clonal expansion of one genotype, and thus further
90 investigated the spatial spread of the isolates and identified changes in virulence factors that may
91 explain this expansion. Our study offers new insights into the population structure and dynamics of *S.*
92 *Paratyphi A* in this highly endemic region.

93

94 **Materials and methods**

95 *Ethics*

96 Samples were obtained from studies that were approved by the institutional ethical review boards of
97 Patan Hospital, The Nepal Health Research Council, and The Oxford University Tropical Research
98 Ethics Committee. All enrollees were required to provide written informed consent for the collection
99 and storage of all samples and subsequent data analysis. In the case of those under 18 years of age, a
100 parent or guardian was asked to provide written informed consent.

101

102 *Study site*

103 Bacterial isolates were obtained from blood cultures obtained from patients enrolled on clinical trials
104 (see below) conducted at Patan hospital in Lalitpur Sub-Metropolitan City (LSMC), which is one of
105 the three urban districts that constitute the Kathmandu Valley. The Bagmati River separates the
106 LSMC from the Kathmandu Metropolitan City at the North and West sides of the city. According to
107 the 2001 Nepalese census, it has an area of 15.43 km² and is divided into 22 municipal wards with a
108 population of 126,991 people living in 68,922 households with an average household size of 4.66
109 persons. At the time of sampling most people living in the area were poor and used municipal water
110 and traditional stone spouts as the main water sources. Patan Hospital is the only hospital in LSMC
111 with 320 beds and provides treatment for about 320,000 outpatients and 20,000 inpatients every year.

112

113 *Bacterial isolates*

114 A total of 223 *S. Paratyphi A* isolates were collected from a wide range of studies conducted at Patan
115 hospital, Nepal between 2005 and 2014; from these, 206 blood isolates were collected from four
116 randomized control trials: gatifloxacin vs cefixime (2005) [25], gatifloxacin versus chloramphenicol
117 (2006-2008) [26], gatifloxacin versus ofloxacin (2008-2011) [27], gatifloxacin versus ceftriaxone
118 (2011-2014)[28][28] and a matched case-control study (2011) [23]; 17 chronic isolates was recovered
119 from a chronic gallbladder carriage study (2007-2010) [24].

120 All isolates originated from acute enteric fever patients were recovered from blood cultures collected
121 at the time of enrolment. Briefly, blood samples (5–10 mL for adults; 1–3 mL for children) were
122 inoculated into media containing tryptone soya broth and sodium polyanethol sulphonate to a total
123 volume of 50 mL. The collecting vials were incubated at 37 °C and examined daily for bacterial
124 growth for 7 days. Positive samples were sub-cultured on MacConkey agar and suspected *Salmonella*
125 colonies were identified using standard biochemical tests and serotype-specific antisera (Murex
126 Biotech, Dartford, England).

127

128 Chronic isolates were collected from stool and bile extracts from patients undergoing
129 cholecystectomy. Bile samples were inoculated in Selenite broth and Peptone broth overnight before
130 being sub-cultured onto MacConkey and Xylene Lysine Deoxycholate agar. *Salmonella* colonies were
131 identified using API20E (bioMerieux, Inc) and confirmed using the same serotype-specific antisera as
132 above.

133

134 Epidemiological and clinical information was collected from the above studies where available. Major
135 reported variables included age, sex, main water source(s), water treatment method(s), clinical
136 symptom(s) at presentation, antimicrobial resistance information and individual GPS location.

137

138 The minimum inhibitory concentrations (MICs) of nalidixic acid, ofloxacin, ciprofloxacin,
139 gatifloxacin, azithromycin, chloramphenicol, amoxicillin, cotrimoxazole and ceftriaxone were
140 determined by E-test (AB Biodisk, Solna, Sweden) according to the manufacturer's instructions.
141 MICs were interpreted by using the 2020 CLSI guidelines [29].

142

143 *Genome sequencing*

144 Genomic DNA was extracted using the Wizard Genomic DNA Extraction Kit (Promega, Wisconsin,
145 USA) and 2 µg of DNA was subjected to indexed-tagged paired-end sequencing on an Illumina HiSeq
146 2000 platform to generate 100 bp paired-end reads following the manufacturer's recommendations.
147 Seven genomes were excluded from analyses due to low quality scores based on abnormal genome

148 size (<4 or >5 million base pairs), >1000 contigs, positive status for contamination using confindr
149 [30], and/or low score in bactinspector (available at: <https://gitlab.com/antunderwood/bactinspector>),
150 resulting in a final total number of 216 isolates, of which 200 acute isolates and 16 carrier isolates
151 (Supplementary Table S1). Similar quality screening criteria were applied to the global collection (see
152 below).

153

154 *Phylogenetic analysis*

155 Sequencing reads were mapped to *S. Paratyphi* A AKU12601 (accession: FM200053) reference
156 genome using the nf-core bactmap pipeline v1.0.0 (<https://nf-co.re/bactmap/1.0.0>; [31]) to make
157 pseudogenomes based on high quality positions in VCF files. In all analyses, *S. Typhi* CT18
158 (accession: GCF_000195995.1_ASM19599v1) was used as an outgroup; the sequence was shredded
159 using fasta2fastqshredder.py (<https://github.com/sanger-pathogens/bact-gen-scripts>) to mimic the data
160 format of the sequenced isolates. Using the pseudogenome multi-alignment file obtained from
161 bactmap, repetitive regions, prophages, and recombinant regions were removed using remove-blocks
162 (https://github.com/sanger-pathogens/remove_blocks_from_aln) and a pre-defined coordinates list
163 specific for Paratyphi A (PARAREPEAT; <https://github.com/katholt/typhoid>) followed by a screen
164 and removal of recombinant regions with Gubbins v3.2.0 [32]. The multi-sequence alignment file
165 obtained from gubbins was fed into RAxML-ng v1.1.0 [33] to infer a maximum-likelihood
166 phylogenetic tree using a GTRG+gamma model and the autoMRE convergence test that resulted in
167 750 bootstrap pseudo-analyses of the alignment. For the global collection analysis, following file
168 selection based on quality checks as above, 828 publicly available *S. Paratyphi* A sequences were
169 used ([34], Supplementary Table S2) and repetitive regions, prophages and recombinant regions were
170 excluded as above. The filtered Nepali collection and global collection alignment files were combined
171 and used to run Gubbins. The maximum-likelihood phylogeny tree was inferred as above with 200
172 bootstraps. Outgroup rooting was performed in FigTree v1.4.4 (<https://github.com/rambaut/figtree>)
173 and trees were visualized and annotated using iTOL [35]. SNP distance matrices were produced using
174 snp-dists v0.8.2 (<https://github.com/tseemann/snp-dists>) using the Gubbins output files.

175

176 *Genotyping, antimicrobial resistance genes, plasmid sequences, and virulence factors*

177 Using the raw sequencing reads, or assemblies for the global collection, genotypes were assigned
178 using the Paratype genotyping tool [7], which also detects mutations in *gyrA*, *parC*, *acrB*, and O-
179 antigen region genes (*rmlB*, *rmlD*, *rmlC*, *ddhB*, *prt*, *wzx*).

180

181 Using SRST2 v.2.0.2 [36], known AMR genes were detected with CARD [37] and ARGannot [38]
182 databases, and plasmid replicons were detected with the PlasmidFinder [39] database. Different
183 alleles and/or mismatches in virulence genes were screened using SRST2 and the Virulence Factor
184 Database (VFDB) [40]. Genes that had a difference (either allele or mismatch) within more than 20%
185 of organisms in each clade were considered for visualisation.

186

187 *Clustering analysis*

188 Precise GPS and collection date data were used to detect spatiotemporal clusters were using a Space
189 Time Permutation model [41] in SatScan™ v10.1.2 (available at: <https://www.satscan.org>) at the
190 month level with 2005-06-17 as start study date and 2014-12-01 as end study date and using the
191 high/low rate detection and default settings.

192

193 *Data visualization and statistical analysis*

194 When needed, data were formatted using the tidyverse v2.0.0 [42] and dplyr v1.1.2 [43] in R [44].
195 Statistical analysis and graphs were produced in Graphpad Prism v9.5.1 and Rstudio v2023.06.0+421
196 (Posit software, PBC) using ggplot2 v3.4.2 [45] and maps were annotated using leaflet v2.1.2
197 (<https://CRAN.R-project.org/package=leaflet>) and captured using webshot v0.5.4 ([https://CRAN.R-](https://CRAN.R-project.org/package=webshot)
198 [project.org/package=webshot](https://CRAN.R-project.org/package=webshot)).

199

200 **Results**

201 *Epidemiological and clinical observations*

202 The genomes of 216 *S. Paratyphi A* isolates originating from studies conducted in Patan Hospital in
203 Lalitpur within the Kathmandu valley, Nepal, were whole genome sequenced. Organisms included in

204 this analysis were isolated from patients presenting with acute enteric fever (referred to as enteric
205 fever hereafter) (200/216) [23,25–28] or from the bile/gallbladder of patients undergoing
206 cholecystectomy (i.e., chronic carriers and referred to as carriers hereafter; 16/216) [24]. As dictated
207 by the design of the clinical studies contributing these organisms, most patients were resident within a
208 2.5 km² radius from Patan Hospital, while carriers were more dispersed (Figure 1A). Most acute
209 enteric fever patients suffered from headache (92.5%; 185/200), with anorexia being the second most
210 common symptom reported (68.5%; 137/200) (Figure 1B). More than a third of patients reported
211 being nauseous (37%, 75/200) and/or having abdominal pain (32.5%, 65/200); diarrhoea (13.5%,
212 27/200) and constipation (12%, 24/200) were less common (Figure 1B). A fraction of patients
213 reported none of these symptoms (2.5%, 5/200). The median age of enteric fever patients was 20 (IQR
214 13–24), while the median age of carriers was 38 (IQR 33–43.5) (see also Figure S1). Most enteric
215 fever patients were male (72%; 144/200), while less than a third were female (28%; 56/200). In
216 comparison, a quarter of carriers were male (4/16) and the majority were female (68.8%, 11/16); this
217 information was missing for one patient. Collectively, the representation of *S. Paratyphi A* from acute
218 enteric fever patients spanned a nine-year period, from 2005 to 2014. During this time, cases peaked
219 in the summer months (June–August; total of 126/200), while remained relatively low in the winter
220 months (October–February; total of 17/200, Figure 1C), which is comparable to previous observations
221 in this area [46].

