Multiple introductions of multidrug-resistant typhoid associated with acute infection and asymptomatic carriage, Kenya

*Samuel Kariuki[1,2], *Zoe A Dyson[2,3,4,5], Cecilia Mbae[1], Ronald Ngetich[1], Susan M Kavai[1], Celestine Wairimu[1], Stephen Anyona[1], Naomi Gitau[1], Robert Onsare[1], Beatrice Ongandi[1], Sebastian Duchene[6], Mohamed Ali[7], John Clemens[8], *Kathryn E Holt[4,5], *Gordon Dougan[3]

Affiliations:

1. Centre for Microbiology Research, Kenya Medical Research Institute, Kenya.
2. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, United Kingdom
3. Cambridge Institute of Therapeutic Immunology & Infectious Disease (CITIID), Department of Medicine, University of Cambridge, Cambridge, UK.
4. London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK
5. Department of Infectious Diseases, Central Clinical School, Monash University, Melbourne, Victoria 3004, Australia
6. Department of Microbiology and Immunology, The University of Melbourne at The Peter Doherty Institute for Infection and Immunity, Melbourne, VIC, Australia
7. Department of International Health, John's Hopkins University, USA.
8. International Diarrheal Diseases Research Centre, Dhaka, Bangladesh.

*Indicates equal contribution

Corresponding author:

Prof Samuel Kariuki

Director, Research and Development, KEMRI, Kenya

Address: Off Mbagathi Road, Nairobi, PO Box 54840-00200, Nairobi, Kenya

Mobile: +254-722-232467; Email address: Samkariuki2@gmail.com/skariuki@kemri.org

Abstract

Understanding the dynamics of infection and carriage of typhoid in endemic settings is critical to finding solutions to prevention and control. In a 3 year case-control study, we investigated typhoid among children aged <16 years (4,670 febrile cases and 8,549 age matched controls) living in an informal settlement, Nairobi, Kenya. 148 *S*. Typhi isolates from cases and 95 from controls (stool culture) were identified; a carriage frequency of 1%. Whole-genome sequencing showed 97% of cases and 88% of controls were genotype 4.3.1 (Haplotype58), with the majority of each (76% and 88%) being multidrug-resistant strains in 3 sublineages of H58 genotype (East Africa 1 (EA1), EA2, and EA3), with sequences from cases and carriers intermingled. The high rate of multidrug-resistant H58 *S.* Typhi, and the close phylogenetic relationships between carriers and controls, provides evidence for the role of carriers as a reservoir for the community spread of typhoid in this setting.

Introduction

Typhoid fever, caused by *Salmonella enterica* serovar Typhi (*S*. Typhi) is estimated to involve ~21.7 million illnesses and 216,000 deaths annually [1,2], with most of these occurring in lower and middle-income countries. In Africa, overall typhoid is now estimated to have an average annual pooled incidence rate of 112.1 (95% CI, 46.7–203.5) cases per 100,000 people [3,4] with a case fatality rate (CFR) of 5.4% (2.7-8.9)[5].

Control of typhoid is impeded by asymptomatic carriage, which historically was estimated to account for 2-5% of individuals infected [6-8]. However, there is a paucity of recent data on the frequency of carriers in different settings including sub-Saharan Africa (SSA) as well as the extent to which they contribute to disease transmission [9]. A recent modelling study using data generated in Blantyre, Malawi, identified multidrug resistant (MDR) *S*. Typhi and/or the emergence of the lineage known as H58 (genotype 4.3.1) as a primary driver of an increasing number of typhoid fever cases. In this study, an estimated 45-95% of typhoid transmission was attributed to carriers [10,11]. *S*. Typhi H58 [12] is a globally disseminated clade frequently associated with MDR (defined as resistance to chloramphenicol, ampicillin and co-trimoxazole) and an increasing frequency of reduced susceptibility to fluoroquinolones. H58 *S.* Typhi are rapidly displacing other lineages in many endemic areas [12-15] and a new subclade that is extensively drug resistant (XDR), displaying resistance to ciprofloxacin and fluoroquinolones in addition to MDR, has been described in Pakistan [16].

Recent reports of epidemics of typhoid fever in SSA suggest that the disease may be becoming more widespread in the region [1,15,17-20]. In Kenya, the rapid growth of population has led to a huge rural-to-urban migration with people increasingly living in informal settlements where clean water and good sanitation are a major challenge [21,22]. The incidence of typhoid in one such informal settlement, Kibera in Nairobi, was estimated at 247 cases per 100,000 with the highest rates in children 5-9 years old (596 per 100,000) [23]. For the last two decades the majority of cases of typhoid in Kenya have been MDR, with reduced susceptibility to fluoroquinolones rising in frequency [14,15,24].

Previously, we showed that *S*. Typhi H58 gained a foothold in Kenya in the 1990s, constituting >75% of the circulating *S*. Typhi we have characterized since 2001 [14]. Two H58 lineages were detected; lineage I being isolated between 1988-2008 and lineage II from 2004 onwards. We have previously observed carriage rates of 6% in households where typhoid cases were detected [14], however these *S*. Typhi isolates were not characterized genetically and the role of asymptomatic carriers in transmission dynamics of typhoid in the community is still poorly understood. Over the past 7 years we have been intensively studying typhoid and other invasive bacterial diseases in Mukuru, an informal settlement 15 km east of the city of Nairobi, Kenya. The prevalence of *S.* Typhi infections among 16,236 children was 1.4% (CI: 1.2-1.6%), and higher amongst males (1.8% vs. 1.2% for females), with a high proportion of infections noted among older children 5-8 years in age[25]. Risk factors predictive of *S.* Typhi infection in Mukuru were multiple but were predominantly associated with contaminated water sources and sanitation issues [25]. Here, we analysed typhoid cases in Mukuru clinically and microbiologically, and identified frequent asymptomatic carriage among children below 16 years of age. By exploiting whole genome sequencing (WGS) and geospatial mapping we characterised the population structure and transmission dynamics of *S*. Typhi in this location.

Materials and Methods

Study site

Mukuru informal settlement is situated East of Nairobi city, about 15 km from the city centre. It is one of the largest slums in the city with a population of around 250,000

people [26]. The informal settlement is made up of improvised temporary dwellings often made from scrap materials, such as corrugated metal sheets, plywood, and polythene-sheets [27]. In addition to poverty, a number of factors associated with informal settlements, including overcrowding, substandard housing, unclean and insufficient quantities of water, and inadequate sanitation, contribute to a high incidence of infectious diseases and increased mortality among children under five years [21,28]. Mukuru informal settlement is divided into eight villages; Mukuru Lunga-Lunga, Mukuru kwa Sinai, Mukuru kwa Ruben, Mukuru kwa Njenga, Mukuru Kayaba, Fuata Nyayo, Jamaica, and Mukuru North. This study was carried out in two of the large villages, Mukuru kwa Njenga and Mukuru kwa Ruben, with a combined population of 150,000. Spatial mapping of the two villages was conducted using the Universal Transverse Mercator system [29], and patient details collected as described previously [25].

The two villages in the informal settlement are served by three outpatient clinics: Ruben Health Centre located in the Ruben village (zone named Simba cool, serves approximately 30% of the population), Missionaries of Mary Located in Kwa Njenga village (zone named Vietnam, serves approximately 45% of the population), and County Government Clinic in Kwa Njenga village (Zone named MCC and serves approximately 25% of the population). The fourth site, Mbagathi District Hospital, is located on the western side of Nairobi city, 5 km from city centre and was used as a referral facility. Participants living outside of the mapped demographic surveillance site (DSS) who came to seek medical services in any of the three study site health facilities or at

Mbagathi District Hospital were included for the purpose of tracking typhoid cases and carriers treated at the facilities, but are reported separately in the results section.

Recruitment of clinical typhoid fever cases and asymptomatic typhoid carriers

Typhoid fever cases and asymptomatic carriers presented in this study were identified and recruited as part of a larger study on surveillance and genomics of invasive *Salmonella* disease in children and young adults less than 16 years of age [25,30]. Children presenting as outpatients at the three study clinics and Mbagathi District Hospital between August 2013 and November 2016 were triaged to identify those with fever, headache and/or diarrhoea for recruitment into the study as potential cases. Patients with current fever (≥38°C) and reportedly febrile for ≥3 days were considered potential typhoid cases and assessed via blood culture. The primary typhoid case definition (data presented in **Table 1**) was children aged 0-16 years with ≥3 days fever ≥38°C and positive blood or stool culture for *S*. Typhi (see bacterial culture methods below).

During the study period, age-matched controls were recruited from children without current fever or diarrhoea attending the same health facilities for healthy mother and child clinics (e.g. for vaccination and nutritional advice). Those with *S*. Typhi positive stool culture were designated as asymptomatic typhoid carriers as described previously [25]. Hence, the inclusion criteria for asymptomatic typhoid carriers (data presented in **Table 1**) were children aged 0-16 years with no diarrhoea, no current fever, and no recent fever history, with stool culture positive for *S*. Typhi (see bacterial culture

methods below). The total number of participants was computed on the basis of a 4% prevalence rate of typhoid from previous study[14] . (A structured questionnaire was used to collect demographic data for both cases and controls recruited into the study as described previously [25].

All isolates cultured from participants and identified as *Salmonella* were archived and later revived for WGS as detailed below. The sequence data revealed some mis-identification of *Salmonella* serotypes (**Fig. 1 and Table S1**), hence for genomic analyses we included all cases and controls whose cultures were found to be *S*. Typhi positive by WGS rather than those identified as *S*. Typhi positive by serotype in the microbiology laboratory.

Bacterial culture

For blood culture, 1-3 mL for children <5 years of age and 5-10 mL for those 5-16 years of age was collected in a syringe, placed into Bactec media bottles (Becton-Dickinson, New Jersey, USA), incubated at 37^{o}C in a computerized BACTEC™ 9050 Blood Culture System (Becton-Dickinson), and subcultured after 24-48 h onto blood, chocolate and MacConkey agar (Oxoid, Basingstoke, UK) plates. For stool culture, rectal swabs or stool samples were obtained from each potential carrier and cultured on selenite F (Oxoid) broth aerobically at 37^{o}C overnight. Broth cultures were then subcultured on MacConkey agar and *Salmonella-Shigella* agar (Oxoid) and incubated at 37^{o}C overnight. Blood and stool isolates were identified using a series of standard biochemical and serological tests as described previously [25].

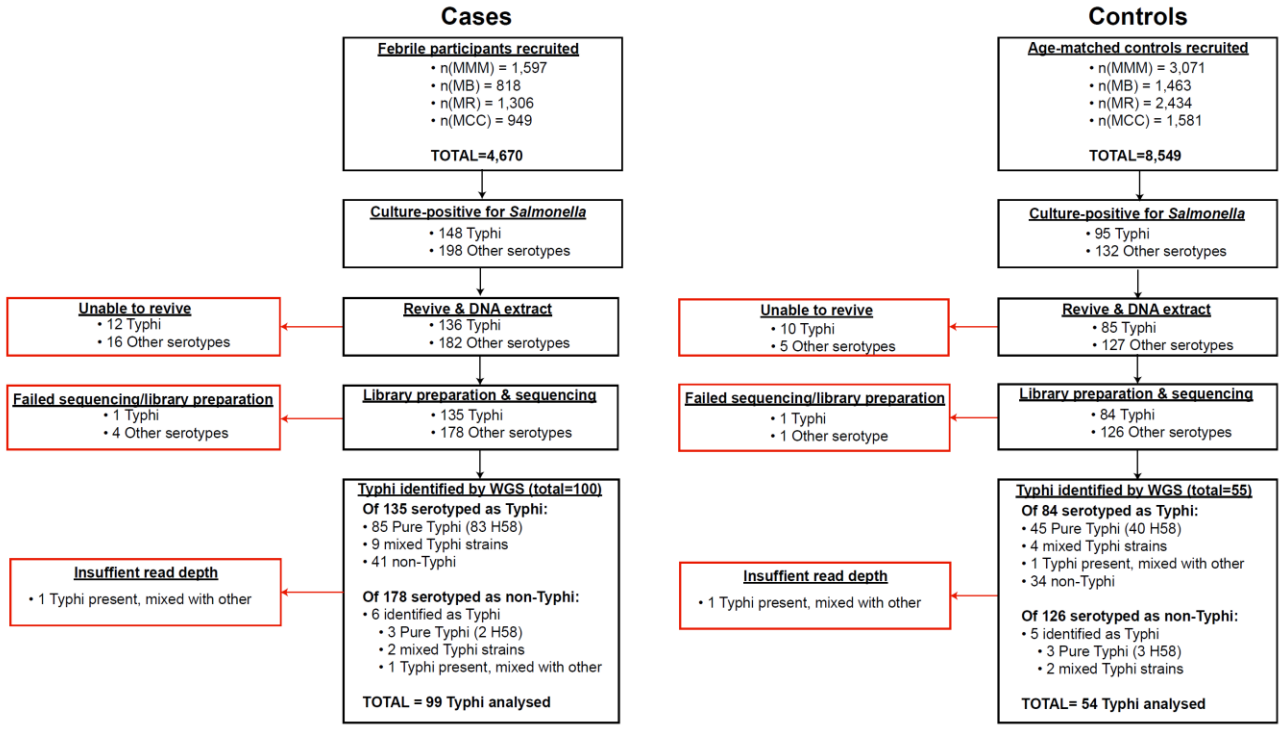
Antimicrobial susceptibility testing

Antimicrobial susceptibility testing was performed using the disk diffusion technique for ampicillin 10 µg, tetracycline 30 µg, co-trimoxazole 25 µg, chloramphenicol 30 µg, cefpodoxime 30 µg, ceftazidime 30 µg, ceftriaxone 30 µg, cefotaxime 30 µg, ciprofloxacin 5 µg and nalidixic acid 10 µg as described previously [25]. Results were interpreted according to the 2017 guidelines provided by the Clinical and Laboratory Standards Institute (CLSI) [31].

Whole genome sequencing

All *Salmonella* isolated from cases and controls were subcultured at the end of the study for DNA extraction and WGS. These included 243 cultures identified as *S*. Typhi from cases (85 from blood and 63 from stool) and 95 from controls (all from stool) (see **Fig. 1** and **Table S1)**, which are the subject of this study (non-typhoidal *Salmonella* data is reported elsewhere [30]). Twelve *S*. Typhi case isolates and 10 control isolates could not be revived and were not further analysed. DNA was extracted using the Wizard Genomic DNA Extraction Kit (Promega, Wisconsin, USA) and shipped on ice to the Wellcome Sanger Institute for sequencing using the Illumina platform as described previously [15]. A total of 217 *S*. Typhi DNA samples were successfully sequenced (two were of insufficient quality to construct sequencing libraries, or failed sequencing). Non-*S*. Typhi bacterial DNA sequences were detected in 75 samples (34.6%; organisms detected are shown in **Table S1**), and 11 sequences originally identified as other *Salmonella* serotypes were later found to be *S.* Typhi with genomic data. Two

188 sequences showed the presence of *S.* Typhi, but at low depth, and were subsequently
189 omitted from further genomic analyses, leaving genome data for 153 *S.* Typhi isolates
190 for further analysis.



191

192 **Fig. 1. Flow chart of samples collected and analysed**. Red boxes indicate bacterial
193 isolates that could not be included in downstream genetic analyses, grouped by reason
194 for exclusion.
195

196

197 ***Phylogenetic and SNP analysis of S.* Typhi *isolates***

198 For SNP analysis, paired-end reads from 153 *S.* Typhi isolates were mapped to the
199 reference sequence of *S.* Typhi CT18 (accession number: AL513382) [32] using the
200 RedDog mapping pipeline (v1beta.10.3), available at http://githib.com/katholt/reddog
201 and detailed in **supplementary methods.** Read alignments were used to assign
202 isolates to previously defined lineages according to the extended genotyping framework

[33,34] with the GenoTyphi pipeline (available at http://github.com/katholt/genotyphi). Unique SNPs defining three novel lineages were identified from the genome-wide SNP allele table and added to the GenoTyphi scheme to facilitate easy identification of these lineages in future studies (details in **supplementary methods and results**).

Phylogenetic analyses were restricted to WGS-confirmed pure cultures of *S*. Typhi H58 (genotype 4.3.1, n=128). For some analyses, an additional 1,076 *S*. Typhi H58 genomes from previously published WGS studies of global and African isolates [12,15,33,35] were also included for context, along with 61 non-H58 genomes for phylogenetic outgroup rooting, using the same mapping approach detailed above (see **Table S2** for full list of genomes analysed and their public data accessions). SNPs called in phage regions or repetitive sequences were filtered from the alignment (details in **supplementary methods**), and any further recombinant regions identified and removed with Gubbins (v2.3.2) [36]. This resulted in a final set of 8,635 SNPs. From this global alignment we extracted a separate SNP alignment for the set of 239 Kenyan *S*. Typhi 4.3.1 genomes (n=128 from this study and n=111 from published studies, see **Table S3** [15,33]), the resulting alignment of length 489 SNPs was used for temporal analyses (described below and in **supplementary methods**).

Maximum likelihood (ML) phylogenetic trees were inferred from SNP alignments using RAxML (v8.2.9) [37] (as detailed in **supplementary methods**) and the resulting trees were visualized using Microreact (interactive global H58 phylogeny available at: https://microreact.org/project/wViqmaRdZuFVEb6yk4i1jU) [38].

Pairwise SNP distances were calculated from the SNP alignment using the dist.dna() function in the R package *ape* (v5.4.1) [39]. Terminal branch lengths were extracted from phylogenies using R package *ggtree* (v2.2.4) [40]. Non-synonymous mutations occurring in terminal branches were detected using SNPPar (v0.4.2dev) [41] and grouped by function based on the gene in which they were found, according to the functional classification scheme in the genome annotation of *S.* Typhi CT18 [8,32].

Phylodynamic analysis

To investigate temporal signal and date the introduction of *S*. Typhi H58 into Kenya based on the 239 available Kenyan genomes (n=128 from this study, and n=111 from previous studies [15,33]), we used several methods. First, we used TempEst (v1.5.1) [42] to assess temporal structure (i.e. clock-like evolution) by conducting a regression analysis of the root-to-tip branch distances of the ML tree as a function of sampling date, and later a date-randomisation test (full details of temporal signal assessment and model selection are provided in **supplementary methods**). To estimate divergence dates for the three *S.* Typhi H58 sublineages we detected in Kenya (EA1-3), we used BEAST (v1.10) [43] to fit a phylodynamic model to the SNP alignment and isolation dates as described in **supplementary methods**. The resultant MCC tree was visualized using *ggtree* (v2.2.4) [40] and Microreact [38] (interactive phylogeny available at: https://microreact.org/project/I2KUoasUB).

Genomic determinants of antimicrobial resistance

The read mapping-based allele typer SRST2 (v0.2.0) [44] was used to detect the presence of plasmid replicons (PlasmidFinder database [45]) and antimicrobial resistance

(AMR) genes (ARGannot database [46]). Where AMR genes were observed without evidence of a known AMR plasmid, raw read data was assembled using Unicycler (v0.4.7) [47] and then examined using Bandage (v0.8.1) [48] to confirm the chromosomal location and composition of AMR-associated transposons. ISMapper (v2.0) [49] was also used to identify the location of IS*1* insertion sequences in the *S*. Typhi chromosome as described in **supplementary methods**. Point mutations located within the quinolone resistance determining region (QRDR) of genes *gyrA, gyrB,* and *parC* associated with reduced susceptibility to fluoroquinolones [35] were detected using GenoTyphi [33,34] as detailed in **supplementary methods**.

Statistical and spatial analysis

All statistical analyses unless otherwise stated were carried out using R (v4.0.2). Details of specific functions within R packages used for individual analyses are available in **supplementary methods**.

Nucleotide sequence and read data accession numbers

Raw Illumina sequence reads have been submitted to the European Nucleotide Archive (ENA) under accession PRJEB19289. Individual sequence accession numbers are listed in **Table S1**.

Ethical Considerations

The study was approved by the Scientific and Ethics Review Unit (SERU) of the Kenya Medical Research Institute (KEMRI) (Scientific Steering Committee No. 2076). All

parents and/or guardians of participating children were informed of the study objectives and voluntary written consent was sought and obtained before inclusion. A copy of the signed consent was filed and stored in password protected cabinets at KEMRI.

Results

Detection of S. Typhi cases and asymptomatic carriers

From August 2013-November 2016, a total of 4,670 febrile children were recruited across the four study sites and subjected to blood and/or stool culture. *S*. Typhi was identified in cultures from 148 children (3.2%); the annual rate was steady over the study period but significantly higher amongst males (4.0% vs 2.3%, p=0.0008, see **Table 1**). The odds of *S*. Typhi positive culture increased significantly with age (OR 1.08, p=0.0005) but the effect was restricted to males (see **Table 1**), amongst whom the isolation rate was 1.3% in those ≤1 year, 2.0% in those aged 1-7 years, and 3.4% in those >7 years old (compared with 0.95%, 1.1% and 0.94%, respectively amongst females). A total of 8,549 age-matched control participants (with no current diarrhoea and no recent fever history) were recruited and subjected to stool culture. *S*. Typhi was identified in cultures from n=95 (1.1%); these are considered asymptomatic carriers. *S*. Typhi culture positivity amongst controls was not significantly associated with age or sex and was stable over the study period (see **Table 1 and Table S4**). No significant statistical association was found between phenotypic or genotypic AMR patterns and case/control status, age, or sex.

Table 1. Culture positive typhoid cases and asymptomatic carriers

	Cases	Controls
Participants tested, N	4,670	8,549
Male, N (%)	2,497 (53.5%)	4,260 (49.8%)
Female, N (%)	2,173 (46.5%)	4,289 (50.2%)
***S*. Typhi culture positive, N (%)**	148 (3.2%)	95 (1.1%)
Male, N (%)	99 (4.0%)	49 (1.15%)
Female, N (%)	49 (2.3%)	46 (1.1%)
WGS confirmed *S*. Typhi, N (%)	100 (2.1%)	55 (0.64%)
Logistic regression for *S*. Typhi culture positive		
Year of isolation, OR (p-value)	1.19 (0.072)	0.94 (0.586)
Male Sex, OR (p-value)	1.81 (0.0008*)	1.08 (0.699)
Age in years, OR (p-value)	1.08 (0.0005*)	1.02 (0.403)
Logistic regression for *S*. Typhi culture positive, males only		
Year of isolation, OR (p-value)	1.19 (0.147)	1.09 (0.576)
Age in years, OR (p-value)	1.11 (0.0001*)	1.06 (0.082)
Logistic regression for *S*. Typhi culture positive, females only		
Year of isolation, OR (p-value)	1.19 (0.296)	0.81 (0.158)
Age in years, OR (p-value)	1.03 (0.551)	0.98 (0.534)

Note the values reported for logistic regressions are from multivariate models including all indicated covariates, fit separately for cases and controls.

Global population structure and antimicrobial resistance profiles of Kenyan S. Typhi

The presence of *S*. Typhi was confirmed by WGS in 94 cases (64%) and 50 controls (53%) that were originally identified as *S*. Typhi via microbiological culture (**Fig. 1, Table S1**). *S.* Typhi genotype 4.3.1 (H58) was dominant throughout the study (n=145, 95%), amongst both cases and controls **(Table 2)**. Five other genotypes were detected: 2.2.2 (n=1), 2.5.0 (n=3), 3.0.0 (n=3), and 4.1.1 (n=1), see **Table 2**.

Table 2. Genotypes and AMR profiles for 153 sequenced *S.* Typhi isolates

Genotype	Cases	Controls	MDR	GyrA mutation	GyrB

								mutation
			Plasmid	Chromosome	S83F	S83Y	D87G	S464F
All	99	54	83	33	3	17	2	75
2.2.2	0	1 (1.9%)	0	0	0	0	0	0
2.5.0	1 (1.0%)	2 (3.7%)	0	0	0	0	0	0
3.0.0	2 (2.0%)	1 (1.9%)	0	0	0	0	0	0
4.1.1	0	1 (1.9%)	0	0	0	0	0	0
4.3.1 (H58)	96 (97%)	49 (91%)	83 (57%)	33 (23%)	4 (2.8%)	19 (13%)	2 (1.4%)	75 (51.7%)
H58 subgroups								
EA1 (L1)	35 (35%)	20 (37%)	29 (53%)	17 (31%)	4 (7.3%)	2 (3.6%)	2 (3.6%)	2 (3.6%)
EA2 (L2)	46 (46%)	27 (50%)	54 (74%)	0	0	0	0	73 (100%)
EA3 (L2)	15 (15%)	2 (3.7%)	0	16 (94%)	0	17 (100%)	0	0

Percentages indicate genotype frequencies amongst cases or controls (first two columns); or frequency of antimicrobial resistance determinants amongst isolates of a given genotype (remaining columns). MDR, multi-drug resistant; L1, lineage I; L2, lineage II.

The few non-H58 isolates (**Table 2**) lacked any known AMR determinants. In contrast, the majority of H58 isolates were MDR (n=116, 80%), often carrying acquired genes conferring resistance to ampicillin (*bla*TEM-1), chloramphenicol (*catA1*), co-trimoxazole (*dfrA7* plus *sul1* and/or *sul2*) and streptomycin (*strAB*). In 33 genomes (23% of H58), these genes were carried by a Tn*2670*-like complex transposable element inserted in the chromosome as reported previously in the region [12,13,15]. The remaining 83 MDR genomes (57% of H58) carried a closely related Tn*2670*-like transposon located within an IncHI1 plasmid, which in all but one isolate also carried an additional tetracycline resistance gene (*tetB*). The IncHI1 plasmids were genotyped as plasmid sequence type 6 (PST6), which is associated with MDR H58 in East Africa and South Asia [12,15,50]. Thus, the observed AMR phenotypes (n=128 H58 and n=8 non-H58 genome sequences) corresponded to the presence of known molecular determinants of AMR. Estimates of sensitivity and specificity of AMR genotyping are presented in **Table S5**

and **supplementary results**. No statistical association was observed between the presence of MDR genes or QRDR mutations shown in **Table 2** and case/control status, age, or sex.

Local subpopulations of *S*. Typhi H58

S. Typhi H58 (genotype 4.3.1) can be subdivided into lineages I (genotype 4.3.1.1) and II (genotype 4.3.1.2). Lineage II was more common in this setting than lineage I: n=90 (62.1% of H58) vs n=55 (37.9%). Examination of the global phylogeny (**Fig. 2**, and online interactive version https://microreact.org/project/wViqmaRdZuFVEb6yk4i1jU) revealed all H58 lineage I isolates from this study shared a most recent common ancestor (mrca) whose descendants form a monophyletic clade that exclusively comprised *S*. Typhi from East African countries (see **Fig. 2**), here defined as H58 sublineage EA1 (East Africa 1) with genotype designation 4.3.1.1.EA1 (labelled in **Fig. 2**).

S. Typhi H58 lineage II (genotype 4.3.1.2) isolates from our study belonged to two distinct clades of the global phylogeny (**Fig. 2**), which were each exclusively populated by East African isolates. The largest of these clades (n=80 isolates, of which 81.3% derive from the current study) formed a monophyletic group nested within a deeper clade of diverse South Asian isolates (see **Fig. 2**), and corresponds to the previously reported introduction of H58 lineage II into Kenya from South Asia [12]. This lineage, here defined as H58 sublineage EA2 (East Africa 2) is designated genotype 4.3.1.2.EA2 (labelled in **Fig. 2**). The smaller East African H58 lineage II clade (n=43 isolates) is

designated genotype 4.3.1.2.EA3 (labelled in **Fig. 2**) and comprised two sister clades, separated by ≥13 SNPs: one involving isolates from Kenya (n=13, all from this study) and the other isolates from Uganda (n=30), which accounted for 100% of the typhoid burden at the Ugandan site where they were identified (see **Fig. 2**). All three East African H58 genotypes have been added to the GenoTyphi scheme using unique marker SNPs and further details on these are provided in **supplementary results**.

The three East African H58 subgroups circulating in our setting all had high rates of MDR (84%, 74% and 94%, respectively); however, in EA2, MDR was exclusively associated with the PST6-IncHI1 plasmid, and in EA3 exclusively with the chromosomal insertion (see **Table 2**, **Fig. 2** and **supplementary results**). In EA1, most MDR was associated with the PST6-IncHI1 plasmid. However, a subclade of isolates (associated with spread to Tanzania and Malawi) carried the chromosomal insertion instead (see **Table 2**, **Fig. 2, supplementary methods**).

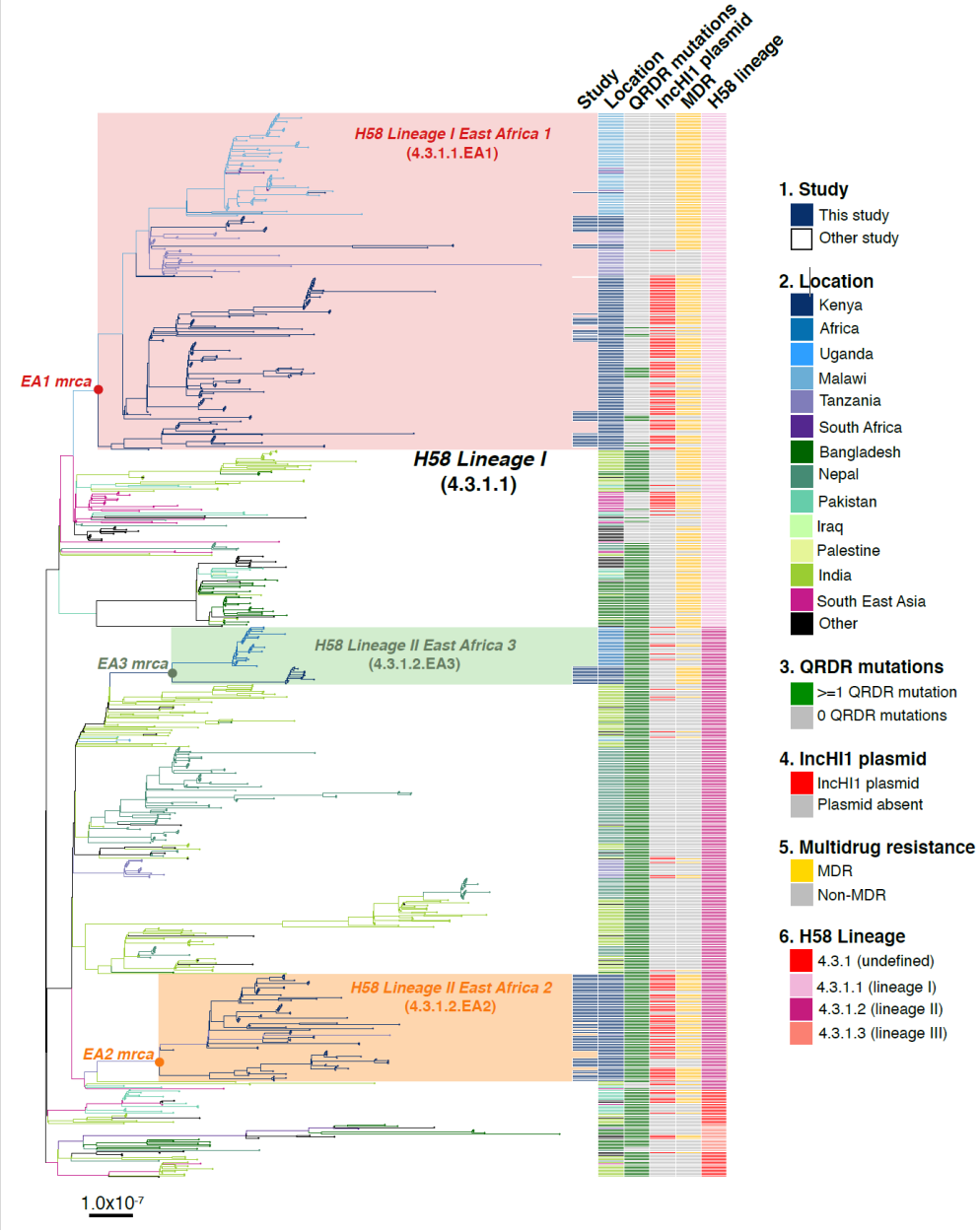


Fig. 2. Global population structure of H58 (4.3.1) *S*. Typhi showing Kenyan isolates cluster into three East African clades. Whole genome phylogeny of 1,204 H58 isolates, including all available Kenyan genomes (n=128 from this study, n=111 from prior studies) and globally distributed genomes for context (n=965, see **Methods**). Branch lengths are in substitutions per core-genome site, branches are coloured to indicate geographical origin (see inset legend), shaded boxes highlight the three East African H58 clades defined in this study. Colour bars to the right indicate (as per inset legend): 1, Kenyan strains isolated and sequenced during this study; 2, geographical location; 3, mutation(s) in the quinolone resistance determining region (QRDR) of genes *gyrA, gyrB,* and *parC*; 4, presence of multidrug resistance (MDR) IncHI1 plasmid; 5, presence of MDR genes; 6, H58 lineage. Interactive version available at https://microreact.org/project/wViqmaRdZuFVEb6yk4i1jU.