222

223 *S. Paratyphi A* population structure in Nepal

224 To explore the structure of this *S. Paratyphi A* population, we inferred a whole genome phylogeny
225 based on 1,082 SNPs and assigned genotypes based on the recently developed paratype scheme [7]
226 (Figure 2 and Supplementary Table S1). As a reference, the lineages of the old genotyping scheme [5]
227 are also shown in Supplementary Table S1. Overall, *S. Paratyphi A* isolates were highly related,
228 having an average pairwise difference of only 10 SNPs (IQR 8–90).

229

230 The most common genotype in the Nepali *S. Paratyphi A* population was 2.4.1 (61.6%; 133/216),
231 while the second most common was 2.4.3 (29.2%; 63/216). Genotypes 2.3.2 (2.8%; 6/216), 2.3.3

232 (1.9%; 4/216), 2.4 (1.4%; 3/216), and 2.4.5 (0.9%; 2/216) were present in small numbers, while 2.3,
233 2.3.4, the newly defined 2.4.6 and 2.4.8, as well as 1.1 were each recorded once (0.5%; 1/216 each).

234

235 The *S. Paratyphi* A population in Nepal split largely into two major subclades, comprised mainly of
236 isolates belonging to genotype 2.4.3, that clustered closely to a small group of 2.4 organisms, and
237 genotype 2.4.1 (Figure 2). In line with the genotyping scheme, genotypes belonging to the 2.3
238 secondary clade of Paratype (2.3, 2.3.2, 2.3.3, 2.3.4) clustered mainly in a separate, smaller clade.

239 Although the number of carrier *S. Paratyphi* A isolates in this collection was small, these belonged
240 mostly to 2.4.1 (9/16), as well as 2.4 (1/16), 2.4.5 (1/16), 2.4.8 (1/16), 2.3 (1/16), 2.3.2 (1/16), 2.3.3

241 (1/16), and 1.1 (1/16), the latter being the most distantly related to the rest. The distribution of

242 genotypes over age in enteric fever patients is shown in Figure S1; we did not observe any correlation

243 between the age brackets and genotype (Pearson's chi squared test $p>0.05$). Furthermore, we

244 visualised the isolates by genotype on a geographical map of the area (Figure S2A-C) and used

245 precise GPS and collection date information to perform spatiotemporal analysis. A space-time

246 permutation model predicted two statistically significant clusters; one high-rate cluster in a 1.66 km

247 radius containing 20 isolates ($p=0.00024$) and one low-rate cluster in a 0.49 km radius containing 33

248 isolates ($p=0.047$) (Figure S2D). Other statistically not significant clusters ($p>0.05$) were also

249 observed (Figure S2D). Notably, we observed a variety of genotypes within these clusters.

250

251 The two major clades observed, that of genotype 2.4.1 and that of genotype 2.4.3, were each highly

252 clonal with isolates of genotype 2.4.3 differing by 2 SNPs on average (IQR 1–3) upon pairwise

253 comparison; isolates of genotype 2.4.1 differed by a median of 7 SNPs (IQR 6–8). Overall, the carrier

254 isolates carried on average 16 SNPs (IQR 14–92) when compared to acute isolates, but, as suggested

255 by the assigned genotypes, carrier isolates were closely related to acute cases that shared the same

256 genotype, such as that genotype 2.4.1 carrier isolates differed by 6 SNPs (IQR 5–8). The distantly

257 related isolate of genotype 1.1 showed a pairwise difference of 438 SNPs with all other isolates (IQR

258 436–438) irrespective of whether they were from acute or carrier cases.

259

260 Most organisms isolated in the first half of the study period (2005–2009) belonged to genotype 2.4.1
261 with a few cases caused by *S. Paratyphi A* belonging to 2.3.2, 2.3.3, 2.4 or 2.4.5 genotypes, while
262 genotype 2.3 was sampled only in 2008 (Figure 2 and 3). In the second half of the study period
263 (2011–2014), we observed that genotype 2.3 and 2.3.2 were not represented. Additionally, genotype
264 2.3.3 organisms were diminished, and there was an extensive decrease in isolation of genotype 2.4.1
265 organisms. Conversely, there was a substantial increase in the isolation of organisms belonging to
266 genotype 2.4.3, which accounted for most organisms isolated during this time period (Figure 3). A
267 limited number of 2.3.4 organisms were identified in 2013 only. The temporal distribution was also
268 supported by the pairwise SNP difference within the clades, such that 2.4.1 had accumulated more
269 SNPs.

270

271 *Water source and water treatment methods in acute cases*

272 Water source has been previously linked to enteric fever outbreaks and transmission, with the
273 contamination of water supplies with sewage waste and general sanitation issues likely responsible for
274 enteric fever in areas like Kathmandu [47,48]. We had information regarding the main water source
275 and water treatment method for 156 acute cases. From those, more than half (82/156; 52.6%) used the
276 municipal water supply, while smaller numbers used bottled water, water from wells, stone spout
277 water, tanker water or a mix of these (Table 1). A third of patients did not treat their water in any way
278 (55/156; 35.3%), another third filtered their water prior to consumption (54/156; 34.6%), while less
279 than a fifth boiled their water (23/156; 14.7%) and very few cases chlorinated, used another way of
280 treatment, or did not treat water at all. For each water source, filtering and no treatment were the more
281 common treatment methods, while other methods (chlorine, boiling, other) were associated with fewer
282 cases (Table 1).

283

284

285

286

287

288 Table 1. Main water source and treatment method of 156 Nepali *S. Paratyphi A* patients.

Main water source	Count (%)	Water treatment method	Count (%)
Municipal	82/156 (52.6)	Boiling	12/82 (14.6)
		Chlorine	3/82 (3.7)
		Filter	38/82 (46.3)
		Mix	8/82 (7.8)
		Other	1/82 (1.2)
		Untreated	19/82 (23.2)
		Unknown	1/82 (1.2)
Stone spout	23/156 (14.7)	Boiling	4/23 (17.4)
		Chlorine	1/23 (4.4)
		Filter	3/23 (13.0)
		Mix	3/23 (13.0)
		Untreated	12/23 (52.2)
Bottled	17/156 (10.9)	Boiling	1/17 (5.9)
		Filter	5/17 (29.4)
		Mix	1/17 (5.9)
		Other	1/17 (5.9)
		Untreated	9/17 (53.0)
Well	17/156 (10.9)	Boiling	4/17 (23.5)
		Chlorine	1/17 (5.9)
		Filter	5/17 (29.4)
		Mix	1/17 (5.9)
		Other	1/17 (5.9)
		Untreated	5/17 (29.4)
Tanker	2/156 (1.3)	Untreated	2/2 (100)
Mix	15/156 (9.6)	Boiling	2/15 (13.3)
		Chlorine	1/15 (6.7)
		Filter	3/15 (20)
		Other	1/15 (6.7)
		Untreated	8/15 (53.3)

289

290 Although the group of the 156 *S. Paratyphi A* cases for which we had water source and treatment
 291 information did not contain comparable numbers of genotypes, there was no correlation between
 292 water source and genotype (genotypes 2.3.2, 2.3.3, 2.3.4, 2.4, 2.4.1, 2.4.3, 2.4.5, 2.4.6; Chi squared
 293 $p=0.248$) (Figure 4). There were a comparable number of *S. Paratyphi A* of genotype 2.4.3 and
 294 genotype 2.4.1 associated with people reporting access to municipal water, stone spout and tanker
 295 water, while 2.4.3 was more common in those reporting using bottled water and 2.4.1 in those
 296 reporting using well water (Figure S3A,C). The majority of cases were associated with municipal
 297 water and hence the greater variety of genotypes identified (2.3.3, 2.4, 2.4.1, 2.4.3, and 2.4.6) (Figure
 298 S3A,C). The low numbers of *S. Paratyphi A* organisms detected following reporting chlorinated,

299 unknown or ‘other’ treatment were all genotype 2.4.1 (Figure S3B, D), while multiple genotypes were
300 detected even after reporting of boiling, filtering or not treating water (Figure 4, S3B,D).

301

302 *The Nepali S. Paratyphi A population in a global context*

303 To investigate the *S. Paratyphi A* population dynamics in a global context, we constructed a global
304 phylogeny based on 6,811 SNPs by including an additional 828 publicly available *S. Paratyphi A*
305 genomes from a wide range of geographical regions and chronological periods (Figure 5). Genomes
306 from our collection clustered not only with those of the same genotype in the global phylogenetic tree,
307 but also with those originating from the same period (2005–2014) and more recently (2015–2019). In
308 the context of this study, the more recent (2011–2014) 2.4.3 genotype clade clustered with other
309 organisms of the same genotype and was closely related to a clade containing 2.4.2 and 2.4.4
310 organisms from more recent years; all these genomes originated from countries bordering Nepal;
311 India, Pakistan, and Bangladesh. The majority of organisms from our collection clustered closely to
312 organisms originating from countries within South and Southeast Asia.