Distribution of S. Typhi *genotypes amongst individuals*

No statistically significant differences in genotype distribution were observed between cases and controls (p=0.077, using Chi-squared test, data in **Table 2**), or between males and females (p=0.37, using Chi-squared test, data in **Table S6**), consistent with symptomatic and asymptomatic infections being drawn from the same general circulating pool of pathogens. The distribution of genotypes amongst cases varied by age group (p=0.01, using Chi-square test), with the frequency of EA1 declining with age and the overall diversity increasing with age (**Table 3**). No significant differences in age groups was evident amongst controls (p=0.9 using Chi-square test, see **Table 3**).

Table 3. *S.* Typhi genotypes associated with n=153 cases and controls among different age groups

	Age group		
	≤1 year	**1-7 years**	**>7 years**
WGS-confirmed cases	**7**	**66**	**26**
EA1	5 (71%)	24 (36%)	6 (23%)
EA2	1 (14%)	34 (52%)	11 (42%)
EA3	0	8 (12%)	7 (27%)
non-H58	1 (14%)	0	2 (78%)
Shannon diversity	0.80	0.97	1.25
WGS-confirmed carriers	**4**	**30**	**20**
EA1	1 (25%)	10 (33%)	9 (45%)
EA2	3 (75%)	16 (53%)	8 (40%)
EA3	0	1 (3%)	1 (5%)
non-H58	0	3 (10%)	2 (10%)
Shannon diversity	0.56	1.05	1.11

Spatiotemporal distribution of S. Typhi *cases and carriers*

We examined the spatial and temporal distribution of all *S.* Typhi isolates collected at the study clinics (see **methods; Fig. S1 and Table S7**), and the subset of 96 *S*. Typhi from cases and 67 from carriers living within the demographic surveillance site (DSS) (**Fig. 3 and Table 4**). A number of peaks in monthly *S.* Typhi case and carrier numbers are apparent in both cohorts (**Fig. 3 and Fig. S1**), with fewer cases and carriers observed in warmer months. Carrier counts remained relatively consistent throughout the study period. We tested for association between case or carrier peaks (>2 positives per month) and high rainfall or temperature in the same month, previous month, or two months prior to the month of observation (**Table 4, and Table S7**). For those *S.* Typhi from within the DSS, high temperatures were associated with lower case and carrier counts in the same month, and in the subsequent month ($p<0.05$, using Fisher's exact test), however no associations between high rainfall and elevated case or carrier counts was observed.

GPS coordinates were available for n=139 (55%) *S.* Typhi isolates, and we endeavoured to look for geographic hotspots suggestive of major point-source single-genotype outbreaks in the informal settlement. However, our data revealed that the three H58 genotypes and non-H58 genotypes were co-circulating throughout the study area, with no evidence of geographic restriction of specific genotypes (see **Fig. S2**). Further, we did not observe any spatially linked phylogenetic clusters of closely related sequences.

418 **Table 4. Climatic predictors of elevated case and control counts inside the DSS**

Typhoid Cases						
Month	**Same month**		**Previous month**		**2 months prior**	
	OR (95% CI)	**p-value**	**OR (95% CI)**	**p-value**	**OR (95% CI)**	**p-value**
Rainfall (precipitation) > 75 mm	0.21 (0.019-1.2)	0.079	1.4 (0.26-6.9)	0.73	3.7 (0.73-22.3)	0.08
Minimum temperature >14°C	0.21 (0.041-0.95)	0.025*	0.61 (0.14-2.6)	0.52	2.2 (0.49-10.5)	0.33
Maximum temperature >26°C	0.85 (0.20-3.6)	1	0.37 (0.080-1.6)	0.20	0.67 (0.15-2.80)	0.75
Asymptomatic Controls						
Month	**Same month**		**Previous month**		**2 months prior**	
	OR (95% CI)	**p-value**	**OR (95% CI)**	**p-value**	**OR (95% CI)**	**p-value**
Rainfall (precipitation) > 75 mm	1.2 (0.21-6.5)	1	0.43 (0.038-2.7)	0.45	0.43 (0.038-2.7)	0.45
Minimum temperature >14°C	0.12 (0.016-0.64)	0.005*	0.41 (0.078-1.9)	0.30	0.65 (0.13-3.2)	0.73
Maximum temperature >26°C	0.10 (0.0090-0.61)	0.005*	0.19 (0.027-1.0)	0.04*	0.63 (0.12-3.0)	0.73

419 Values in cells are odds ratios and p-values for Fisher's exact test between high case or control count (>2
420 per month) and high rainfall/temperature. * highlights p-values <0.05.
421

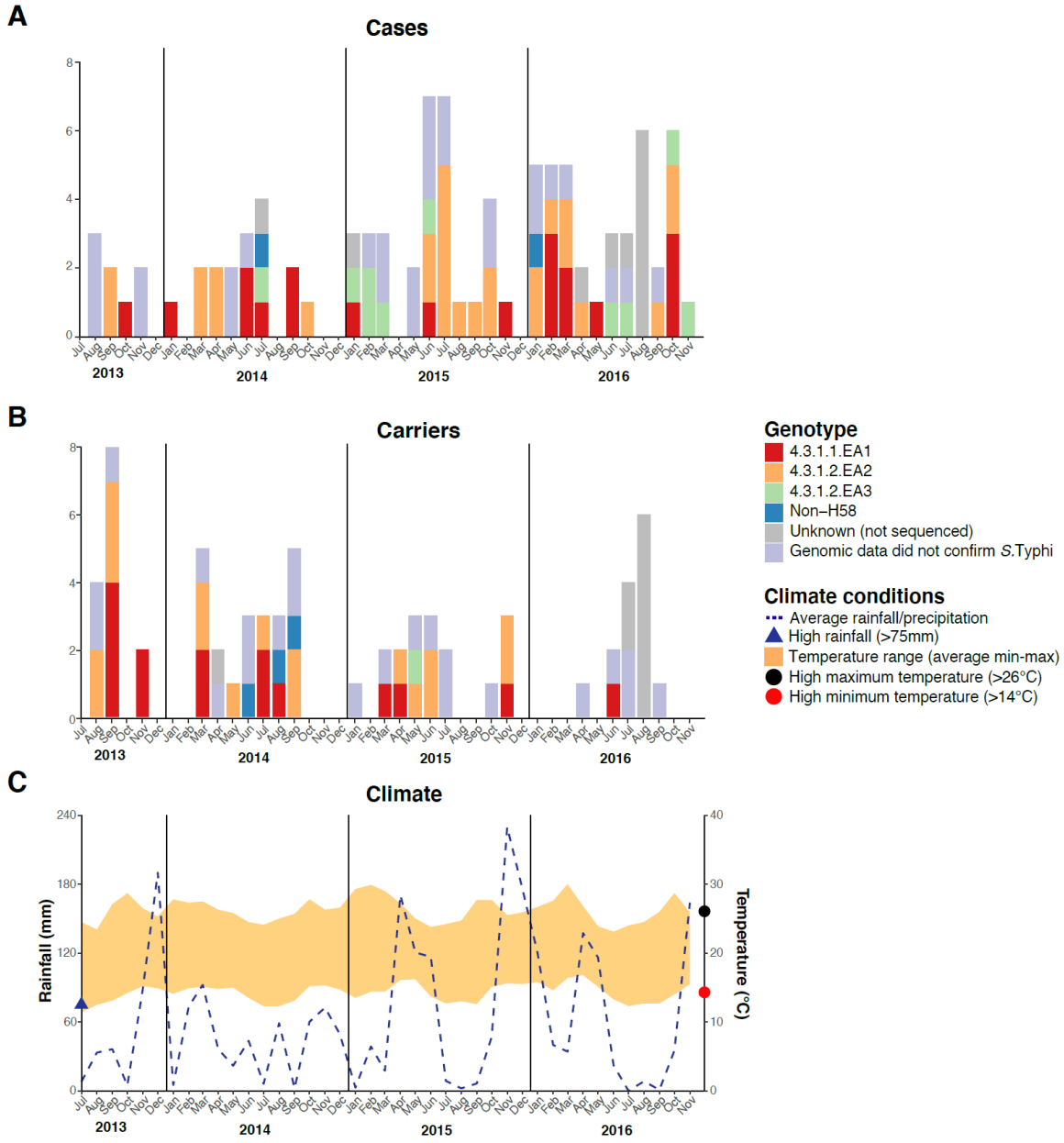


Figure 3. Epidemic curve of all *S.* Typhi cases and controls per month inside the DSS. (A) Monthly distribution of *S.* Typhi genotypes from cases. (B) Monthly distribution of *S.* Typhi genotypes from carriers. Note that the counts include all participants who were culture-positive for *S*. Typhi and also those who were culture-positive for other *Salmonella* but identified later by WGS as *S*. Typhi. (C) Weather conditions throughout the study period. Blue dashed line indicates precipitation level per month (rainfall), shaded orange polygon indicates the temperature range, red circle indicates threshold for high minimum temperature for statistical testing, black circle indicates threshold for high maximum temperature for statistical testing, blue triangle indicates threshold for high rainfall for statistical testing.

435 ***Evolutionary history of S.* Typhi *cases and controls***

436 We applied Bayesian phylodynamic analysis to all available Kenyan H58 genomes to
437 estimate the dates of emergence of each of the East African lineages. The data showed
438 temporal structure (see **methods** and **Fig. S3**), and we estimated a genome-wide
439 substitution rate of 0.8 SNPs per genome per year (95% HPD, 0.1-1.0). This translates
440 to a rate of $1.9x10^{-7}$ genome-wide substitutions per site per year (95% HPD = $1.5x10^{-7}$-
441 $2.2x10^{-7}$). The novel EA1 isolates from this study (accounting for 35% of cases and 37%
442 of controls) were intermingled with those sequenced previously from Kenya and were
443 genetically diverse (median pairwise distance ~16 SNPs, interquartile range 12-27).
444 This is indicative of a well-established EA1 *S*. Typhi population in Nairobi for which we
445 estimate the mrca existed circa 1990 (95% HPD, 1981-1999) (see **Fig. 4a**). The most
446 common lineage was EA2 (48%), which also showed extensive diversity and we
447 estimate emerged circa 1988-1990 (95% HPD, 1978-1997) (see **Fig. 4a**), earlier than
448 the first recorded H58 Lineage II isolation in Kenya in 2004 [14]. We estimate the MDR
449 fluoroquinolone non-susceptible lineage EA3, which accounts for just 11% of isolates,
450 arrived much more recently (Kenyan mrca circa 2012, 95% HPD 2009-2014) (see **Fig.**
451 **4a**). The topology of the global H58 tree (**Fig. 2**) supports South Asia as the most likely
452 origin for EA3, with EA3 strains spreading between Kenya and Uganda, probably
453 through the shared transport systems.
454
455 The Bayesian tree of Kenyan H58 isolates (**Fig. 4a)** shows intermingling of sequences
456 from acute cases and asymptomatic carriers. Sequences from carriers appeared more
457 deeply branched than those of cases (**Fig. 4a**), which we tested by comparing the

458 terminal branch lengths (estimated in units of time in the Bayesian phylogeny) and
459 isolate-specific SNP counts, for high-quality H58 sequences from acute cases (n=85) vs
460 those of asymptomatic carriers (n=43) (**Fig. 4b**). The mean values were higher for
461 carriers vs cases (**Fig. 4b-c**), with the trend being more pronounced among samples
462 from within the DSS (see **Fig. S4)**, but these trends were not statistically significant
463 (p=0.42 for unique SNPs and p=0.57 for terminal branches for all samples, using one-
464 sided Wilcoxon rank sum test; p=0.051 for unique SNPs and p=0.58 for terminal
465 branches in DSS). The mean number of non-synonymous (NS) mutations detected in
466 terminal branches was greater for carrier isolates than those from cases, but again this
467 difference was not statistically significant (0.72 vs 0.54 for all sequences, p=0.53 using
468 Wilcox rank sum test; 0.81 vs. 0.39, p=0.20 inside the DSS; see **Fig. 4d** and **Fig. S4**).
469 There was also no significant difference in terminal branch lengths or unique SNP
470 counts between genomes carrying vs lacking MDR genes or QRDR mutations (data not
471 shown).
472
473 Examination of the location of terminal-branch NS mutations revealed that certain
474 functional categories of genes carried more NS mutations arising on terminal branches
475 associated with carriage samples vs those from acute cases (**Fig. 5, Fig. S5,** and **Table**
476 **S8)**. Notably, carriage samples were associated with significantly higher frequencies of
477 terminal-branch NS mutations in genes responsible for the synthesis of surface
478 polysaccharides and antigens (9.3% of carriers vs 1.2% of acute cases, p=0.043,
479 Fisher's exact test). Notably, in the *viaB* operon (responsible for Vi capsule
480 biosynthesis) we identified n=2/43 carriage isolates that harboured NS mutations (*tviD*-

R159C and *tviE*-P263S) compared with only n=1/85 case isolate (*tviB*-V1M); and in the *wba* cluster genes (responsible for O-antigen biosynthesis) we identified n=2/43 carriage isolates that harboured NS mutations (*wza*-V137G and *wzxC*-L26F) whilst none were detected among case samples (**Table S8**). Non-significant excesses of mutations in carriage isolates were also observed for pathogenicity-island related functions and for periplasmic and exported lipoproteins (**Fig 5**).