313

314 *Antimicrobial susceptibility*

315 The isolates were phenotypically tested for their antimicrobial susceptibility. All acute isolates were
316 susceptible to amoxicillin, cotrimoxazole, ceftriaxone, and gatifloxacin. The majority of acute isolates
317 were susceptible (MIC \leq 8 μ g/ml; 80.6%; 87/108) or exhibited intermediate resistance (MIC 8–32
318 μ g/ml; 18.5%; 20/108) against chloramphenicol and azithromycin (susceptible MIC \leq 16 μ g/ml
319 90.1%; 136/150 and intermediate resistance MIC 16–32 μ g/ml 6.7%; 10/150, respectively). It is worth
320 noting that the small number of acute isolates displaying resistance to azithromycin had MICs of 32–
321 48 μ g/ml. Similarly, the majority of carrier isolates (15/16) were susceptible to chloramphenicol
322 and/or azithromycin. Conversely, 94% of acute isolates tested (141/150) were resistant to nalidixic
323 acid (MIC $>$ 256 μ g/ml), while we recorded 14.7% (22/150) resistant (MIC $>$ 1 μ g/ml) and 79.3%
324 (119/150) intermediate resistance (MIC 0.5–0.12 μ g/ml) to ciprofloxacin, as well as 48% (72/150)
325 resistant (MIC $>$ 2 μ g/ml) and 46.7% (70/150) intermediate resistance (MIC 0.25–1 μ g/ml) to

326 ofloxacin. Overall, 62.5% (10/16) of carrier isolates were resistant to nalidixic acid, which was
327 significantly lower than acute isolates (Fisher's exact test $p=0.0009$). Significantly fewer (Fisher's
328 exact test $p<0.05$) carrier isolates were ciprofloxacin-resistant (12.5%; 2/16) or intermediate (50%;
329 8/16) compared to acute isolates.

330

331 We also evaluated the presence of genetic markers associated with AMR in the Nepali collection of *S.*
332 *Paratyphi A*. Mutations in *gyrA* are associated with reduced susceptibility to fluoroquinolones in *S.*
333 *Paratyphi A* [18]. The majority of *S. Paratyphi A* isolates carried a mutation in *gyrA*; the most
334 common mutation was a substitution of serine (S) to phenylalanine (F) at position 83 (S83F; 199/216;
335 92.1%), while only two isolates carried a serine to tyrosine (Y) substitution at the same position
336 (S83Y; 0.9%) and two carried an aspartic acid (D) to tyrosine substitution at position 87 (D87Y;
337 0.9%) in *gyrA* (Figure 2). Notably, none of the isolates carried more than one mutation in *gyrA*, unlike
338 more distantly related global isolates that were studied here (Figure 5). We did not find any mutations
339 in *parC* that has previously been linked to fluoroquinolone resistance [18]. Resistance to azithromycin
340 is an increasing problem in typhoid treatment, with a few azithromycin resistant cases reported in *S.*
341 *Paratyphi A* [20,21,49]. However, we did not detect any mutations in *acrB* that are commonly
342 associated with this phenotype [21]. All samples also carried the cryptic *aac6'-Iy* gene, which,
343 however, is generally silent and is only expressed after rare transcriptional modifications [50].

344

345 *Genetic differences in virulence determinants between clades*

346 To identify genetic markers associated with the expansion of either 2.4.1 or 2.4.3 clade, we next
347 investigated the presence of plasmids, differences in gene content, and/or differences in virulence
348 factors. By screening the collection against the PlasmidFinder database (see Methods), we found only
349 three isolates carrying a plasmid, two genotype 2.4.3 organisms (alleles ColBS512_1_NC_010656,
350 and Col156_1_NC_009781) and one genotype 2.3.4 organism (allele CP036184_1_p5025970) that
351 we have previously found circulating in India [34]. We did not identify any other known plasmid
352 replicons in this collection.

353

354 An important virulence factor for *S. Paratyphi A* is the O-antigen of the lipopolysaccharide (LPS).
355 While O-antigen has important roles in bacterial physiology and virulence, there is added interest due
356 to its use as a target antigen in several vaccines in development against *S. Paratyphi A* [11]. For this
357 reason, the paratype genotyping tool also records SNP mutations for genes in the O-antigen encoding
358 region and three of such mutations were detected in our collection (*wzx*-G329D, *rmlB*-H176Y, *prt*-
359 C249C) (Figure 2). Interestingly, genotype 2.4.1 isolates commonly had a *rmlB*-H176Y mutation,
360 while the *wzx*-G329D substitution was only found in 2.4.3 isolates; the frequency of these mutations
361 did not change with time. It is worth noting that several different O-antigen gene mutation profiles
362 were observed in a global context as well (Figure 5).

363

364 We lastly inspected the collection for differences in other virulence factors, as defined by the
365 Virulence Factor database [40]. We identified several differences in genes associated with the
366 *Salmonella* pathogenicity island (SPI) 1 and 2 to be associated with the 2.4.3 clade (Figure 6). More
367 specifically, changes in SPI-2 genes included a SNP in the SPI-2 effector protein *sifA* that was
368 associated with 2.4.3 clade and its closely related isolates (2.4, 2.4.5, 2.4.6, 2.4.8) and was not found
369 in the rest of the collection. Similarly, clade 2.4.3 exhibited a higher rate of carrying distinct alleles
370 for *iacP* and the SPI-1 effector protein genes *sopA* and *sopD2*, while two different *sopE* alleles
371 clustered with each of the 2.4.3 and 2.4.1 clade. A similar pattern was observed for several SPI-2
372 structural/effector protein genes (*ssaJ*, *ssaN*, *ssaP*, and *steA*), as well as *sinH* that is involved in
373 bacterial adhesion. Finally, we observed higher chance of differences in the 2.4.3 clade for a small
374 number of fimbria-associated genes (*bcfC*, *csfA*), while this clade was more likely to lack a Type 6
375 secretion system (T6SS)-associated gene (STM0283). The genetic diversity identified in these
376 virulence factors included either small mismatches (e.g. *sifA*) or larger scale changes (where different
377 alleles carry multiple mismatches/indels, like in the case of *sopE*).

378

379 **Discussion**

380 The control of enteric fever caused by *S. Paratyphi A* is impacted by AMR, poor hygiene, and
381 sanitation measures, as well as the lack of a prophylactic vaccine and a robust framework for

382 sustained surveillance that would allow investigation of population dynamics. Here, we assessed
383 disease dynamics and defined the *S. Paratyphi A* population structure at the genomic level in the
384 highly endemic Kathmandu valley, Nepal, by performing WGS, utilising a recently developed
385 genotyping scheme and associated metadata.

386

387 The population of *S. Paratyphi A* studied here comprised of isolates spanning a 9-year period (2005–
388 2014). The median age of *S. Paratyphi A* patients was 20 years, in line with identified at-risk
389 populations [51]. Carriers were older in age and mostly females, in accordance with associated risk
390 factors for gallbladder removal [3]. Overall, the rate of *S. Paratyphi A* isolation peaked in the summer
391 months, consistent with previous observations regarding the seasonality of enteric fever, and may be
392 associated with increased rainfall and flooding, and/or warmer temperatures, which may promote
393 water contamination [23,52]. Contamination of water sources with coliforms of typhoidal *Salmonella*
394 has been reported in Nepal [47], which may explain the higher counts of *S. Paratyphi A*-infected
395 patients using municipal water irrespective of the treatment method of choice. In addition, the
396 majority of *S. Paratyphi A* in Kathmandu was punctuated by mutations in *gyrA* associated with
397 reduced susceptibility to fluoroquinolones, which was supported by the MIC values of ciprofloxacin.
398 These data further highlight the need for improved access to clean water in endemic areas.

399

400 We have very limited understanding of the role of carriers in shaping the population of *S. Paratyphi A*.
401 Although the number of isolated organisms was small, we found a considerable genetic diversity
402 within the carrier isolates, as suggested by the number of genotypes assigned to this group of
403 organisms. The genetic variation observed between *S. Paratyphi A* from enteric fever and those from
404 carriers was associated with the assigned genotype. We found one carrier isolate belonging to
405 genotype 1.1 which was very distant from the rest of the population, differing by a markedly high
406 number of SNPs from all other organisms in this study. This genetic distance from other circulating
407 genotypes serves also as an indication of the length of carriage time in this patient. Importantly, we
408 found carrier isolates to not be restricted to a single genotype and several carrier isolates belonged to
409 genotypes that included several acute isolates circulating in the area. Specifically, carriers of genotype

410 2.4.1 had many common characteristics with acute isolates, such as antimicrobial susceptibility
411 profiles and mutations in *gyrA* and *rmlB*, suggesting not only that 2.4.1 has been circulating within the
412 Nepali population for some time, but also that carriers reflect the general circulating population.
413 Gallbladder carriage in *S. Typhi* has been suggested to be responsible for enriching the genetic
414 diversity in the circulating population in such a way that carrier isolates show increased numbers of
415 SNPs compared to acute even within the same genotype [4]. In contrast, we did not make such an
416 observation here, suggesting that either these carriers have been infected comparatively recently and
417 the bacterial organisms have not undergone adaptive mutations, or that sustained person-to-person
418 transmission may be taking place. Importantly, chronic carriers have limited contribution to the
419 genetic diversity in the population and genotypes that were identified only in the carrier isolates in
420 this collection did not reappear.

421

422 Further exploring the transmission dynamics of the *S. Paratyphi A* population, we observed
423 spatiotemporal clusters incorporating multiple genotypes, indicative of acute transmission of *S.*
424 *Paratyphi A* in Kathmandu. Previous case-control studies in Indonesia and Nepal also found that acute
425 transmission plays an important role in *S. Paratyphi A* infection with flooding and contaminated street
426 food being important risk factors, whereas contaminated drinking water was more associated with *S.*
427 *Typhi* infections [9,53]. Noticeably, all these localized *S. Paratyphi A* outbreaks occurred during
428 monsoon months between May and September corresponding to the seasonal distribution of enteric
429 fever in Asia. As mentioned earlier, there is a high degree of faecal contamination in the water supply
430 in correlation with increased rainfall [47]. Only half of the patients in this study had access to
431 municipal water, further supporting that *S. Paratyphi A* infection spreads from a focal point by direct
432 transmission already suggested in our early studies [46].

433

434 Outbreaks of enteric fever caused by *S. Paratyphi A* are not uncommon. We recently reported an
435 outbreak of *S. Paratyphi A* in Vadodara, India, which was likely associated with environmental
436 contamination and a reduction in hygiene standards [34]. Additionally, an outbreak of *S. Paratyphi A*
437 infection caused by isolates belonging to the same subclade originating from Cambodia spread to

438 several provinces and sickened many returning travellers [15]. Although the source remained
439 undetermined, contaminated water or food is the most reasonable cause that was then facilitated by
440 acute person to person transmission. It is, therefore, essential that preventative measures including
441 improvement of food safety, water quality and hygiene are implemented to control *S. Paratyphi A*
442 dissemination. Considering, however, that these measures are unlikely to serve as an imminent
443 solution in Nepal, the development of vaccines against *S. Paratyphi A* or a bivalent vaccine protecting
444 against both serovars Typhi and Paratyphi A is of high priority.

445

446 Our global *S. Paratyphi A* population analysis suggested that the genotypes observed in Kathmandu in
447 2005–2014 were circulating in neighbouring countries in the Indian subcontinent and were possibly
448 introduced to Nepal. Interestingly, we observed a clonal expansion of genotype 2.4.3 that spread
449 across Kathmandu and quickly overtook genotype 2.4.1 during the time frame studied here. Genotype
450 2.4.3 was closely related to genotypes 2.4.2 and 2.4.4 in the neighbouring countries, which appeared
451 in later years, suggesting an exchange of genotypes that circulate in the subcontinent. While we do not
452 currently have data on the prevalent genotype(s) post 2014, it is likely that there are cycles of clonal
453 expansion that are driven by environmental pressures. It remains unclear whether the genotype
454 expansion observed in Nepal was due to successful local clonal expansion or a result of strain
455 introduction from neighbouring countries. Whether the local phylogenetic dynamics of *S. Paratyphi A*
456 in Nepal, namely dominance of a single genotype over time, has also occurred elsewhere in South
457 Asia warrants further investigation.

458

459 The lower average SNP difference within genotype 2.4.3 compared to 2.4.1 organisms suggests recent
460 transmission and rapid clonal expansion, that led to the replacement of other genotypes within a four-
461 year period. Our data indicate that, while reduced susceptibility to fluoroquinolones is important for
462 the persistence of successful genotypes of *S. Paratyphi A*, DNA gyrase mutations are present in most
463 genotypes, both Nepali and global, and occurred independently due to increased use of
464 fluoroquinolones, and therefore the clonal expansion of 2.4.3 in Nepal was not solely due to AMR. *S.*
465 *Paratyphi A* is highly endemic in Nepal, suggesting that entering an immunologically naïve

466 population is possibly not the main driver for the 2.4.3 clade expansion, unless this replacement was
467 associated with subtle changes in antigenicity in these organisms. For example, mutations in the O-
468 antigen *rmlB* and *wzx* genes were associated with specific genotypes (2.4.1 and 2.4.3, respectively).
469 RmlB is essential for rhamnose biosynthesis, which is one of the O-antigen sugars, and Wzx
470 translocates the synthesised O-units across the inner membrane [54]. It is possible that subtle changes
471 on the O-antigen may interfere with antibody recognition and thus immune protection. Given that
472 aspects of O-antigen structure have been shown to directly affect interactions with factors of the
473 immune system [55,56], it is imperative to investigate whether these genetic changes can have a
474 phenotypic impact that may affect interactions with the host. We additionally detected increased
475 genetic variation between 2.4.3 and 2.4.1 *S. Paratyphi A* organisms in several SPI-1 and SPI-2 genes.
476 These pathogenicity islands encode for two type 3 secretion systems (T3SS), which are essential for
477 virulence by mediating entry (SPI-1) and survival (SPI-2) within host cells [57]. Genetic differences
478 in the genes encoding several SPI-1 effector proteins (*SifA*, *SopA*, *SopD2*, *SopE*) that dictate
479 interactions with host cells, such as cytoskeletal changes and immune signalling [57], in combination
480 with changes in SPI-2 structural components (*SsaJ/N/P*) that may affect SPI-2 T3SS stability and thus
481 intracellular survival, may have been central to the success of genotype 2.4.3 and warrant further
482 investigation. It is likely that the combined genetic diversity in virulence factors has contributed to the
483 expansion of genotype 2.4.3.

484

485 In summary, the *S. Paratyphi A* population in Nepal is largely clonal marked by mutations in DNA
486 gyrase and increased genetic variation in other virulence-associated genetic loci. We observed an
487 unusual clonal expansion and replacement by a single genotype, which may be attributed to both
488 reduced susceptibility to fluoroquinolones and to the genetic variation in virulence genes. Rapid
489 dissemination within the local population and diversification of the dominant clones have resulted in
490 short-term spatiotemporal clusters of various genotypes indicating the importance of direct
491 transmission of *S. Paratyphi A* in this setting. We further suggest that in a highly endemic setting,
492 chronic carriage has a limited contribution to disease transmission and maintenance of the genetic
493 diversity.

494 **Acknowledgements**

495 We wish to thank all patients participating in the contributing clinical studies and the diagnostics
496 services helping collecting samples.

497

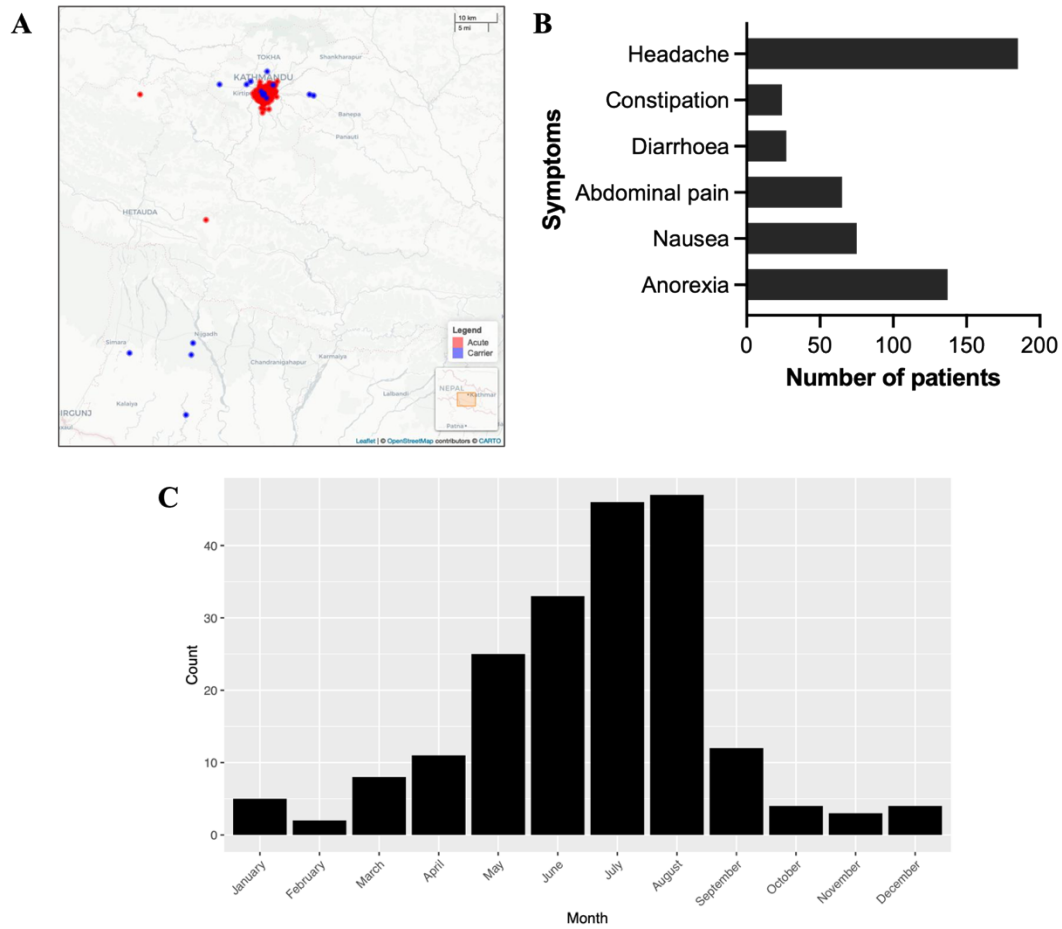
498 **References**

- 499 1. Stanaway JD, Reiner RC, Blacker BF, et al. The global burden of typhoid and paratyphoid
500 fevers: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Infect Dis*
501 **2019**; 19:369–381.
- 502 2. Crump JA, Sjölund-Karlsson M, Gordon MA, Parry CM. Epidemiology, clinical presentation,
503 laboratory diagnosis, antimicrobial resistance, and antimicrobial management of invasive
504 *Salmonella* infections. *Clin Microbiol Rev.* 2015; 28:901–937.
- 505 3. Gunn JS, Marshall JM, Baker S, Dongol S, Charles RC, Ryan ET. *Salmonella* chronic
506 carriage: Epidemiology, diagnosis, and gallbladder persistence. *Trends Microbiol.* 2014;
507 22:648–655.
- 508 4. Duy PT, Thieu NTV, Nguyen TNT, et al. Gallbladder carriage generates genetic variation and
509 genome degradation in *Salmonella* Typhi. *PLoS Pathog* **2020**; 16:e1008998.
- 510 5. Zhou Z, McCann A, Weill F-X, et al. Transient Darwinian selection in *Salmonella enterica*
511 serovar Paratyphi A during 450 years of global spread of enteric fever. *Proceedings of the*
512 *National Academy of Sciences* **2014**; 111:12199–12204.
- 513 6. Zellweger RM, Basnyat B, Shrestha P, et al. A 23-year retrospective investigation of
514 *Salmonella* Typhi and *Salmonella* Paratyphi isolated in a tertiary Kathmandu hospital. *PLoS*
515 *Negl Trop Dis* **2017**; 11:e0006051.
- 516 7. Tanmoy AM, Hooda Y, Sajib MSI, et al. Paratype: a genotyping tool for *Salmonella* Paratyphi
517 A reveals its global genomic diversity. *Nature Communications* 2022 13:1 **2022**; 13:1–10.
- 518 8. Maskey AP, Day JN, Tuan PQ, et al. *Salmonella enterica* Serovar Paratyphi A and *S. enterica*
519 Serovar Typhi Cause Indistinguishable Clinical Syndromes in Kathmandu, Nepal. *Clinical*
520 *Infectious Diseases* **2006**; 42:1247–1253.
- 521 9. Vollaard AM, Ali S, Van Asten HAGH, et al. Risk factors for typhoid and paratyphoid fever in
522 Jakarta, Indonesia. *J Am Med Assoc* **2004**; 291:2607–2615.
- 523 10. Hooda Y, Tanmoy AM, Saha SK, Saha S. Genomic Surveillance of *Salmonella* Paratyphi A:
524 Neglected No More? *Open Forum Infect Dis* **2023**; 10:S53–S57.
- 525 11. Shakya M, Neuzil KM, Pollard AJ. Prospects of Future Typhoid and Paratyphoid Vaccines in
526 Endemic Countries. *J Infect Dis* **2021**; 224:S770.
- 527 12. Rahman SIA, Nguyen TNT, Khanam F, et al. Genetic diversity of *Salmonella* Paratyphi A
528 isolated from enteric fever patients in Bangladesh from 2008 to 2018. *PLoS Negl Trop Dis*
529 **2021**; 15.