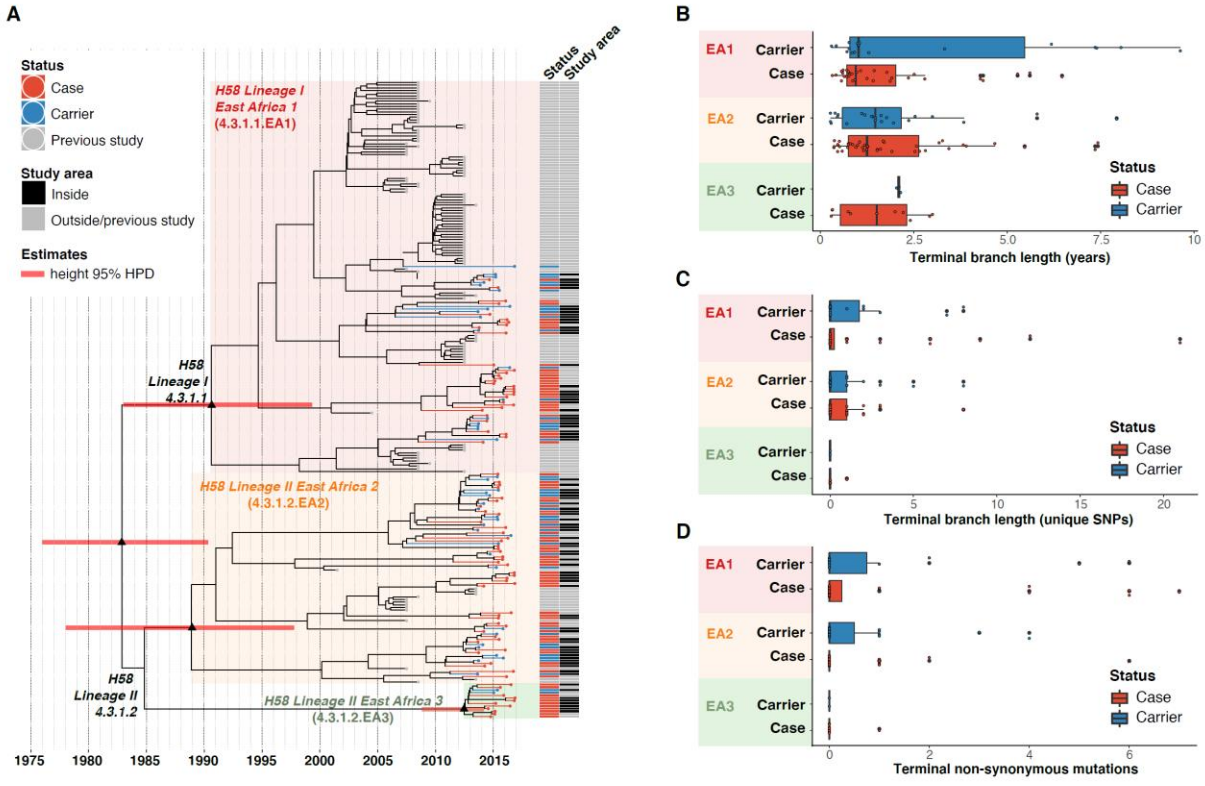


Fig. 4. Temporal distribution of genotypes and among cases and carriers. (A) Dated maximum-clade credibility phylogenetic tree of Kenyan *S.* Typhi genotype 4.3.1 (H58), including 128 isolated from this study. Tip colours & first colour bar indicate symptom status, second colour bar indicates those isolates from children living in the defined survey area. Black triangles demarcate nodes of interest, and the accompanying bars indicate 95% HPD of node heights. Interactive phylogeny available at https://microreact.org/project/I2KUoasUB. (B) Distribution of terminal branch lengths for all sequences, extracted from the Bayesian tree shown in (A). (C) Distribution of isolate-specific SNPs detected in sequences from all cases and controls. (D) Distribution of terminal non-synonymous mutations detected in sequences from all

cases and controls. In the boxplots in panels B, C, and D, black bars indicate median values, boxes indicate interquartile range. Cases and carrier samples indicated as per the inset legend.

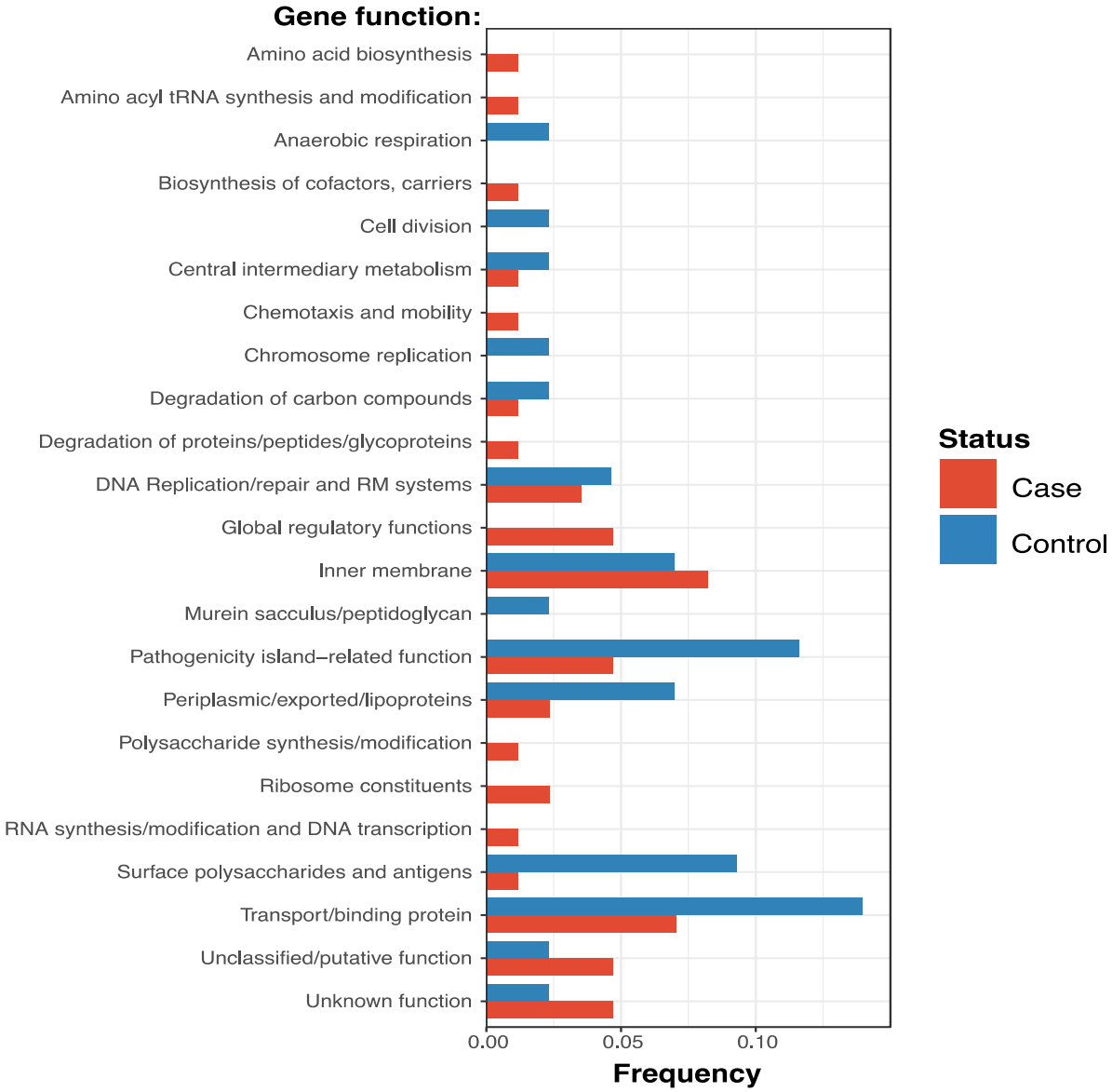


Fig. 5. Frequency of terminal non-synonymous mutations in difference gene functional categories among cases and carriers. Frequency of terminal non-synonymous mutations in all sequences. Red bars indicate the frequency non-synonymous mutations found in acute case samples, and blue bars indicate the frequency of mutations found in carrier samples.

Discussion

In this case-control typhoid surveillance study, we observed an asymptomatic *S.* Typhi carriage rate of 1.1% among children aged 16 years and under from an informal settlement with endemic Water, Sanitation, and Hygiene (WaSH) related enteric diseases [25,30,51]. To our knowledge, there has not been systematic surveillance for typhoid carriage in communities in Africa, but globally carriage and shedding of *S.* Typhi has mostly been associated with older age groups [6,52,53]. Our data highlights a role for paediatric carriage, revealing a lower percentage of carriers amongst infants ≤1 year of age (0.62%), increasing to 1.2% in children between 7 and 16 years (**Table S4**). Thus, carriage and shedding, especially among school age children, is likely an important factor in the onward transmission of typhoid in this setting [25]. Symptomatic typhoid fever is common in school age children, with a case culture positive rate of 4.3% among febrile children 7 to 16 years of age, though our data shows that there is also a substantial burden among younger children 1 to 7 years of age, and infants up to 1 year of age, with culture positive rates among febrile participants of 3.1% and 2.2%, respectively (**Table S4**).

We previously noted a dominance of MDR H58 *S.* Typhi over the last decade, essentially replacing the antimicrobial susceptible genotypes that dominated in the 1980-1990s [54]. The *S.* Typhi circulating in the informal settlement in the present study are largely comprised of descendants of the previously observed H58 sublineages 4.3.1.EA1 (36%) and 4.3.1.2.EA2 (48%) [12,15]. Both EA1 and EA2 appear to be long established genotypes, with the mrca of EA1 existing circa 1990 (95% HPD, 1981-1999)

535 (see **Fig. 4a**), consistent with the earliest recorded detection of H58 Lineage I in Kenya
536 in 1988 [14]. Similarly we predict that the mrca of EA2 existed circa 1988-1990 (95%
537 HPD, 1978-1997) (see **Fig. 4a**), earlier than the first recorded H58 Lineage II isolation in
538 Kenya in 2004 [14]. Our data thus support contemporaneous imports of EA1 and EA2 in
539 the late 1980s or early 1990s, shortly after the emergence of H58 in South Asia [12] (circa
540 1982, 95% HPD 1974-1990), and show that both lineages have persisted and
541 diversified locally alongside one another in the intervening decades. H58 sublineage
542 EA3 was introduced later (we estimate that the Kenyan mrca existed circa ~2012 (95%
543 HPD 2009 to 2014)), and consistent with this the lineage displays less diversity and
544 accounts for a smaller fraction of cases and controls (11%). The topology of our global
545 phylogeny (**Fig. 2**) suggests that South Asia is the most likely origin of EA3 (as it is for
546 EA1 and EA2), and that EA3 appears to have spread between Uganda and Kenya. This
547 in line with multiple reports [13,15] of H58 strains spreading through East Africa, mainly
548 arising from intracontinental and transcontinental travel and concomitant risk factors
549 associated with WASH conditions.
550
551 The sublineages of *S*. Typhi H58 in Kenya exhibit different antibiotic resistance profiles
552 (**Fig. 2, Table 2**). Notably, EA1 has a large Kenyan sublineage of MDR strains with
553 IncHI1 plasmids [12,15] (which are commonly associated with outbreaks in East Africa and
554 Asia [12,15,50]) but also a Kenyan sublineage with chromosomally integrated MDR.
555 Chromosomal integration of MDR has not previously been reported in *S*. Typhi from
556 Kenya (see **supplementary data**), but has been reported in Malawi and Tanzania [12]1
557 and our new data suggests that the variant may have been transferred to these

558 locations from Kenya (see **Figure 2**). MDR H58 isolates are now widespread across
559 East Africa, having been detected in Malawi, Uganda, Rwanda, Tanzania and
560 Mozambique [12,13,15,33,55].
561
562 The three East African lineages differed markedly in their patterns of mutations
563 conferring Decreased Ciprofloxacin Susceptibility (DCS). GyrB-S464F was conserved
564 among all EA2, whereas all EA3 isolates carried the GyrA-S83Y mutation. The GyrA-
565 S464F mutation was also detected at low frequency in EA1 (**Table 2**). This data
566 indicates that ciprofloxacin resistance has been selected independently multiple times
567 and is ongoing. Increasing rates of ciprofloxacin resistance have also been observed
568 following similar introductions of H58 elsewhere in East Africa [15], and likely reflect a
569 change in treatment practise following widespread dissemination of MDR *S.* Typhi
570 elsewhere including South and Southeast Asia [34,56,57].
571
572 The different *S*. Typhi lineages appeared to be fairly evenly distributed between both
573 acute cases and carriers, with the most common subgroup (EA2) accounting for 46% of
574 acute cases and 50% of carriers. Similarly, all *S.* Typhi genotypes were identified
575 throughout the study period and spatially across the study site, with most case/carrier
576 monthly counts and geographic regions containing a diversity of genotypes (**Fig. 3 and**
577 **Fig. S2**). Our data therefore provides no evidence for major point-source single-
578 genotype outbreaks, but is consistent with persistent contamination of water supplies
579 with multiple *S.* Typhi genotypes. Higher temperatures were associated with lower *S.*
580 Typhi case and carrier counts, however, no association with high rainfall was observed.

These findings are in line with previous studies focusing on seasonal trends in nearby Kibera [23]. However they contrast with trends previously observed in other settings including Malawi, where higher temperatures and rainfall were associated with increased risk of disease albeit with a time lag of multiple months [58], and South Asia [59-61].

In our phylogenetic trees, branch lengths and SNP counts are measures of evolutionary time, and their terminal or isolate-specific quantities represent an upper bound on the time since acquisition of the infection from which each isolate was sampled. Hence, our data suggest that the *S*. Typhi isolated from asymptomatic controls may have had, on average, a longer duration from acquisition of the infection to sampling in the clinic. This supports the interpretation that *S*. Typhi-positive controls identified in this study represent genuine medium- to long-term typhoid carriers, rather than simply reflecting transient presence in the gut. The greater diversity observed here amongst controls (**Table 2**) further supports this interpretation. Longer branch lengths among carrier samples were also observed in a recent study [8] of *S.* Typhi isolated from bile samples from the gallbladders of cholecystectomy patients in Nepal. The differences in terminal branch lengths were non-significant in our study, possibly reflecting low statistical power or, perhaps less likely, that our control data constitute a mix of multiple carriage types including convalescent (three weeks to three months), temporary (three to twelve months), and chronic (more than one year) carriers. Also in line with previous findings [8], our analyses provide evidence of positive selection among carriage isolates, with a higher proportion of non-synonymous mutations detected among carriers in specific

biological pathways including surface polysaccharides and antigens, transport/binding proteins, pathogenicity, and anaerobic respiration (**Fig. 5, Fig S5, Table S8**). This is exemplified in the genes encoding surface antigens, notably those responsible for biosynthesis of Vi capsule and O-antigen lipopolysaccharide.

Our study is not without limitations, firstly, our data are from a single informal settlement community in Nairobi, and thus may not be representative of the overall population structure and AMR patterns of typhoid in Kenya more broadly, or in older age groups. Similarly, our sample size yielded a relatively small number of isolates for WGS, and we thus lack statistical power for some genetic analyses.

Conclusion

Our study is the first case-control study to identify and sequence both typhoid carriers and cases contemporaneously in an endemic community setting. High rates of AMR among both infection types in Kenya combined with high carriage and case rates, especially in the younger age groups, highlight the need for enhanced AMR and genomic surveillance in this region to inform both treatment guidelines and control strategies that keep pace with the local evolution and spread of AMR. Intervention strategies are urgently needed including the introduction of the new Vi conjugate vaccine in a programme that includes targeting of paediatric age groups in the short term, and improvements to WaSH infrastructure in the long term.

627 **Funding**

628 Research reported in this publication was supported by the National Institute of Allergy
629 and Infectious Diseases of the National Institutes of Health under Award Number
630 R01AI099525 to S.K. and the Wellcome Trust. ZAD was supported by a grant funded by
631 the Wellcome Trust (STRATAA; 106158/Z/14/Z and TyVac), and received funding from
632 the European Union's Horizon 2020 research and innovation programme under the
633 Marie Skłodowska-Curie grant agreement TyphiNET No 845681. GD was supported by
634 the Cambridge Biomedical Research Council NIHR AMR theme. KEH was supported by
635 a Senior Medical Research Fellowship from the Viertel Foundation of Australia, and the
636 Bill and Melinda Gates Foundation, Seattle (grant #OPP1175797).

637

638 **Author contributions**

639 SK, GD, JC and MA designed the project. ZAD, SD, KEH performed bioinformatic
640 analysis of the *S.* Typhi genomes. CM, SA, NG and BO performed field work and
641 patient recruitment and sampling, RN, CW, SRO and SMK performed laboratory sample
642 processing and microbiological data analysis. SK, ZAD, KEH and GD wrote the
643 manuscript. All authors contributed to the manuscript editing.

644

645 **Competing interests**

646 The authors declare no competing interests.

647

648 **Additional information**

649 Correspondence and requests for materials should be addressed to SK.

References

1 Crump JA, Mintz ED. Global trends in typhoid and paratyphoid Fever. *Clin Infect Dis* 2010; **50**: 241–6.

2 Mogasale V, Maskery B, Ochiai RL, *et al.* Burden of typhoid fever in low-income and middle-income countries: a systematic, literature-based update with risk-factor adjustment. *Lancet Glob Health* 2014; **2**: e570–80.

3 Hopewell MR, Graham JP. Trends in access to water supply and sanitation in 31 major sub-Saharan African cities: an analysis of DHS data from 2000 to 2012. *BMC Public Health* 2014; **14**: 208–12.

4 Marchello CS, Hong CY, Crump JA. Global Typhoid Fever Incidence: A Systematic Review and Meta-analysis. *Clin Infect Dis* 2019; **68**: S105–16.

5 Marchello CS, Birkhold M, Crump JA. Complications and mortality of typhoid fever: A global systematic review and meta-analysis. *J Infect* 2020; **81**: 902–10.

6 Levine MM, Black RE, Lanata C. Precise estimation of the numbers of chronic carriers of Salmonella typhi in Santiago, Chile, an endemic area. *J Infect Dis* 1982; **146**: 724–6.

7 Parry CM, Hien TT, Dougan G, White NJ, Farrar JJ. Typhoid fever. *N Engl J Med* 2002; **347**: 1770–82.

8 Thanh DP, Thieu NTV, Nguyen Thi Nguyen T, *et al.* Gallbladder carriage generates genetic variation and genome degradation in Salmonella Typhi. *PLoS Pathog* 2020; **16**: e1008998.

9 Gauld JS, Hu H, Klein DJ, Levine MM. Typhoid fever in Santiago, Chile: Insights from a mathematical model utilizing venerable archived data from a successful disease control program. *PLoS Negl Trop Dis* 2018; **12**: e0006759.

10 Pitzer VE, Feasey NA, Msefula C, *et al.* Mathematical Modeling to Assess the Drivers of the Recent Emergence of Typhoid Fever in Blantyre, Malawi. *Clin Infect Dis* 2015; **61 Suppl 4**: S251–8.

11 Saad NJ, Bowles CC, Grenfell BT, *et al.* The impact of migration and antimicrobial resistance on the transmission dynamics of typhoid fever in Kathmandu, Nepal: A mathematical modelling study. *PLoS Negl Trop Dis* 2017; **11**: e0005547.

12 Wong VK, Baker S, Pickard DJ, *et al.* Phylogeographical analysis of the dominant multidrug-resistant H58 clade of Salmonella Typhi identifies inter- and intracontinental transmission events. *Nat Genet* 2015; **47**: 632–9.

13 Feasey NA, Gaskell K, Wong V, *et al.* Rapid emergence of multidrug resistant, H58-lineage Salmonella typhi in Blantyre, Malawi. *PLoS Negl Trop Dis* 2015; **9**: e0003748.

14 Kariuki S, Revathi G, Kiiru J, *et al.* Typhoid in Kenya is associated with a dominant multidrug-resistant Salmonella enterica serovar Typhi haplotype that is also widespread in Southeast Asia. *J Clin Microbiol* 2010; **48**: 2171–6.

15 Park SE, Pham DT, Boinett C, *et al.* The phylogeography and incidence of multi-drug resistant typhoid fever in sub-Saharan Africa. *Nat Commun* 2018; **9**: 5094.

16 Klemm EJ, Shakoor S, Page AJ, *et al.* Emergence of an Extensively Drug-Resistant Salmonella entericaSerovar Typhi Clone Harboring a Promiscuous Plasmid Encoding Resistance to Fluoroquinolones and Third-Generation Cephalosporins. *mBio* 2018; **9**: 346–344.

17 Hendriksen RS, Leekitcharoenphon P, Mikoleit M, *et al.* Genomic dissection of travel-associated extended-spectrum-beta-lactamase-producing Salmonella enterica serovar typhi isolates originating from the Philippines: a one-off occurrence or a threat to effective treatment of typhoid fever? *J Clin Microbiol* 2015; **53**: 677–80.

18 Lutterloh E, Likaka A, Sejvar J, *et al.* Multidrug-resistant typhoid fever with neurologic findings on the Malawi-Mozambique border. *Clin Infect Dis* 2012; **54**: 1100–6.

19 Marks F, Kalckreuth von V, Aaby P, *et al.* Incidence of invasive salmonella disease in sub-Saharan Africa: a multicentre population-based surveillance study. *Lancet Glob Health* 2017; **5**: e310–23.

20 Neil KP, Sodha SV, Lukwago L, *et al.* A large outbreak of typhoid fever associated with a high rate of intestinal perforation in Kasese District, Uganda, 2008-2009. *Clin Infect Dis* 2012; **54**: 1091–9.

21 Kyobutungi C, Ziraba AK, Ezeh A, Yé Y. The burden of disease profile of residents of Nairobi's slums: results from a demographic surveillance system. *Popul Health Metr* 2008; **6**: 1.

22 Mberu BU, Haregu TN, Kyobutungi C, Ezeh AC. Health and health-related indicators in slum, rural, and urban communities: a comparative analysis. *Glob Health Action* 2016; **9**: 33163.

23 Breiman RF, Cosmas L, Njuguna H, *et al.* Population-based incidence of typhoid fever in an urban informal settlement and a rural area in Kenya: implications for typhoid vaccine use in Africa. *PLoS ONE* 2012; **7**: e29119.

718 24 Mutai WC, Muigai AWT, Waiyaki P, Kariuki S. Multi-drug resistant Salmonella
719 enterica serovar Typhi isolates with reduced susceptibility to ciprofloxacin in Kenya.
720 *BMC Microbiol* 2018; **18**: 187–5.

721 25 Mbae C, Mwangi M, Gitau N, *et al.* Factors associated with occurrence of
722 salmonellosis among children living in Mukuru slum, an urban informal settlement
723 in Kenya. *BMC Infect Dis* 2020; **20**: 422.

724 26 KNBS KNBOS. Kenya population and housing census 2009. 2009.

725 27 Olack B, Burke H, Cosmas L, *et al.* Nutritional status of under-five children living in
726 an informal urban settlement in Nairobi, Kenya. *J Health Popul Nutr* 2011; **29**: 357–
727 63.

728 28 Mutisya E, Yarime M. Understanding the Grassroots Dynamics of Slums in Nairobi:
729 The Dilemma of Kibera Informal Settlements. *Int Trans J Eng Manag Appl Sci*
730 *Technol* 2011; **2**: 1–18.

731 29 Tinline RR, Gregory D. The Universal Transverse Mercator Code: A location code
732 for disease reporting. *Can Vet J* 1988; **29**: 825–9.

733 30 Kariuki S, Mbae C, Van Puyvelde S, *et al.* High relatedness of invasive multi-drug
734 resistant non-typhoidal Salmonella genotypes among patients and asymptomatic
735 carriers in endemic informal settlements in Kenya. *PLoS Negl Trop Dis* 2020; **14**:
736 e0008440.

737 31 Patel JB, Cockerill FR, Bradford PA. Performance standards for antimicrobial
738 susceptibility testing: twenty-fifth informational supplement. 2015.

739 32 Parkhill J, Dougan G, James KD, *et al.* Complete genome sequence of a multiple
740 drug resistant Salmonella enterica serovar Typhi CT18. *Nature* 2001; **413**: 848–52.

741 33 Wong VK, Baker S, Connor TR, *et al.* An extended genotyping framework for
742 Salmonella enterica serovar Typhi, the cause of human typhoid. *Nat Commun*
743 2016; **7**: 12827.

744 34 Britto CD, Dyson ZA, Duchene S, *et al.* Laboratory and molecular surveillance of
745 paediatric typhoidal Salmonella in Nepal: Antimicrobial resistance and implications
746 for vaccine policy. *PLoS Negl Trop Dis* 2018; **12**: e0006408.

747 35 Pham Thanh D, Karkey A, Dongol S, *et al.* A novel ciprofloxacin-resistant subclade
748 of H58 Salmonella Typhi is associated with fluoroquinolone treatment failure. *Elife*
749 2016; **5**: e14003.

750 36 Croucher NJ, Page AJ, Connor TR, *et al.* Rapid phylogenetic analysis of large
751 samples of recombinant bacterial whole genome sequences using Gubbins.
752 *Nucleic Acids Res* 2015; **43**: e15–5.

753 37 Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis
754 of large phylogenies. *Bioinformatics* 2014; **30**: 1312–3.

755 38 Argimon S, Abudahab K, Goater RJE, *et al.* Microreact: visualizing and sharing
756 data for genomic epidemiology and phylogeography. *Microb Genom* 2016; **2**:
757 e000093.

758 39 Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in
759 R language. *Bioinformatics* 2004; **20**: 289–90.

760 40 Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. ggtree: an rpackage for visualization
761 and annotation of phylogenetic trees with their covariates and other associated
762 data. *Methods Ecol Evol* 2016; **8**: 28–36.

763 41 Edwards DJ, Duchêne S, Pope B, Holt KE. SNPPar: identifying convergent
764 evolution and other homoplasies from microbial whole-genome alignments. *bioRxiv*
765 DOI:10.1101/2020.07.08.194480.

766 42 Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure
767 of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol*
768 2016; **2**: vew007.

769 43 Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian
770 phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol*
771 2018; **4**: vey016.

772 44 Inouye M, Dashnow H, Raven L-A, *et al.* SRST2: Rapid genomic surveillance for
773 public health and hospital microbiology labs. *Genome Med* 2014; **6**: 90.

774 45 Carattoli A, Zankari E, García-Fernández A, *et al.* In silico detection and typing of
775 plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob*
776 *Agents Chemother* 2014; **58**: 3895–903.

777 46 Gupta SK, Padmanabhan BR, Diene SM, *et al.* ARG-ANNOT, a new bioinformatic
778 tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob Agents*
779 *Chemother* 2014; **58**: 212–20.

780 47 Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome
781 assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017; **13**:
782 e1005595.

783 48 Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de
784 novo genome assemblies. *Bioinformatics* 2015; **31**: 3350–2.

785 49 Hawkey J, Hamidian M, Wick RR, *et al.* ISMapper: identifying transposase insertion
786 sites in bacterial genomes from short read sequence data. *BMC Genomics* 2015;
787 **16**: 667–11.

50 Holt KE, Phan M-D, Baker S, *et al.* Emergence of a globally dominant IncHI1 plasmid type associated with multiple drug resistant typhoid. *PLoS Negl Trop Dis* 2011; **5**: e1245.

51 Kariuki S, Mbae C, Onsare R, *et al.* Multidrug-resistant Nontyphoidal Salmonella Hotspots as Targets for Vaccine Use in Management of Infections in Endemic Settings. *Clin Infect Dis* 2019; **68**: S10–5.

52 Senthilkumar B, Senbagam D, Rajasekarapandian M. An epidemiological surveillance of asymptomatic typhoid carriers associated in respect to socioeconomic status in India. *J Public Health* 2012; **22**: 297–301.

53 Gunn JS, Marshall JM, Baker S, Dongol S, Charles RC, Ryan ET. Salmonella chronic carriage: epidemiology, diagnosis, and gallbladder persistence. *Trends Microbiol* 2014; **22**: 648–55.

54 Kariuki S, Gordon MA, Feasey N, Parry CM. Antimicrobial resistance and management of invasive Salmonella disease. *Vaccine* 2015; **33 Suppl 3**: C21–9.

55 Ingle DJ, Nair S, Hartman H, *et al.* Informal genomic surveillance of regional distribution of Salmonella Typhi genotypes and antimicrobial resistance via returning travellers. *PLoS Negl Trop Dis* 2019; **13**: e0007620.

56 Britto CD, Dyson ZA, Mathias S, *et al.* Persistent circulation of a fluoroquinolone-resistant Salmonella enterica Typhi clone in the Indian subcontinent. *J Antimicrob Chemother* 2020; **75**: 337–41.

57 Rahman SIA, Dyson ZA, Klemm EJ, *et al.* Population structure and antimicrobial resistance patterns of Salmonella Typhi isolates in urban Dhaka, Bangladesh from 2004 to 2016. *PLoS Negl Trop Dis* 2020; **14**: e0008036.

58 Thindwa D, Chipeta MG, Henrion MYR, Gordon MA. Distinct climate influences on the risk of typhoid compared to invasive non-typhoid Salmonella disease in Blantyre, Malawi. *Sci Rep* 2019; **9**: 20310–1.

59 Baker S, Holt KE, Clements ACA, *et al.* Combined high-resolution genotyping and geospatial analysis reveals modes of endemic urban typhoid fever transmission. *Open Biol* 2011; **1**: 110008.

60 Dewan AM, Corner R, Hashizume M, Ongee ET. Typhoid Fever and its association with environmental factors in the Dhaka Metropolitan Area of Bangladesh: a spatial and time-series approach. *PLoS Negl Trop Dis* 2013; **7**: e1998.

61 Kanungo S, Dutta S, Sur D. Epidemiology of typhoid and paratyphoid fever in India. *J Infect Dev Ctries* 2008; **2**: 454–60.

SUPPLEMENTARY MATERIAL

Supplementary methods

Phylogenetic and SNP analysis of S. Typhi *isolates*

For SNP analysis, paired-end reads from 153 *S.* Typhi isolates were mapped to the reference sequence of *S.* Typhi CT18 (accession number: AL513382) [1] using the RedDog mapping pipeline (v1beta.10.3), available at http://githib.com/katholt/reddog. Briefly, RedDog uses Bowtie (v2.2.3) [2] to map reads to the reference sequence; SAMtools (v0.1.19) [3] to identify SNPs with phred quality scores above 30; filter out SNPs supported by <5 reads, or with >2.5 times the genome-wide average read depth (representing putative repeated sequences), or with ambiguous (heterozygous) consensus base calls. For each SNP position that passed these criteria in any one isolate, consensus base calls (i.e. alleles) for that position were extracted from all genomes, and used to construct an alignment of alleles across all SNP sites. Ambiguous base calls and those with phred quality <20 were treated as unknown alleles and represented with a gap character in the SNP alignment. Read alignments were used to assign isolates to previously defined lineages according to the *S.* Typhi extended genotyping framework [4-6], by subjecting the alignments (BAM format) to analysis with the GenoTyphi pipeline (available at http://github.com/katholt/genotyphi). Unique SNPs defining three novel lineages were identified from the genome-wide SNP allele table (with SNPs responsible for non-synonymous mutations in highly conserved genes without deletions prioritized for lineage definitions), these were added to the GenoTyphi scheme to facilitate easy identification of these lineages in future studies.

Phylogenetic analyses were restricted to WGS-confirmed pure cultures of *S*. Typhi H58 (genotype 4.3.1, n=128). For some analyses, an additional 1,076 *S*. Typhi H58 genomes from previously published WGS studies of global and African isolates [4,7-9] were also included for context (using the same mapping approach detailed above). Alleles from 61 additional *S*. Typhi genomes representing all non-H58 subclades (listed in **Table S2**), and *S*. Paratyphi A str. AKU_12601 (accession FM200053) [10], were also included in the phylogenetic analysis as outgroups for tree rooting. SNPs called in phage regions or repetitive sequences (354 kbp; ~7.4% of bases in the CT18 reference chromosome, as defined previously [7,11,12] were filtered from the alignment, which was then used with the CT18 reference genome (AL513382) to produce a whole genome pseudoalignment that was subjected to analysis with Gubbins (v2.3.2) [13] to remove any further recombinant regions. This resulted in a final set of 8,635 SNPs identified from an alignment of 4,275,037 sites for the 1,266 isolates.

From the global SNP alignment, maximum likelihood (ML) phylogenetic trees were inferred using RAxML (v8.2.9) [14], with a generalized time-reversible model, a Gamma distribution to model site-specific rate variation (the GTR+ Γ substitution model; GTRGAMMA in RAxML), and 100 bootstrap pseudo-replicates to assess branch support.