- 530 13. Holt KE, Thomson NR, Wain J, et al. Multidrug-Resistant *Salmonella enterica* Serovar
531 Paratyphi A Harbors IncHI1 Plasmids Similar to Those Found in Serovar Typhi. *J Bacteriol*
532 **2007**; 189:4257–4264.
- 533 14. Maskey AP, Basnyat B, Thwaites GE, Campbell JI, Farrar JJ, Zimmerman MD. Emerging
534 trends in enteric fever in Nepal: 9124 cases confirmed by blood culture 1993-2003. *Trans R*
535 *Soc Trop Med Hyg* **2008**; 102:91–95.
- 536 15. Kuijpers LMF, Le Hello S, Fawal N, et al. Genomic analysis of *Salmonella enterica* serotype
537 Paratyphi A during an outbreak in Cambodia, 2013-2015. *Microb Genom* **2016**; 2:e000092.
- 538 16. Britto CD, Dyson ZA, Duchene S, et al. Laboratory and molecular surveillance of paediatric
539 typhoidal *Salmonella* in Nepal: Antimicrobial resistance and implications for vaccine policy.
540 *PLoS Negl Trop Dis* **2018**; 12:e0006408.
- 541 17. Woods CW, Murdoch DR, Zimmerman MD, et al. Emergence of *Salmonella enterica* serotype
542 Paratyphi A as a major cause of enteric fever in Kathmandu, Nepal. *Trans R Soc Trop Med*
543 *Hyg* **2006**; 100:1063–1067.
- 544 18. Nair S, Unnikrishnan M, Turner K, et al. Molecular Analysis of Fluoroquinolone-resistant
545 *Salmonella Paratyphi A* Isolate, India. *Emerg Infect Dis* **2006**; 12:489.
- 546 19. Mawatari M, Kato Y, Hayakawa K, et al. *Salmonella enterica* serotype paratyphi a carrying
547 CTX-M-15 type extended-spectrum beta-lactamase isolated from a Japanese traveller
548 returning from India, Japan, July 2013. *Eurosurveillance* **2013**; 18:20632.
- 549 20. Molloy A, Nair S, Cooke FJ, et al. First report of *Salmonella enterica* serotype paratyphi A
550 azithromycin resistance leading to treatment failure. *J Clin Microbiol* **2010**; 48:4655–4657.
- 551 21. Hooda Y, Sajib MSI, Rahman H, et al. Molecular mechanism of azithromycin resistance
552 among typhoidal *Salmonella* strains in Bangladesh identified through passive pediatric
553 surveillance. *PLoS Negl Trop Dis* **2019**; 13:e0007868.
- 554 22. Nair S, Day M, Godbole G, et al. Genomic surveillance detects *Salmonella enterica* serovar
555 Paratyphi A harbouring blaCTX-M-15 from a traveller returning from Bangladesh. *PLoS One*
556 **2020**; 15:e0228250.
- 557 23. Karkey A, Arjyal A, Anders KL, et al. The Burden and Characteristics of Enteric Fever at a
558 Healthcare Facility in a Densely Populated Area of Kathmandu. *PLoS One* **2010**; 5:e13988.
- 559 24. Dongol S, Thompson CN, Clare S, et al. The Microbiological and Clinical Characteristics of
560 Invasive *Salmonella* in Gallbladders from Cholecystectomy Patients in Kathmandu, Nepal.
561 *PLoS One* **2012**; 7.
- 562 25. Pandit A, Arjyal A, Day JN, et al. An Open Randomized Comparison of Gatifloxacin versus
563 Cefixime for the Treatment of Uncomplicated Enteric Fever. *PLoS One* **2007**; 2:e542.
- 564 26. Arjyal A, Basnyat B, Koirala S, et al. Gatifloxacin versus chloramphenicol for uncomplicated
565 enteric fever: an open-label, randomised, controlled trial. *Lancet Infect Dis* **2011**; 11:445–454.

- 566 27. Koirala S, Basnyat B, Arjyal A, et al. Gatifloxacin Versus Ofloxacin for the Treatment of
567 Uncomplicated Enteric Fever in Nepal: An Open-Label, Randomized, Controlled Trial. *PLoS*
568 *Negl Trop Dis* **2013**; 7.
- 569 28. Arjyal A, Basnyat B, Nhan HT, et al. Gatifloxacin versus ceftriaxone for uncomplicated
570 enteric fever in Nepal: An open-label, two-centre, randomised controlled trial. *Lancet Infect*
571 *Dis* **2016**; 16:535–545.
- 572 29. CLSI. M100 Performance Standards for Antimicrobial Susceptibility Testing. 2020; 30th
573 edition.
- 574 30. Low AJ, Koziol AG, Manninger PA, Blais B, Carrillo CD. ConFindr: Rapid detection of
575 intraspecies and cross-species contamination in bacterial whole-genome sequence data. *PeerJ*
576 **2019**; 2019:e6995.
- 577 31. Ewels PA, Peltzer A, Fillinger S, et al. The nf-core framework for community-curated
578 bioinformatics pipelines. *Nature Biotechnology* 2020 38:3 **2020**; 38:276–278.
- 579 32. Croucher NJ, Page AJ, Connor TR, et al. Rapid phylogenetic analysis of large samples of
580 recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* **2015**;
581 43:e15.
- 582 33. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and
583 user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **2019**;
584 35:4453–4455.
- 585 34. Pereira-Dias J, Taneja N, Mahindroo J, et al. The genomic characterization of *Salmonella*
586 *Paratyphi A* from an outbreak of enteric fever in Vadodara, India. *Microb Genom* **2023**;
587 9:000914.
- 588 35. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree
589 display and annotation. *Nucleic Acids Res* **2021**; 49:W293–W296.
- 590 36. Inouye M, Dashnow H, Raven LA, et al. SRST2: Rapid genomic surveillance for public health
591 and hospital microbiology labs. *Genome Med* **2014**; 6:1–16.
- 592 37. Alcock BP, Raphenya AR, Lau TTY, et al. CARD 2020: antibiotic resistance surveillance with
593 the comprehensive antibiotic resistance database. *Nucleic Acids Res* **2020**; 48:D517.
- 594 38. Gupta SK, Padmanabhan BR, Diene SM, et al. ARG-ANNOT, a New Bioinformatic Tool To
595 Discover Antibiotic Resistance Genes in Bacterial Genomes. *Antimicrob Agents Chemother*
596 **2014**; 58:212.
- 597 39. Carattoli A, Zankari E, Garcíá-Fernández A, et al. In silico detection and typing of plasmids
598 using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother*
599 **2014**; 58:3895–3903.
- 600 40. Chen L, Yang J, Yu J, et al. VFDB: a reference database for bacterial virulence factors.
601 *Nucleic Acids Res* **2005**; 33:D325–D328.
- 602 41. Kuldorff M, Heffernan R, Hartman J, Assunção R, Mostashari F. A Space–Time Permutation
603 Scan Statistic for Disease Outbreak Detection. *PLoS Med* **2005**; 2:e59.

- 604 42. Wickham H. tidyverse: easily install and load the ‘Tidyverse’. 2017;
- 605 43. Wickham H, Francois R, Henry L, Müller K, Vaughan D. dplyr: A Grammar of Data
606 Manipulation. 2023.
- 607 44. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for
608 Statistical Computing Vienna Austria. 2014; 0: {ISBN} 3-900051-07-0.
- 609 45. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
- 610 46. Baker S, Holt KE, Clements ACA, et al. Combined high-resolution genotyping and geospatial
611 analysis reveals modes of endemic urban typhoid fever transmission. *Open Biol* **2011**; 1.
- 612 47. Karkey A, Jombart T, Walker AW, et al. The Ecological Dynamics of Fecal Contamination
613 and Salmonella Typhi and Salmonella Paratyphi A in Municipal Kathmandu Drinking Water.
614 *PLoS Negl Trop Dis* **2016**; 10:4346.
- 615 48. Mogasale V V., Ramani E, Mogasale V, Park JY, Wierzba TF. Estimating Typhoid Fever Risk
616 Associated with Lack of Access to Safe Water: A Systematic Literature Review. *J Environ*
617 *Public Health* **2018**; 2018.
- 618 49. Sajib MSI, Tanmoy AM, Hooda Y, et al. Tracking the Emergence of Azithromycin Resistance
619 in Multiple Genotypes of Typhoidal Salmonella. *mBio* **2021**; 12.
- 620 50. Magnet S, Courvalin P, Lambert T. Activation of the Cryptic *aac(6′)-Iy* Aminoglycoside
621 Resistance Gene of Salmonella by a Chromosomal Deletion Generating a Transcriptional
622 Fusion. *J Bacteriol* **1999**; 181:6650.
- 623 51. Martin LB, Khanam F, Qadri F, Khalil I, Sikorski MJ, Baker S. Vaccine value profile for
624 Salmonella enterica serovar Paratyphi A. *Vaccine* **2023**; 41:S114–S133.
- 625 52. Saad NJ, Lynch VD, Antillón M, Yang C, Crump JA, Pitzer VE. Seasonal dynamics of
626 typhoid and paratyphoid fever. *Scientific Reports* 2018 8:1 **2018**; 8:1–9.
- 627 53. Karkey A, Thompson CN, Tran Vu Thieu N, et al. Differential Epidemiology of Salmonella
628 Typhi and Paratyphi A in Kathmandu, Nepal: A Matched Case Control Investigation in a
629 Highly Endemic Enteric Fever Setting. *PLoS Negl Trop Dis* **2013**; 7:e2391.
- 630 54. Liu B, Knirel YA, Feng L, et al. Structural diversity in Salmonella O antigens and its genetic
631 basis. *FEMS Microbiol Rev* **2014**; 38:56–89.
- 632 55. Mylona E, Sanchez-Garrido J, Hguyen Hoang Thu T, et al. Very long O-antigen chains of
633 Salmonella Paratyphi A inhibit inflammasome activation and pyroptotic cell death. *Cell*
634 *Microbiol* **2021**; :1–14.
- 635 56. Hiyoshi H, Wangdi T, Lock G, et al. Mechanisms to Evade the Phagocyte Respiratory Burst
636 Arose by Convergent Evolution in Typhoidal Salmonella Serovars. *Cell Rep* **2018**; 22:1787–
637 1797.
- 638 57. Johnson R, Mylona E, Frankel G. Typhoidal Salmonella: distinctive virulence factors and
639 pathogenesis. *Cell Microbiol* **2018**; 20:e12939.