From the global alignment we extracted a separate SNP alignment for the set of 239 Kenyan *S*. Typhi H58 genomes (n=128 from this study and n=111 from published studies, see **Table S3** [4,9], which had length 489 SNPs. The same phylogenetic

869 inference methods were used to generate a ML tree from the SNP alignment of 239
870 Kenyan isolates for temporal analyses (described below).
871
872 Phylogenetic trees were visualized using Microreact (interactive global H58 phylogeny
873 available at: https://microreact.org/project/wViqmaRdZuFVEb6yk4i1jU) [15] and the R
874 package *ggtree* v1.14.6 [16]. For the purpose of plotting the global H58 tree, clusters of *S.*
875 Typhi isolates that were members of the same monophyletic clade and isolated from the
876 same location in South East Asian countries (i.e. representing local outbreaks) were
877 reduced to a single representative each using the *drop.tip()* function in the R package
878 *ape* [17].
879
880 Terminal branch lengths were extracted from phylogenies using R package *ggtree* [16].
881 Pairwise SNP distances were calculated from alignments using the *dist.dna()* function in
882 the R package *ape* v5.3 [17]. Non-synonymous mutations were detected using SNPPar
883 (V0.4.2dev) [18] and grouped by function based on the gene in which they were found
884 according to the *S.* Typhi functional classification scheme developed at the Sanger
885 Institute (www.sanger.ac.uk) using the genome annotation of CT18 [1,19].
886
887 ***Phylodynamic analysis***
888 To investigate temporal signal and date the introduction of *S*. Typhi H58 into Kenya
889 based on the 239 available genomes (n=128 from this study, and n=111 from previous
890 studies [4,9], we used several methods. First, we used TempEst (v1.5.1) [20] to assess
891 temporal structure (i.e. clock-like evolution) by conducting a regression analysis of the

root-to-tip branch distances of the ML tree as a function of sampling date (expressed as decimal years at a resolution of days), using the heuristic residual mean squared method with the best fitting root selected. To estimate divergence dates for the three *S.* Typhi H58 sublineages we detected in Kenya (EA1-3), we used BEAST (v1.10) [21] to fit a phylodynamic model to the SNP alignment and isolation dates (as decimal years at a resolution of days). Note that as EA1-3 almost exclusively comprise Kenyan strains in the global H58 tree and the parent node for each of EA1-3 is the parent node for all Kenyan isolates of EA1-3 (see **Fig. 2**), the divergence dates for the EA1-3 parent nodes in the tree of 239 Kenyan isolates is taken as the divergence date for each of EA1-3 generally, as well as the lower bound for the date of introduction of each of these sublineages into Kenya.

We ran separate models using constant-coalescent population size and Bayesian skyline tree priors, in combination with a strict clock model or a relaxed (uncorrelated log normal distribution) clock model, to identify the best fitting model for our data. For BEAST analyses the GTR+Γ substitution model was selected, and sampling times (tip dates) were used to calibrate the molecular clock (for isolates collected in this study the precise day of isolation was used; for the previously published genomes, only the isolation year was known, so tip dates were assigned to the first of July for that year with an uncertainty of 0.5 years). For all tree prior and model combinations, a chain length of 100,000,000 steps with sampling every 5,000 steps was used [22]. The relaxed (uncorrelated lognormal) clock model, that allows for evolutionary rate variations among branches of the tree and the constant-coalescent model were found to best fit our data.

To assess the temporal signal of these Bayesian estimates, we conducted a date-randomisation test where sampling times were assigned randomly to the sequences, and the analysis re-run 20 times with the best fitting models (constant-coalescent demographic and uncorrelated lognormal clock) [22,23]. The date-randomisation test revealed that these data displayed 'strong' temporal structure (meeting the criterion CR2 of Duchene et al 2016). Our preliminary BEAST runs resulted in implausible tree topologies and dated the most recent common ancestor (mrca) of H58 *S.* Typhi in Kenya ~1927 (95% highest posterior density = 1847-1984), conflicting with previously inferred divergence dates for the emergence of H58 (5,7,26). Fixing the tree topology to that obtained from ML inference yielded more plausible date estimates (~1968, 95% HPD = 1957-1977), however, sampling from the prior (without the sequence alignment) using the same model showed that this estimate was driven entirely by the priors provided to the model, with no information contributed by the sequence data. Taken together, these preliminary analyses suggested that while a temporal signal is present in the alignment, the signal is weak and would benefit from the specification of sensible priors to calibrate the root height. Previous analyses of the H58 divergence date, inferred using global data spanning a wider sampling period and with stronger temporal signal, estimate it emerged circa 1989 (95% HPD, 1981-1995) [7], and we have previously reported the presence of H58 lineage 1 in Kenya in the late 1980s [24]. We therefore specified a log-normally distributed root height prior with mean 1989 and standard deviation 4 years. Use of this root height prior (without fixing the tree topology) yielded a plausible tree topology (i.e. consistent with the outgroup-rooted maximum likelihood tree inferred from the same alignment, with the expected separation of H58

sublineages into monophyletic clades), showed evidence of temporal signal (via date randomisation testing), and was not driven by priors alone; hence this approach was used for the final analyses presented here.

For the final analyses, 2 independent runs each conducted with a chain length of 100,000,000 steps sampling every 5,000 iterations were combined using LogCombiner (v1.10.0) [21], after removing the first 10% of steps from each as 'burn-in'. Maximum-clade credibility trees (MCC) trees were generated with 'common ancestor heights' specified for node heights, using TreeAnnotator (v1.10.0) [21]. The effective sample sizes (ESS) from the combined runs were >200 for all reported parameters. The resultant MCC tree was visualized using *ggtree* (v1.14.6) [16] and Microreact [15] (interactive phylogeny available at: https://microreact.org/project/I2KUoasUB).

Genomic determinants of antimicrobial resistance

The read mapping-based allele typer SRST2 (v0.2.0) [25] was used to detect the presence of plasmid replicons (PlasmidFinder database [26]) and antimicrobial resistance (AMR genes) (ARGannot database [27]) and to identify the precise alleles of AMR genes. Where AMR genes were observed without evidence of a known AMR plasmid, raw read data was *de novo* assembled using Unicycler (v0.4.7) [28] and then examined visually using the Bandage (v0.8.1) [29] assembly graph visualizer, in order to interrogate the assembly to confirm the chromosomal location and composition of AMR-associated transposons. ISMapper (v2.0) [30] was used with default parameters to screen all read sets for insertion sites of transposases of IS*1* (accession number J01730) relative to the

CT18 reference chromosome sequence, in order to identify the location of any such insertion sequences in the chromosome of each Kenyan *S.* Typhi genome. Single point mutations located within the quinolone resistance determining region (QRDR) of genes *gyrA, gyrB,* and *parC* associated with reduced susceptibility to fluoroquinolones [8] were detected from the whole genome read alignments (BAM files) described above, using GenoTyphi [4-6]. Where a resistance phenotype was detected in the absence of known molecular determinants of AMR, DBGWAS (v0.5.4) [31 31] was utilised to carry out a bacterial genome-wide association study (GWAS) to screen for genetic loci and/or variants associated with the observed resistance phenotype.

Statistical and spatial analysis

All statistical analyses unless otherwise stated were carried out using R (v4.0.2). Multivariate logistic regression analyses were conducted with the *glm()* function in base R, and linear regression analyses of were carried out using the *lm()* function in base R. Shannon diversity was calculated using the *diversity()* function in the R package *vegan* [32]. Base R function *chisq.test()* was used to conduct a Chi-Squared test of age groups and *S.* Typhi genotypes. Fisher's exact test was carried out using the *fisher.test()* function in base R to investigate the frequency of non-synonymous mutations among cases and carriers, as well as associations between binary categories ('high' vs 'low') for monthly case counts or carrier counts and weather conditions (monthly rainfall, minimum and maximum temperature) in either the same month, the previous month, or two months prior. Thresholds for the 'high' categories were: case or carrier counts >2; rainfall >75 mm; minimum temperature >14°C; maximum temperature >26°C.

GPS coordinate data were visualized using Microreact [15] .

Supplementary results

Precision of AMR phenotypes from genotypes

Observed AMR phenotypes for all high quality WGS data (n=128 H58 and n=8 Non-H58 genome sequences) were largely explained by the presence of known molecular determinants of AMR (**Table S5**). Sensitivity was generally high for previous first line drugs for treating typhoid including chloramphenicol, ampicillin, co-trimoxazole, as well as tetracyclines (>94%), as was specificity (>74%), and low very major errors (VME; failure to predict a resistance phenotype) were determined for these drugs (<4%). Major errors (ME; failure to predict a susceptible resistance phenotype) ranged between 5.1-7.4% for these drugs, and for three sequences, this is potentially explained by the loss of an IncHI1 plasmid (n=2) or Tn*2670*-like transposon (n=1) associated with the MDR phenotype in culture, as resistance phenotypes were not observed for any of these drugs.

VME for nalidixic acid (32.4%), and ME (23.5%) for ciprofloxacin were both high. The sensitivity and specificity for these drugs (**Table S5**) are confounded by the presence of the GyrB-S464F mutation (**Table 2**), conserved in all n=73 genome sequences of genotype 4.3.1.2.EA2 (H58 lineage 2 sublineage East Africa 2; discussed below in detail). The GyrB-S464F mutation has been previously reported to cause Decreased Ciprofloxacin Susceptibility (DCS), but not nalidixic acid resistance [33], however, in our data the relationship between GyrB-S464F Quinolone/Fluoroquinolone resistance in this strain background remains unclear as 50.7% of EA2 isolates showed reduced susceptibility to nalidixic acid and 47.9% ciprofloxacin, with the rest testing sensitive to

these drugs. Neither comparative analysis of resistant and sensitive EA2 sequences, nor further bacterial GWAS analysis, revealed any evidence of either causal or compensatory mutations, or the acquisition/loss of genetic loci responsible for these phenotypes, perhaps due to low power.

Specificity estimates were high (100%), but sensitivity low (0%), for third generation cephalosporins including ceftazidime cefotaxime ceftriaxone and cefpodoxime (see **Table S5**). This was the result of phenotypic resistance to 3rd generation cephalosporins detected at low frequency among our high quality sequenced genomes (n=1-4, 0.73-2.9%) in the absence of any known molecular determinants for resistance to these drugs. A GWAS was again carried out using susceptibility data for these drugs, but no candidate molecular mechanisms significantly associated with the observed resistance phenotypes were identified, again possibly due to low power.

Local subpopulations of *S*. Typhi H58

Salmonella Typhi H58 (genotype 4.3.1) is subdivided into lineages I (genotype 4.3.1.1) and II (genotype 4.3.1.2). Lineage II was more common in this setting than lineage I: n=90 (62.1% of H58) vs n=55 (37.9%). Examination of the global phylogeny (**Fig. 2**, and online interactive version https://microreact.org/project/wViqmaRdZuFVEb6yk4i1jU) revealed all H58 lineage I isolates from this study shared a most recent common ancestor (mrca) whose descendants form a monophyletic clade that exclusively comprised *S*. Typhi from East African countries (see **Fig. 2**). This clade corresponds to the previously reported introduction of H58 lineage I from South Asia into Eastern

Africa, which appears to have arrived in Kenya before spreading to Tanzania and on to Malawi and South Africa [4,34] (**Fig. 2**). Here, we define this as H58 sublineage EA1 (East Africa 1) with genotype designation 4.3.1.1.EA1 (labelled in **Fig. 2**), which can be identified by the presence of a synonymous marker SNP STY0750-G1407A (position 751854 in CT18; this genotype has been added to the GenoTyphi scheme available at http://github.com/katholt/genotyphi).

S. Typhi H58 lineage II (genotype 4.3.1.2) isolates from our study belonged to two distinct clades of the global phylogeny (**Fig. 2**), which were each exclusively populated by East African isolates. The largest of these clades (n=80 isolates, of which 81.3% derive from the current study) formed a monophyletic group in which isolates from this study were intermingled with those obtained from previous studies in Kenya (n=14, 17.5%), and a single Tanzanian strain isolated in 2012 (see **Fig. 2**), suggestive of a single inter-country transmission event from Kenya into Tanzania. This clade is nested within a deeper clade of diverse South Asian isolates (see **Fig. 2**), and corresponds to the previously reported introduction of H58 lineage II into Kenya from South Asia [7,9]. This lineage, here defined as H58 sublineage EA2 (East Africa 2) and designated genotype 4.3.1.2.EA2 (labelled in **Fig. 2**), can be identified by the presence of a synonymous marker SNP STY4818-C1069T (position 4680610 in CT18). The smaller East African H58 lineage II clade (n=43 isolates) comprised two sister clades, separated by ≥13 SNPs, one involving isolates from Kenya (n=13, all from this study) and the other isolates from Uganda (n=30, which accounted for 100% of the typhoid burden at the Ugandan site in the TSAP study [9]) (see **Fig. 2**). This clade (including both the Ugandan

and Kenyan subgroups) is here defined as H58 sublineage EA3 (East Africa 3) and designated genotype 4.3.1.2.EA3 (labelled in **Fig. 2**), identified by synonymous marker SNP STY2750-G96A (position 2587488 in CT18). Both EA2 and EA3 genotypes have been added to the GenoTyphi scheme.

The three East African H58 subgroups circulating in our setting all had high rates of MDR (84%, 74% and 94%, respectively); however in EA2, MDR was exclusively associated with the PST6-IncHI1 plasmid, and in EA3 exclusively with the chromosomal insertion (see **Table 2** and **Fig. 2**). In EA1, most MDR was associated with the PST6-IncHI1 plasmid. However, a subclade of isolates (associated with spread to Tanzania and Malawi) carried the chromosomal insertion instead (see **Table 2** and **Fig. 2**). Chromosomal integration of MDR strains has not previously been reported in *S*. Typhi from Kenya, although it has been reported in the Malawi and Tanzanian sublineages of EA1 [7] which our data suggests transferred to those locations from Kenya. Here, the integration of the MDR composite transposon into the chromosome of Kenyan EA1 and EA3 isolates was supported by both examination of genome assembly graphs and analysis of IS*1* insertion sites (see **Methods**), both of which supported integration near gene *cyaA* as has been reported previously [5,7]. Interestingly, MDR in the Ugandan subclade of EA3 was associated with the IncHI1-PST6 plasmid, suggesting that migration of the transposon from plasmid to chromosome may have occurred *in situ* in Kenya after divergence from the Ugandan branch. The three East African lineages also differed markedly in their patterns of mutations associated with reduced susceptibility to fluoroquinolones: GyrB-S464F was conserved among all EA2, whereas all EA3 isolates

carried the GyrA-S83Y mutation, and three distinct GyrA mutations, and the GyrA-S464F mutation were detected at low frequency in EA1 (see **Table 2**).

Supplementary Table 1 – Sequences used in this study (excel file)

Lab_ID	Study_ID	Isolation_da	Sample_sour	Location_clir	Location_zor	Within_study	Patient_age	Patient_sex	Status	Bacteriology	Genomic_typ	Analysis_incl	Lane_and_tr	Accession_n	Typhi_genot	IncHI1
MR4478	MR2001A	13-07-16	Stool	Ruben	Rurii	Yes	1 yrs. 3 mo.	F	Control	E. coli spp+ S	Enterobacter	Non-Typhi or	22306_3_25	ERS4397906	-	-
MCC2873	MCC2140	29-06-16	Stool	City council	Wapewape	Yes	1 yrs. 3 mo.	F	Control	Proteus spp -	E. coli	Non-Typhi or	22306_3_24	ERS4397905	-	-
MR3470	MR1350A	31-07-15	Blood	Ruben	Simbacool	Yes	3 yrs. 6 mo.	F	Case	Salmonella e	S. Typhi	Included	22204_7_17	ERS4397835	4.3.1.2.EA2	Yes
MMM5660	MMM3858B	26-09-16	Blood	Njenga	Quarry	Yes	6 yrs.	M	Case	Salmonella T	S. Agona ST1	Non-Typhi or	22204_7_19	ERS4397756	-	-
MCC3187	MCC2365	14-09-16	Stool	City council	Plot 10	No	2 yrs. 1 mo.	M	Control	Klebsiella sp	Enterobacter	Non-Typhi or	22204_7_19	ERS4397755	-	-
MR4670	MR2154A	21-09-16	Stool	Ruben	Vietnam	Yes	5 yrs. 10 mo.	M	Control	Klebsiella sp	E. coli ST93	Non-Typhi or	24276_3_27	ERS4397732	-	-
MCC3298	MCC2442	17-10-16	Blood	City council	Tassia	No	3 yrs. 4 mo.	M	Case	Salmonella T	Acinetobacte	Non-Typhi or	24276_3_19	ERS4397724	-	-
MCC3379	MCC2504	09-11-16	Blood	City council	Plot 10	No	3 yrs. 3 mo.	M	Case	Salmonella T	Acinetobacte	Non-Typhi or	24276_3_19	ERS4397722	-	-
MMM191	MMM0186	04-09-13	Stool	Njenga	AA/MCC	Yes	1 yrs. 6 mo.	M	Control	E. coli spp + S	S. Typhi	Included in H	22204_8_17	ERS3403506	4.3.1.1.EA1	Yes
MR835	MR0215C	23-01-14	Blood	Ruben	Lungalunga	No	4 yrs. 6 mo.	M	Case	Salmonella T	S. Typhi	Included	22204_8_60	ERS3403486	4.3.1.2.EA2	Yes
MR1063	MR0316B	05-03-14	Stool	Ruben	Vietnam	Yes	1 yrs. 6 mo.	F	Control	Salmonella T	S. Typhi	Included	22204_8_59	ERS3403485	4.3.1.1.EA1	Yes
MB1425	MB1043	03-09-15	Blood	Mbagathi ho	Gatwikira	No	8 yrs.	M	Case	Salmonella T	S. Typhi	Included	22204_8_29	ERS3403466	4.3.1.1.EA1	Yes
MB1638	MB1213	16-10-15	Blood	Mbagathi ho	Pumwani	No	12 yrs. 10 mo	F	Case	Salmonella T	S. Typhi	Included	22204_8_16	ERS3403456	4.3.1.2.EA3	-
MMM4044	MMM2759A	12-10-15	Stool	Njenga	Quarry	Yes	6 yrs.	M	Case	Salmonella T	S. Typhimuri	Non-Typhi or	22204_8_10	ERS3403452	-	-
MB2386	MB1770	10-05-16	Stool	Mbagathi ho	Kibera	No	11 yrs. 1 mo.	F	Control	Klebsiella sp	S. Typhi	Included	22204_8_16	ERS3403444	4.3.1.2.EA2	Yes
MCC800	MCC0602	25-08-14	Stool	City council	Motomoto	Yes	5 mo.	M	Control	No growth	S. Typhi	Included	22204_8_91	ERS3403391	4.3.1.1.EA1	Yes
MR4799	MR2243A	25-10-16	Blood	Ruben	Mombasa	Yes	5 yrs. 4 mo.	M	Case	Salmonella T	S. Typhi	Included in H	24276_3_27	ERR3332782	4.3.1.2.EA2	-
MMM5877	MMM3913B	01-11-16	Stool	Njenga	Tassia	No	10 yrs.	F	Case	Salmonella T	S. Typhi	Included	24276_3_24	ERR3332781	2.5.0	-
MMM5742	MMM3924A	10-10-16	Blood	Njenga	Vietnam	Yes	1 yrs.	M	Case	Salmonella T	S. Typhi	Included in H	24276_3_23	ERR3332780	4.3.1.1.EA1	-
MB3007	MB2245	03-11-16	Stool	Mbagathi ho	Korogocho	No	14 yrs. 4 mo.	M	Control	E. coli spp+ K	S. Typhi	Included in H	24276_3_15	ERR3332779	4.3.1.1.EA1	Yes

Supplementary Table 2 – Outgroup sequences used in this study (excel file)

Lane_id	Accession_number	Genotype	Year	Country	Publication
10593_2_44	ERR360657	0.0.1	1998	Cameroon	Wong et al 2015
10562_2_20	ERR357775	0.0.2	2011	Indonesia	Wong et al 2015
10349_1_89	ERR343337	0.0.3	2002	India	Wong et al 2015
10592_2_36	ERR360484	0.1.0	2009	Algeria	Wong et al 2015
10592_2_57	ERR360505	0.1.1	2001	Cameroon	Wong et al 2015
10593_2_55	ERR360668	0.1.2	1999	Algeria	Wong et al 2015
10592_2_38	ERR360486	0.1.3	1998	Peru	Wong et al 2015
10593_2_14	ERR360627	1.1.1	2005	Algeria	Wong et al 2015
10071_3_32	ERR338008	1.1.2	2007	South Africa	Wong et al 2015
10592_2_63	ERR360511	1.1.3	1976	DRC	Wong et al 2015
10493_1_28	ERR352453	1.1.4	2011	Unknown	Wong et al 2015
10561_2_77	ERR357652	1.2.1	1993	Vietnam	Wong et al 2015
10592_2_5	ERR360453	2.0.0	2003	Pakistan	Wong et al 2015
10592_2_2	ERR360450	2.0.1	2003	Pakistan	Wong et al 2015
10349_1_73	ERR343321	2.0.2	2011	Mexico	Wong et al 2015
10541_2_8	ERR357445	2.1.0	2011	East Timor	Wong et al 2015
10060_5_18	ERR331221	2.1.1	2006	Indonesia	Wong et al 2015
10349_1_3	ERR343251	2.1.2	2002	Western Asia	Wong et al 2015
10071_3_76	ERR338052	2.1.3	2004	Indonesia	Wong et al 2015
10349_1_2	ERR343250	2.1.4	2002	Western Asia	Wong et al 2015
10493_1_42	ERR352467	2.1.5	2011	Indonesia	Wong et al 2015
10060_5_49	ERR331252	2.1.6	2009	Indonesia	Wong et al 2015
10562_2_49	ERR357804	2.1.7	1993	Papua New Guinea	Wong et al 2015
9953_5_54	ERR326650	2.1.8	2006	Indonesia	Wong et al 2015
9953_5_57	ERR326653	2.1.9	2007	Indonesia	Wong et al 2015
10060_6_73	ERR331370	2.2.0	2006	Tanzania	Wong et al 2015
10493_1_15	ERR352440	2.2.1	2012	India	Wong et al 2015
10562_2_12	ERR357767	2.2.2	2011	India	Wong et al 2015
10209_5_36	ERR340792	2.2.3	2002	Laos	Wong et al 2015
9475_4_57	ERR279353	2.2.4	2004	India	Wong et al 2015
10593_2_56	ERR360669	2.3.1	2007	Cameroon	Wong et al 2015
10592_2_59	ERR360507	2.3.2	1999	Mali	Wong et al 2015
10593_2_49	ERR360662	2.3.3	2000	Bangladesh	Wong et al 2015
9953_5_69	ERR326665	2.3.4	2010	Laos	Wong et al 2015
10493_1_70	ERR352495	2.3.5	1983	Fiji	Wong et al 2015
10071_3_68	ERR338044	2.4.0	2009	South Africa	Wong et al 2015
10071_3_40	ERR338016	2.4.1	2006	South Africa	Wong et al 2015
10349_1_5	ERR343253	2.5.0	2002	India	Wong et al 2015
10561_2_5	ERR357580	2.5.1	2011	DRC	Wong et al 2015
10060_5_34	ERR331237	3.0.0	2011	Indonesia	Wong et al 2015
10593_2_45	ERR360658	3.0.1	2000	Morocco	Wong et al 2015
10541_2_6	ERR357443	3.0.2	2012	India	Wong et al 2015
10593_2_18	ERR360631	3.1.0	2002	China	Wong et al 2015
10593_2_58	ERR360671	3.1.1	2004	Benin	Wong et al 2015
9953_5_15	ERR326611	3.1.2	2010	Indonesia	Wong et al 2015
10060_6_66	ERR331363	3.2.1	2010	Laos	Wong et al 2015
10592_2_25	ERR360473	3.2.2	2003	Pakistan	Wong et al 2015
10592_2_17	ERR360465	3.3.0	2003	Pakistan	Wong et al 2015
7468_7_78	ERR108677	3.3.1	2009	Tanzania	Wong et al 2015
9953_5_93	ERR326689	3.4.0	2008	Laos	Wong et al 2015
8490_6_44	ERR213262	3.5.0	1996	Fiji	Wong et al 2015
10493_1_76	ERR352501	3.5.1	1985	Fiji	Wong et al 2015
10060_6_15	ERR331312	3.5.2	2010	Laos	Wong et al 2015
10492_1_5	ERR352258	3.5.3	2010	Samoa	Wong et al 2015
10071_8_55	ERR338124	3.5.4	2012	Samoa	Wong et al 2015
10071_8_60	ERR338129	4.1.0	2012	Samoa	Wong et al 2015
9475_6_92	ERR279189	4.1.1	2006	Malawi	Wong et al 2015
8490_6_49	ERR213267	4.2.0	1996	Fiji	Wong et al 2015
8447_8_28	ERR204271	4.2.1	2008	Fiji	Wong et al 2015

8490_5_25	ERR213243	4.2.2	2011	Fiji		Wong et al 2015
8490_6_61	ERR213279	4.2.3	1994	Fiji		Wong et al 2015

Supplementary Table 3 – Kenyan sequences used in temporal analyses (excel file)

Lane.and.tree_ID	Accession_number	Genotype	Year	Date	Country	Publication
16404_4_5	ERR984707	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16404_4_75	ERR984776	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16404_5_36	ERR984827	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16404_5_43	ERR984834	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16404_5_67	ERR984858	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16549_7_4	ERR1010003	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16404_4_46	ERR984748	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16404_4_59	ERR984761	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16404_4_65	ERR984767	4.3.1.1	2013	01-07-13	Kenya	Park et al. 2018
16404_4_66	ERR984768	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16404_4_73	ERR984774	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16404_4_80	ERR984781	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16404_4_87	ERR984788	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16404_4_89	ERR984790	4.3.1.1	2013	01-07-13	Kenya	Park et al. 2018
16404_5_35	ERR984826	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16404_5_37	ERR984828	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16404_5_44	ERR984835	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16404_5_45	ERR984836	4.3.1.1	2013	01-07-13	Kenya	Park et al. 2018
16404_5_51	ERR984842	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16404_5_52	ERR984843	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16404_5_53	ERR984844	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16404_5_60	ERR984851	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16404_5_68	ERR984859	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16404_5_76	ERR984867	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16404_5_77	ERR984868	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16404_5_84	ERR984875	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018

16404_5_85	ERR984876	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16549_7_18	ERR1010016	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16549_7_20	ERR1010018	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16549_7_23	ERR1010021	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16549_8_1	ERR1010089	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16549_8_10	ERR1010098	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16549_8_16	ERR1010104	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16549_8_17	ERR1010105	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16549_8_18	ERR1010106	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16549_8_2	ERR1010090	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16549_8_31	ERR1010119	4.3.1.1	2013	01-07-13	Kenya	Park et al. 2018
16549_8_33	ERR1010121	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16549_8_47	ERR1010133	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16549_8_48	ERR1010134	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16549_8_54	ERR1010140	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16549_8_63	ERR1010148	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16549_8_8	ERR1010096	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16549_8_9	ERR1010097	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16404_5_75	ERR984866	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16404_4_82	ERR984783	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16404_4_53	ERR984755	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16549_8_26	ERR1010114	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16404_5_59	ERR984850	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16404_5_61	ERR984852	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16404_6_3	ERR984885	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16404_6_35	ERR984917	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16549_8_25	ERR1010113	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
16549_8_7	ERR1010095	4.3.1.1	2013	01-07-13	Kenya	Park et al. 2018
16549_8_71	ERR1010156	4.3.1.1	2012	01-07-12	Kenya	Park et al. 2018
10071_8_62	ERR338131	4.3.1.1	2007	01-07-07	Kenya	Wong et al. 2016

10071_8_63	ERR338132	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016
10071_8_64	ERR338133	4.3.1.1	2007	01-07-07	Kenya	Wong et al. 2016
10071_8_67	ERR338136	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016
10071_8_68	ERR338137	4.3.1.1	2007	01-07-07	Kenya	Wong et al. 2016
10071_8_70	ERR338139	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016
10071_8_71	ERR338140	4.3.1.1	2009	01-07-09	Kenya	Wong et al. 2016
10071_8_72	ERR338141	4.3.1.1	2007	01-07-07	Kenya	Wong et al. 2016
10071_8_74	ERR338143	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016
10071_8_76	ERR338145	4.3.1.1	2007	01-07-07	Kenya	Wong et al. 2016
10071_8_78	ERR338147	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016
10071_8_79	ERR338148	4.3.1.1	2007	01-07-07	Kenya	Wong et al. 2016
10071_8_82	ERR338151	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016
10071_8_83	ERR338152	4.3.1.1	2007	01-07-07	Kenya	Wong et al. 2016
10071_8_88	ERR338157	4.3.1.1	2007	01-07-07	Kenya	Wong et al. 2016
10071_8_90	ERR338159	4.3.1.1	2009	01-07-09	Kenya	Wong et al. 2016
10071_8_91	ERR338160	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016
10071_8_94	ERR338163	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016
10071_8_95	ERR338164	4.3.1.1	2007	01-07-07	Kenya	Wong et al. 2016
8525_2_54	ERR212564	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016
8525_3_30	ERR212636	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016
8525_3_32	ERR212638	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016
8525_3_33	ERR212639	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016
8525_3_34	ERR212640	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016
8525_3_37	ERR212643	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016
8525_3_40	ERR212646	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016
8525_3_42	ERR212648	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016
8525_3_45	ERR212651	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016
8525_3_46	ERR212652	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016
8525_3_47	ERR212653	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016
8525_3_48	ERR212654	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016

8525_3_49	ERR212655	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016	
8525_3_51	ERR212657	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016	
8525_3_53	ERR212659	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016	
8525_3_54	ERR212660	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016	
8525_3_55	ERR212661	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016	
8525_3_56	ERR212662	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016	
8525_3_58	ERR212664	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016	
8525_3_59	ERR212665	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016	
8525_3_60	ERR212666	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016	
8525_3_61	ERR212667	4.3.1.1	2008	01-07-08	Kenya	Wong et al. 2016	
8615_4_29	ERR217406	4.3.1.1	2004	01-07-04	Kenya	Wong et al. 2016	
16404_4_39	ERR984741	4.3.1.2	2013	01-07-13	Kenya	Park et al. 2018	
16404_5_69	ERR984860	4.3.1.2	2013	01-07-13	Kenya	Park et al. 2018	
16549_7_5	ERR1010004	4.3.1.2	2013	01-07-13	Kenya	Park et al. 2018	
16549_8_34	ERR1010122	4.3.1.2	2013	01-07-13	Kenya	Park et al. 2018	
10071_8_65	ERR338134	4.3.1.2	2007	01-07-07	Kenya	Wong et al. 2016	
10071_8_66	ERR338135	4.3.1.2	2007	01-07-07	Kenya	Wong et al. 2016	
10071_8_69	ERR338138	4.3.1.2	2007	01-07-07	Kenya	Wong et al. 2016	
10071_8_77	ERR338146	4.3.1.2	2007	01-07-07	Kenya	Wong et al. 2016	
10071_8_80	ERR338149	4.3.1.2	2007	01-07-07	Kenya	Wong et al. 2016	
10071_8_81	ERR338150	4.3.1.2	2007	01-07-07	Kenya	Wong et al. 2016	
10071_8_89	ERR338158	4.3.1.2	2007	01-07-07	Kenya	Wong et al. 2016	
10071_8_92	ERR338161	4.3.1.2	2008	01-07-08	Kenya	Wong et al. 2016	
8525_3_31	ERR212637	4.3.1.2	2008	01-07-08	Kenya	Wong et al. 2016	
8525_3_71	ERR212677	4.3.1.2	2001	01-07-01	Kenya	Wong et al. 2016	
22204_7_274	ERR3332596	4.3.1.1	2014	16-01-14	Kenya	This study	
24276_3_143	ERR3332776	4.3.1.2	2016	18-07-16	Kenya	This study	
22306_3_253	ERR3332773	4.3.1.2	2016	11-07-16	Kenya	This study	
22306_3_248	ERR3332770	4.3.1.2	2014	22-09-14	Kenya	This study	
22306_3_247	ERR3332769	4.3.1.1	2014	15-09-14	Kenya	This study	