641

642 **Figure 1.** Epidemiological observations of *S. Paratyphi A* patients in Nepal

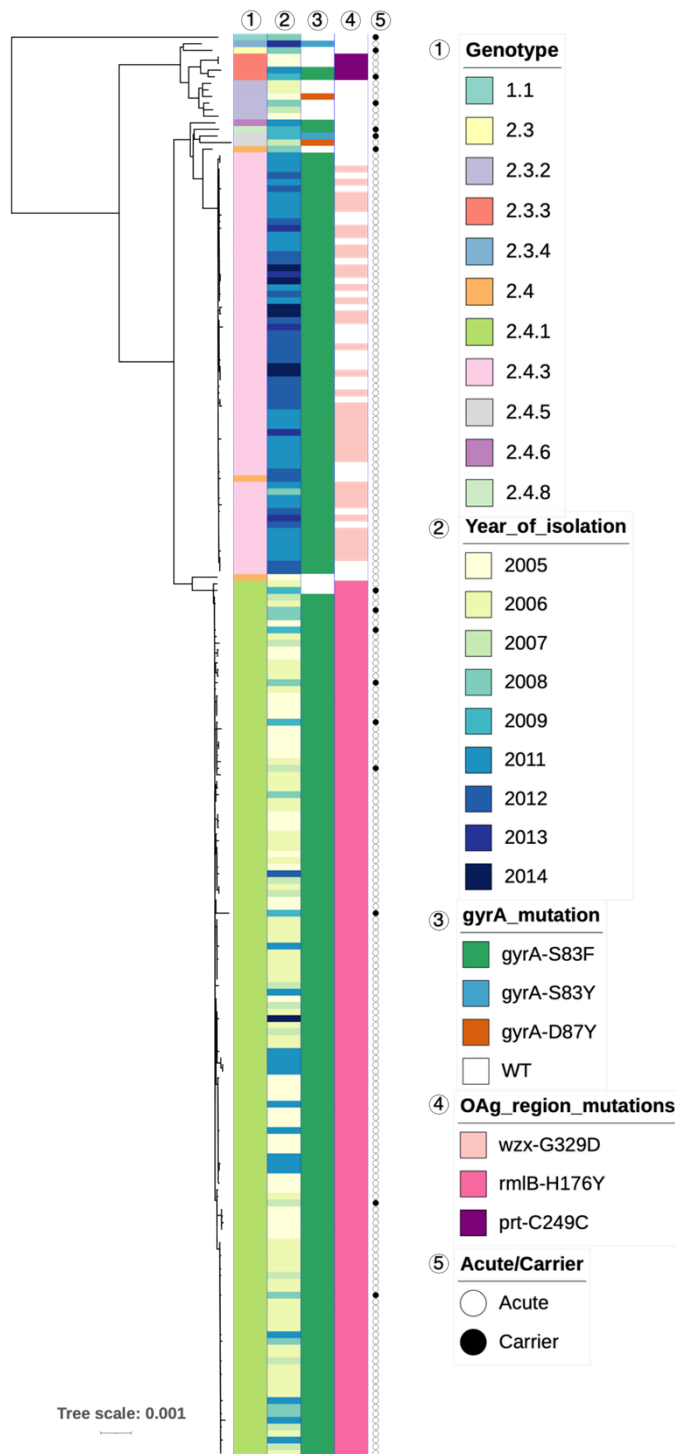
643 (A) Geographic map of Nepal showing the precise residence location of acute enteric fever patients

644 (red) and carriers (blue). (B) Numbers of acute enteric fever patients (200) reporting symptoms of

645 headache, constipation, diarrhoea, abdominal pain, nausea, and/or anorexia. (C) Monthly distribution

646 of acute enteric fever cases during 2005–2014.

647



648

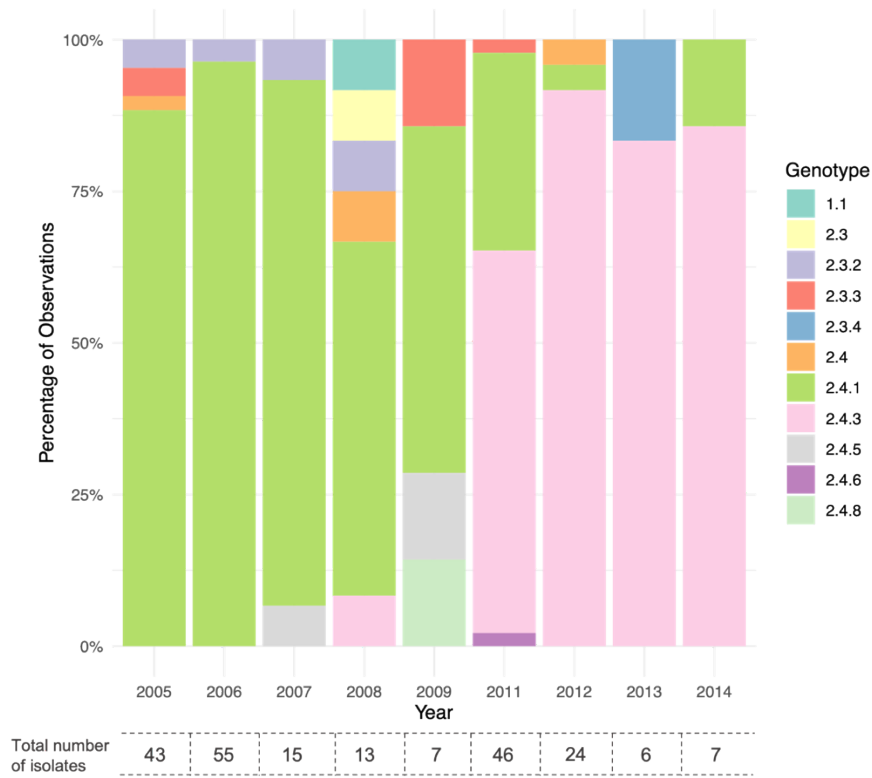
649 **Figure 2.** The population structure of *S. Paratyphi A* in Nepal in 2005–2014

650 Rooted SNP-based maximum likelihood phylogenetic tree of 200 acute and 16 carrier *S. Paratyphi A*

651 isolates from Nepal. Annotations are marked with a number on top that corresponds to the numbers on

652 the key. Tree was constructed with the reference strain included, which was removed at the

653 visualisation stage.



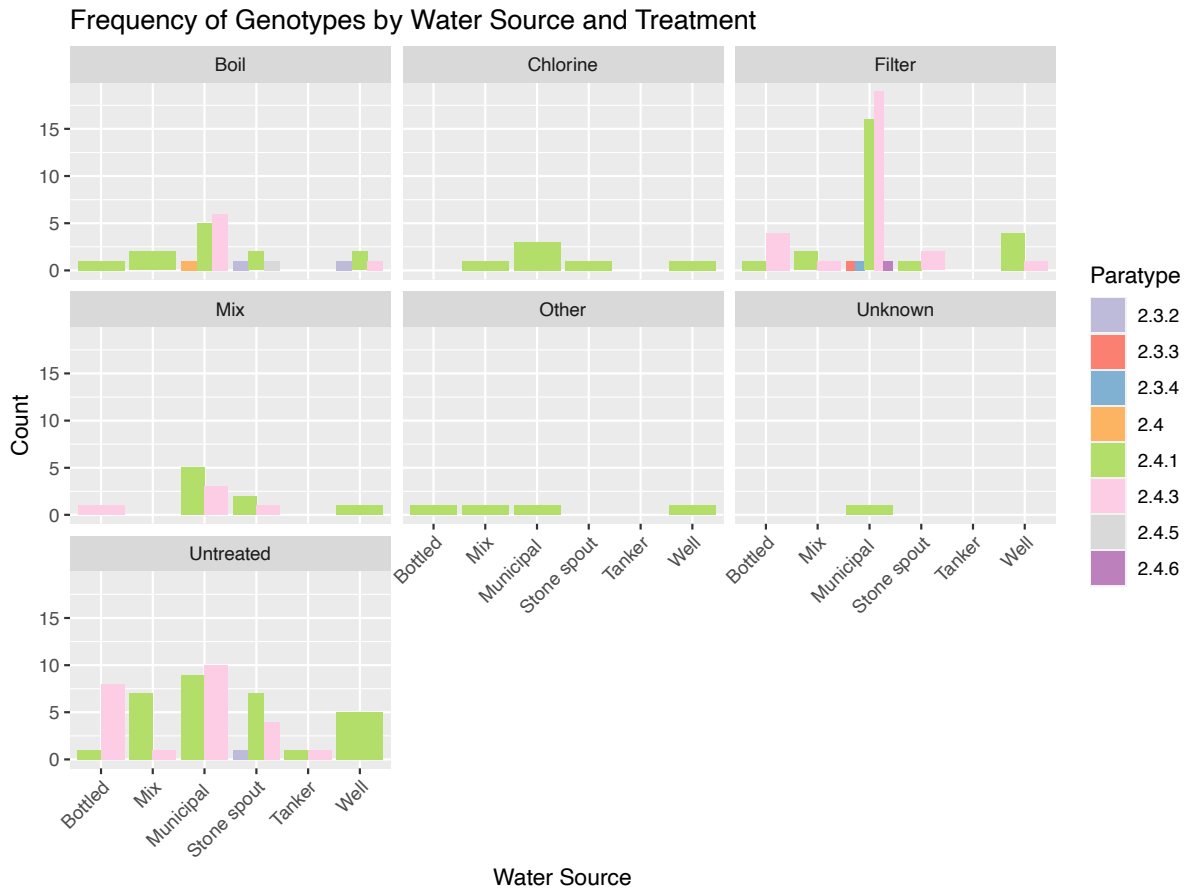
654

655 **Figure 3.** Annual distribution of *S. Paratyphi A* genotypes in Nepal, 2005-2014

656 Stacked bar plot showing the percentage of each genotype in each year. No isolates sampled in 2010.

657 Numbers below bar correspond to the absolute total numbers of isolates in each year.

658

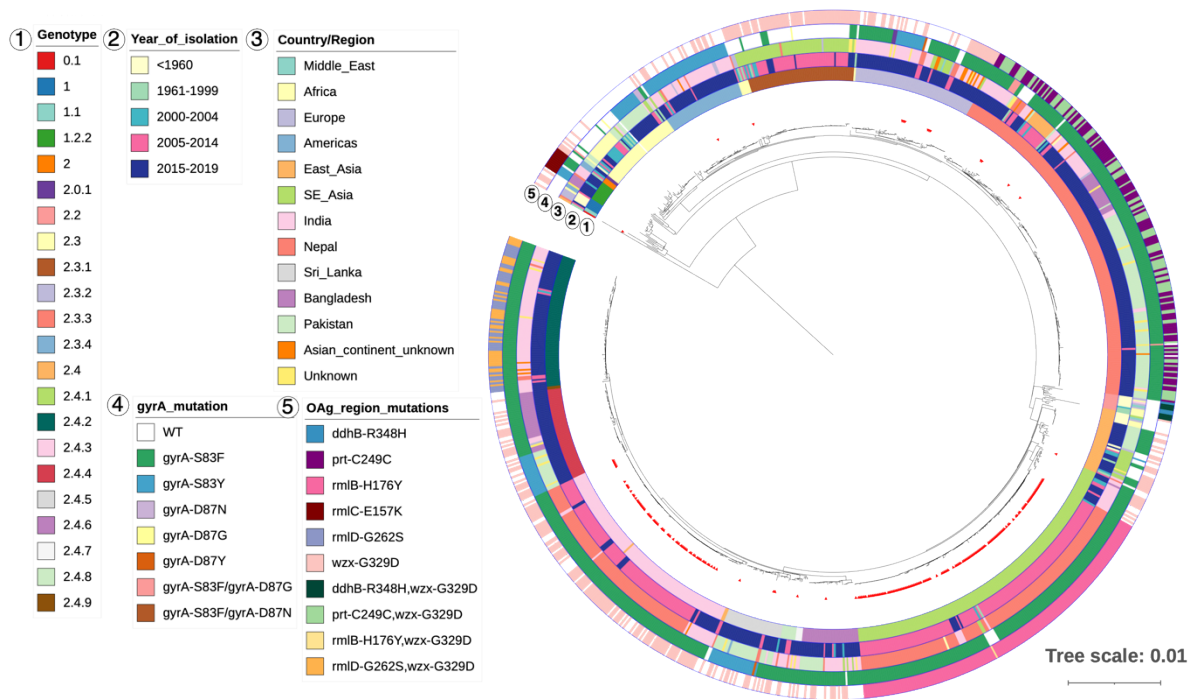


659

660 **Figure 4.** Frequency of each genotype in each water source, grouped by treatment method

661 Bar plots showing the absolute numbers of each genotype per water source for each treatment method.

662



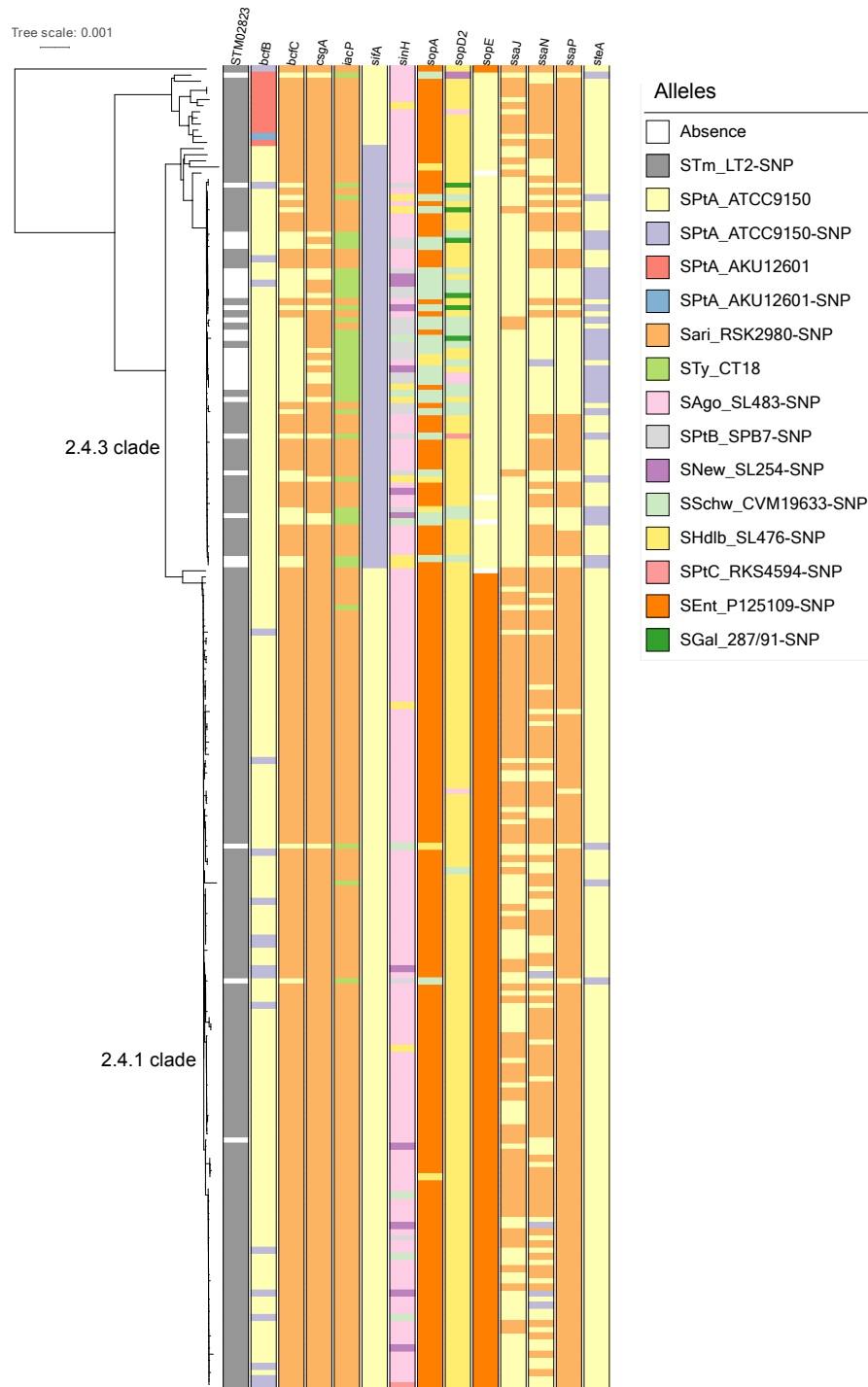
663

664 **Figure 5.** The Nepali *S. Paratyphi A* collection in a global context

665 Rooted SNP-based maximum likelihood phylogenetic tree of 828 global isolates and the 216 Nepali
 666 isolates (marked with red arrows). Annotation rings are marked with a number that corresponds to the
 667 numbers on the key. NB the tree was re-rooted on midpoint on the longest branch for visualisation
 668 purposes. Tree was constructed with the reference strain included, which was removed at the
 669 visualisation stage.

670

671



672

673 **Figure 6.** Genetic diversity in virulence factors in the Nepali *S. Paratyphi A* population

674 Virulence factors with genetic diversity in >20% of isolates as identified using the Virulence Factor

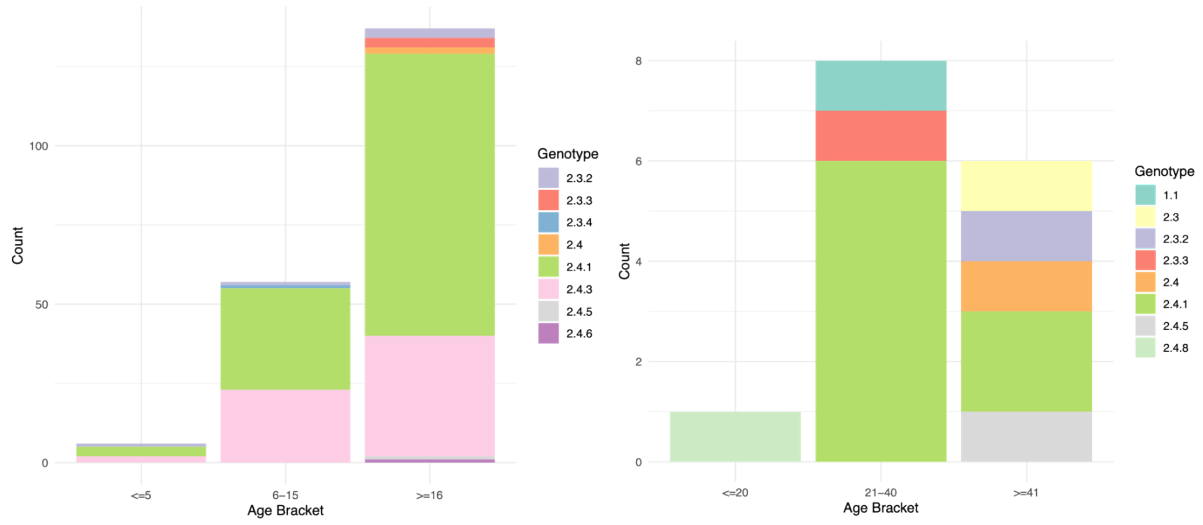
675 Database. Alleles identified are shown in the key (Serotype_strain ID), SNP indicates a mismatch in

676 the allele. STm, *S. Typhimurium*; SPtA, *S. Paratyphi A*; Sari, *S. Arizona*; STy, *S. Typhi*; SAgo, *S.*

677 Agona; SPtB, *S. Paratyphi B*; SNew, *S. Newport*; SSchw, *S. Schwarzengrund*; SHdlb, *S. Heidelberg*;

678 SPtC, *S. Paratyphi C*; SEnt, *S. Enteritidis*; SGal, *S. Gallinarum*.

679

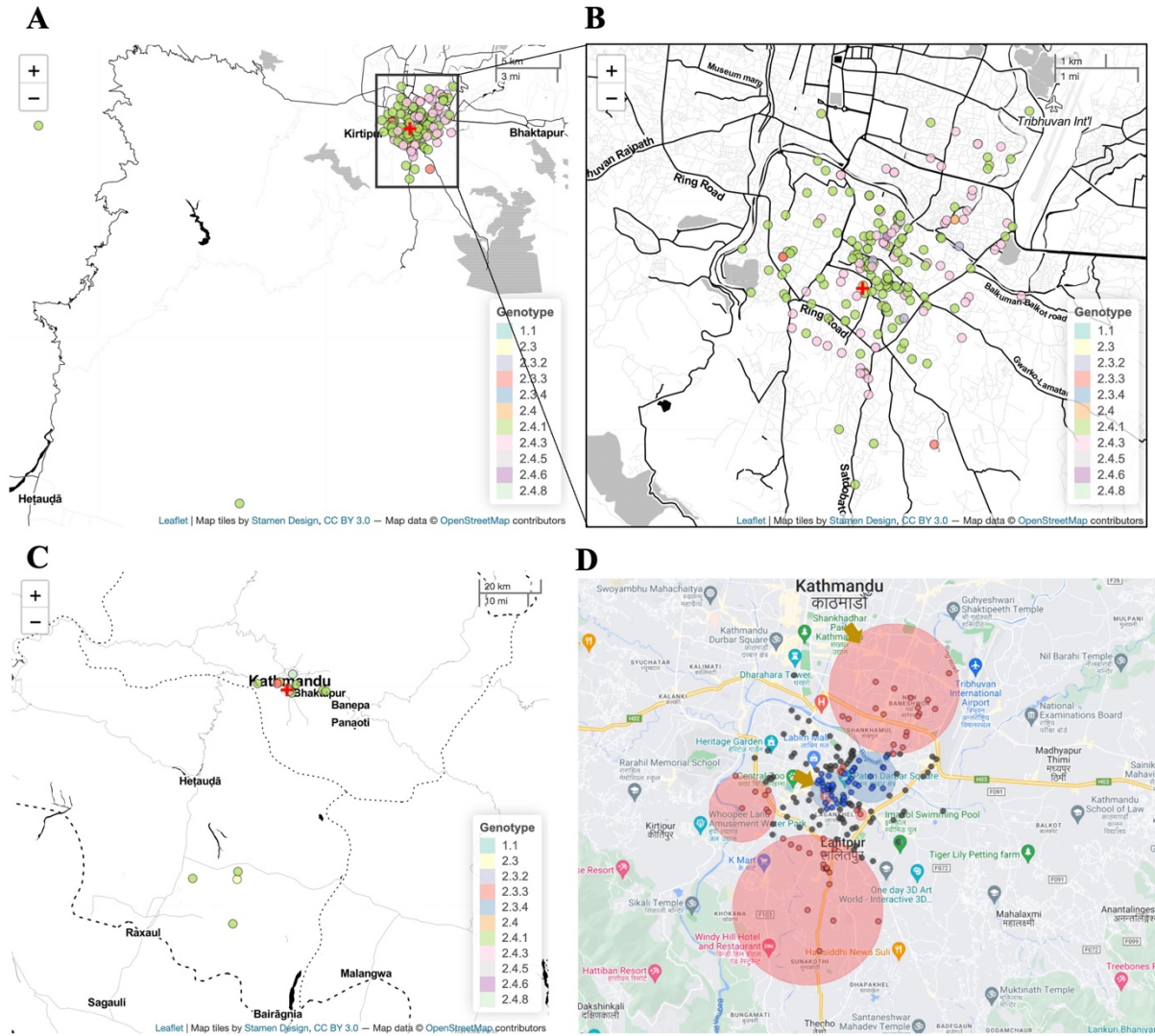


680

681 **Figure S1. *S. Paratyphi A* genotype distribution by age of acute enteric fever patients (left) and**

682 **chronic carriers (right). Bar charts show counts of patients in each age bracket.**

683



684

685 **Figure S2. Geographical distribution of *S. Paratyphi A* genotypes. (A) Genotypes of acute *S.***

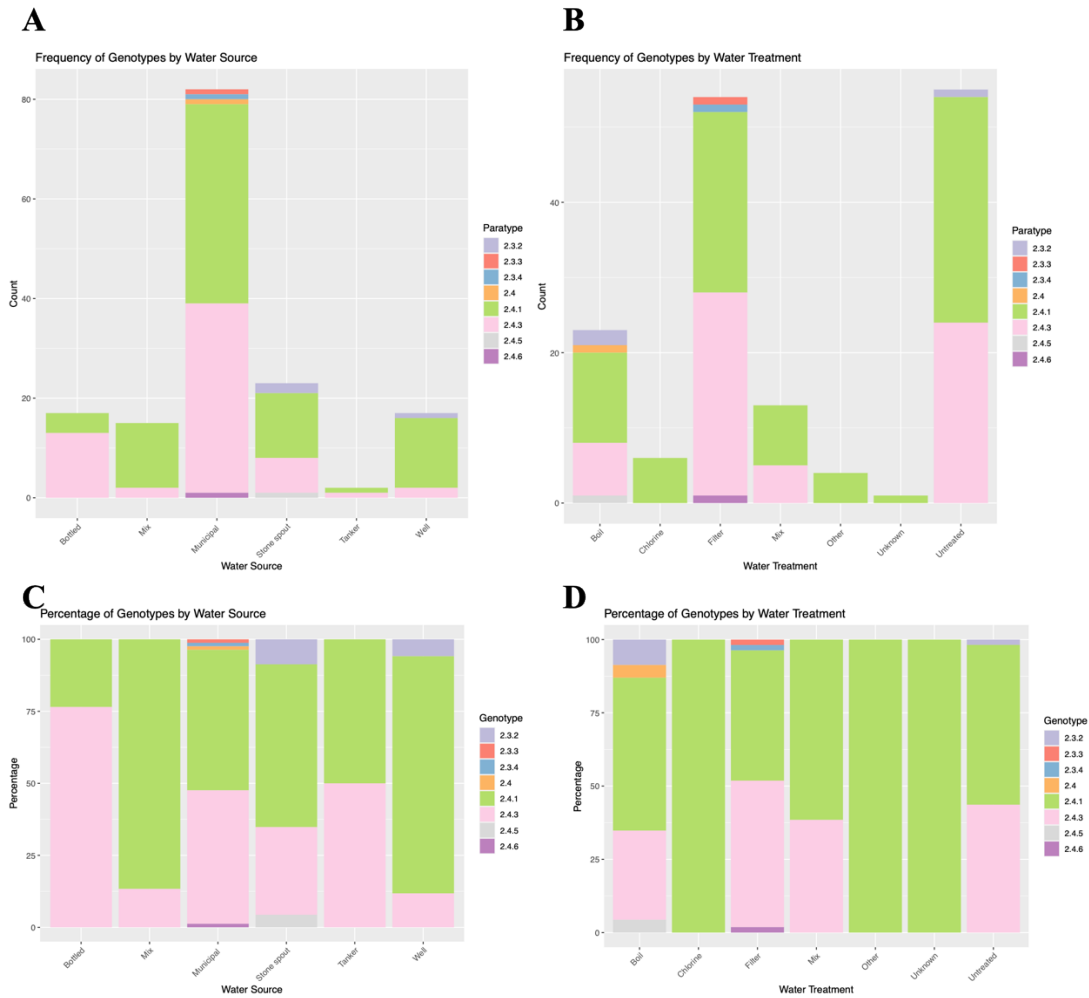
686 *Paratyphi A* organisms with (B) a zoomed in inset of the majority of cases. (C) Genotypes of carrier *S.*

687 *Paratyphi A* organisms. (D) Spatiotemporal clustering of *S. Paratyphi A* with red circles showing high

688 rate clusters and the blue circles indicating low rate clusters (statistically significant clusters, $p < 0.05$,

689 are indicated with yellow arrows). Red cross indicates the Patan Hospital.

690



691

692

693

694

695

Figure S3. Genotype distribution with water source and water treatment from 156 enteric fever patients. Stacked bar plots showing the (A-B) count and (C-D) percentages of each genotype for each associated water source (A-C) and water treatment (B-D).