22306_3_244	ERR3332766	4.3.1.2	2016	13-07-16	Kenya	This study	
22306_3_228	ERR3332758	4.3.1.2	2014	08-09-14	Kenya	This study	
22306_3_225	ERR3332755	4.3.1.1	2014	10-03-14	Kenya	This study	
22306_3_218	ERR3332750	4.3.1.1	2015	12-03-15	Kenya	This study	
22306_3_207	ERR3332746	4.3.1.1	2014	21-07-14	Kenya	This study	
22306_3_206	ERR3332745	4.3.1.1	2013	03-09-13	Kenya	This study	
22306_3_205	ERR3332744	4.3.1.1	2013	23-09-13	Kenya	This study	
22306_3_204	ERR3332743	4.3.1.2	2014	05-03-14	Kenya	This study	
22306_3_200	ERR3332740	4.3.1.2	2014	26-05-14	Kenya	This study	
22306_3_197	ERR3332737	4.3.1.1	2014	01-09-14	Kenya	This study	
24276_3_244	ERR2909580	4.3.1.2	2014	03-03-14	Kenya	This study	
22306_3_191	ERR3332731	4.3.1.2	2013	07-12-13	Kenya	This study	
22306_3_188	ERR3332729	4.3.1.2	2014	03-04-14	Kenya	This study	
22306_3_185	ERR3332727	4.3.1.2	2014	05-03-14	Kenya	This study	
22306_3_184	ERR3332726	4.3.1.1	2014	16-06-14	Kenya	This study	
22204_8_178	ERS3403506	4.3.1.1	2013	04-09-13	Kenya	This study	
22306_3_177	ERR3332719	4.3.1.2	2013	04-09-13	Kenya	This study	
24276_3_194	ERR2909540	4.3.1.1	2014	08-07-14	Kenya	This study	
22306_3_85	ERR3332652	4.3.1.2	2014	06-03-14	Kenya	This study	
22306_3_75	ERR3332648	4.3.1.2	2013	30-09-13	Kenya	This study	
22306_3_68	ERR3332643	4.3.1.1	2014	03-07-14	Kenya	This study	
22306_3_67	ERR3332642	4.3.1.2	2014	09-07-14	Kenya	This study	
22306_3_66	ERR3332641	4.3.1.1	2014	13-02-14	Kenya	This study	
22306_3_56	ERR3332634	4.3.1.1	2013	03-09-13	Kenya	This study	
22306_3_55	ERR3332633	4.3.1.1	2014	11-06-14	Kenya	This study	
22306_3_38	ERR3332623	4.3.1.1	2013	26-09-13	Kenya	This study	
24276_3_113	ERR2909474	4.3.1.1	2015	28-08-15	Kenya	This study	
22306_3_27	ERR3332617	4.3.1.2	2015	11-08-15	Kenya	This study	
22306_3_164	ERR3332709	4.3.1.2	2015	23-07-15	Kenya	This study	
24276_3_184	ERR2909536	4.3.1.2	2016	29-03-16	Kenya	This study	

22306_3_155	ERR3332701	4.3.1.1	2015	08-07-15	Kenya	This study	
22306_3_147	ERR3332694	4.3.1.1	2016	03-05-16	Kenya	This study	
24276_3_168	ERR2909520	4.3.1.2	2015	02-07-15	Kenya	This study	
22306_3_140	ERR3332691	4.3.1.2	2015	02-07-15	Kenya	This study	
22306_3_138	ERR3332690	4.3.1.2	2015	03-07-15	Kenya	This study	
24276_3_116	ERR2909477	4.3.1.1	2015	28-09-15	Kenya	This study	
22306_3_129	ERR3332685	4.3.1.2	2016	13-06-16	Kenya	This study	
22306_3_126	ERR3332684	4.3.1.2	2014	23-07-14	Kenya	This study	
24276_3_185	ERR2909537	4.3.1.2	2016	06-04-16	Kenya	This study	
22306_3_116	ERR3332677	4.3.1.1	2016	07-06-16	Kenya	This study	
22306_3_115	ERR3332676	4.3.1.1	2016	02-06-16	Kenya	This study	
22306_3_108	ERR3332669	4.3.1.2	2015	08-07-15	Kenya	This study	
22306_3_96	ERR3332662	4.3.1.2	2014	04-02-14	Kenya	This study	
22306_3_93	ERR3332660	4.3.1.2	2016	11-01-16	Kenya	This study	
22306_3_92	ERR3332659	4.3.1.2	2015	02-11-15	Kenya	This study	
24276_3_171	ERR2909523	4.3.1.2	2015	01-10-15	Kenya	This study	
24276_3_170	ERR2909522	4.3.1.2	2015	13-08-15	Kenya	This study	
24276_3_214	ERR2909557	4.3.1.1	2013	12-10-13	Kenya	This study	
24276_3_115	ERR2909476	4.3.1.2	2015	14-09-15	Kenya	This study	
24276_3_277	ERR2909611	4.3.1.1	2013	21-11-13	Kenya	This study	
24276_3_249	ERR2909585	4.3.1.2	2013	23-09-13	Kenya	This study	
24276_3_129	ERR2909490	4.3.1.2	2016	10-02-16	Kenya	This study	
24276_3_126	ERR2909487	4.3.1.1	2016	26-01-16	Kenya	This study	
24276_3_180	ERR2909532	4.3.1.1	2016	08-02-16	Kenya	This study	
22204_7_161	ERR3332562	4.3.1.1	2013	11-11-13	Kenya	This study	
24276_3_278	ERR2909612	4.3.1.2	2013	27-11-13	Kenya	This study	
24276_3_248	ERR2909584	4.3.1.2	2013	23-09-13	Kenya	This study	
24276_3_197	ERR2909543	4.3.1.2	2013	20-08-13	Kenya	This study	
24276_3_256	ERR2909592	4.3.1.1	2015	03-06-15	Kenya	This study	
24276_3_177	ERR2909529	4.3.1.2	2016	28-01-16	Kenya	This study	

24276_3_155	ERR2909508	4.3.1.1	2015	17-03-15	Kenya	This study	
24276_3_247	ERR2909583	4.3.1.2	2013	23-09-13	Kenya	This study	
24276_3_159	ERR2909511	4.3.1.2	2015	05-05-15	Kenya	This study	
24276_3_193	ERR2909539	4.3.1.2	2014	28-04-14	Kenya	This study	
24276_3_164	ERR2909516	4.3.1.2	2014	16-10-14	Kenya	This study	
22204_7_146	ERR3332556	4.3.1.2	2013	29-08-13	Kenya	This study	
22204_7_145	ERR3332555	4.3.1.1	2016	01-03-16	Kenya	This study	
24276_3_122	ERR2909483	4.3.1.1	2015	17-11-15	Kenya	This study	
22204_7_142	ERR2525481	4.3.1.1	2015	09-11-15	Kenya	This study	
24276_3_135	ERR2909496	4.3.1.2	2016	24-02-16	Kenya	This study	
24276_3_226	ERR2909568	4.3.1.1	2016	08-02-16	Kenya	This study	
24276_3_153	ERR2909506	4.3.1.2	2015	04-03-15	Kenya	This study	
24276_3_157	ERR2909509	4.3.1.2	2015	21-04-15	Kenya	This study	
22204_7_130	ERR3332551	4.3.1.1	2015	24-04-15	Kenya	This study	
22204_7_122	ERR3332548	4.3.1.1	2015	23-06-15	Kenya	This study	
22204_7_115	ERR3332546	4.3.1.2	2015	08-04-15	Kenya	This study	
22204_7_111	ERR3332545	4.3.1.2	2015	11-06-15	Kenya	This study	
22204_7_107	ERR3332543	4.3.1.2	2015	23-11-15	Kenya	This study	
22204_7_105	ERR3332542	4.3.1.2	2015	13-11-15	Kenya	This study	
22204_7_104	ERR3332541	4.3.1.2	2015	11-11-15	Kenya	This study	
22204_7_97	ERR3332539	4.3.1.1	2016	08-02-16	Kenya	This study	
24276_3_272	ERR2909608	4.3.1.2	2016	09-03-16	Kenya	This study	
22204_7_95	ERR2525467	4.3.1.1	2016	09-03-16	Kenya	This study	
22204_7_262	ERR2525516	4.3.1.2	2015	07-05-15	Kenya	This study	
22204_7_261	ERR2525515	4.3.1.1	2015	29-06-15	Kenya	This study	
22204_7_258	ERR3332591	4.3.1.2	2015	17-02-15	Kenya	This study	
22204_7_256	ERR3332589	4.3.1.2	2015	27-04-15	Kenya	This study	
22204_7_251	ERR2525508	4.3.1.2	2015	18-03-15	Kenya	This study	
22204_7_243	ERR3332587	4.3.1.2	2015	26-02-15	Kenya	This study	
24276_3_152	ERR2909505	4.3.1.1	2015	19-02-15	Kenya	This study	

22204_7_239	ERR3332586	4.3.1.2	2015	03-06-15	Kenya	This study	
24276_3_160	ERR2909512	4.3.1.2	2015	26-05-15	Kenya	This study	
24276_3_162	ERR2909514	4.3.1.2	2015	08-06-15	Kenya	This study	
22204_7_237	ERR3332585	4.3.1.2	2015	12-06-15	Kenya	This study	
22204_7_236	ERR2525501	4.3.1.2	2015	12-06-15	Kenya	This study	
24276_3_163	ERR2909515	4.3.1.2	2015	15-06-15	Kenya	This study	
22204_7_232	ERR3332584	4.3.1.2	2015	21-05-15	Kenya	This study	
22204_7_223	ERR2525493	4.3.1.2	2015	13-04-15	Kenya	This study	
22204_7_219	ERR2525491	4.3.1.2	2015	12-05-15	Kenya	This study	
24276_3_142	ERR2909503	4.3.1.1	2015	15-01-15	Kenya	This study	
24276_3_141	ERR2909502	4.3.1.2	2015	13-01-15	Kenya	This study	
22204_7_216	ERR2525488	4.3.1.1	2015	14-01-15	Kenya	This study	
22204_7_215	ERR2525487	4.3.1.1	2016	25-02-16	Kenya	This study	
24276_3_161	ERR2909513	4.3.1.2	2015	08-06-15	Kenya	This study	
22204_7_210	ERR2525484	4.3.1.2	2015	19-10-15	Kenya	This study	
24276_3_167	ERR2909519	4.3.1.2	2015	04-06-15	Kenya	This study	
22204_7_209	ERR2525483	4.3.1.2	2015	26-10-15	Kenya	This study	
24276_3_134	ERR2909495	4.3.1.1	2016	24-02-16	Kenya	This study	
24276_3_182	ERR2909534	4.3.1.2	2016	22-02-16	Kenya	This study	
22204_7_208	ERR3332581	4.3.1.2	2016	09-11-16	Kenya	This study	
22204_7_207	ERR3332580	4.3.1.2	2016	17-11-16	Kenya	This study	
24276_3_146	ERR3332777	4.3.1.1	2016	03-10-16	Kenya	This study	
22204_7_193	ERR3332578	4.3.1.2	2016	26-09-16	Kenya	This study	
22204_7_188	ERR3332577	4.3.1.2	2016	25-10-16	Kenya	This study	
24276_3_149	ERR3332778	4.3.1.1	2016	26-10-16	Kenya	This study	
22204_7_185	ERR3332576	4.3.1.1	2016	11-10-16	Kenya	This study	
22204_7_184	ERR3332575	4.3.1.1	2016	11-10-16	Kenya	This study	
22204_7_181	ERR3332574	4.3.1.2	2016	24-10-16	Kenya	This study	
24276_3_276	ERR3332782	4.3.1.2	2016	25-10-16	Kenya	This study	
24276_3_150	ERR3332779	4.3.1.1	2016	03-11-16	Kenya	This study	

24276_3_238	ERR3332780	4.3.1.1	2016	10-10-16	Kenya	This study	
22204_7_180	ERR3332573	4.3.1.1	2016	13-10-16	Kenya	This study	
22306_3_57	ERR3332635	4.3.1.2	2014	10-10-14	Kenya	This study	

Supplementary Table 4. Culture positive typhoid cases and asymptomatic carriers

Typhoid Cases	**Total**	**<=1 year**	**1-7 years**	**7-16 years**	**Linear regression association with age (p-value)**
Febrile participants subjected to stool and blood culture	4,670	630	3,299	741	-
S. Typhi positive	148 (3.2%)	14 (2.2%)	102 (3.1%)	32 (4.3%)	0.000525*
- Female	49 (1.0%)	6 (0.95%)	36 (1.1%)	7 (0.94%)	0.536
- Male	99 (2.1%)	8 (1.3%)	66 (2.0%)	25 (3.4%)	0.0000773*
WGS confirmed *S.* Typhi positive	100 (2.1%)	7 (1.1%)	67 (2.0%)	26 (3.5%)	0.0001*
- Female	36 (0.77%)	4 (0.63%)	25 (0.76%)	7 (0.94%)	0.508
- Male	64 (1.4%)	3 (0.48%)	42 (1.27%)	19 (2.6%)	0.000009*
Asymptomatic Carriers	**8,549**	**641**	**5,495**	**2,413**	**-**
Total stool cultures from non-febrile age-matched controls	8,530	641	5,480	2,409	-
S. Typhi positive	95 (1.1%)	4 (0.62%)	62 (1.1%)	29 (1.2%)	0.402
- Female	46 (0.54%)	3 (0.47%)	31 (0.57%)	12 (0.50%)	0.585
- Male	49 (0.57%)	1 (0.16%)	31 (0.57%)	17 (0.71%)	0.081
WGS confirmed *S.* Typhi positive	55 (0.64%)	4 (0.62%)	31 (0.57%)	20 (0.83%)	0.256
- Female	29 (0.34%)	3 (0.47%)	16 (0.29%)	10 (0.42%)	0.909
- Male	26 (0.30%)	1 (0.16%)	15 (0.27%)	10 (0.42%)	0.127

Supplementary Table 5 – Comparison of phenotypic and genotypic AMR profiles of 136 (n=128 H58, n=8 Non-H58) high quality *S.* Typhi genome sequences

Drug_class	Resistant phenotype*	Resistant genotype	Very major error	Susceptible phenotype	Susceptible genotype	Major error	Sensitivity	Specificity	PPV
Beta-lactamases									
Ampicillin	100	95	5 (3.7%)	36	28	8 (5.9%)	95%	77.8 %	92.2%
Ceftazidime	1	0	1 (0.74%)	135	135	0	0%	100%	0%
Cefotaxime	4	0	4 (2.9%)	132	132	0	0%	100%	0%
Ceftriaxone	1	0	1 (0.74%)	135	135	0	0%	100%	0%
Cefpodoxime	4	0	4 (2.9%)	132	132	0	0%	100%	0%
Chloramphenicol									
Chloramphenicol	93	92	1 (0.74%)	43	32	10 (7.4%)	98.9%	74.4%	89.3%
Tetracyclines									
Tetracycline	69	65	4 (2.9%)	67	59	8 (5.9%)	94.2%	88.1%	89.0%
Folate pathway inhibitors									
Co-trimoxazole	98	96	2 (1.5%)	38	31	7 (5.1%)	98.0%	81.6%	93.2%
Quinolones and Fluoroquinolones									
Ciprofloxacin	63	55	8 (5.9%)	73	41	32 (23.5%)	87.3%	56.2%	63.2%
Nalidixic acid	65	21	44 (32.4%)	71	71	0	32.3%	100%	100%

*indicates where resistant and intermediate phenotypes have been combined as resistant

Supplementary Table 6 – Distribution of n=153 *S.* Typhi genotypes among each sex for cases and controls

	Female	Male

WGS-confirmed cases	**35**	**64**
EA1	12 (34.3%)	23 (35.9%)
EA2	16 (45.7%)	30 (46.9%)
EA3	4 (11.4%)	11 (17.9%)
non-H58	3 (8.57%)	0
WGS-confirmed carriers	**28**	**26**
EA1	10 (35.7%)	10 (38.5%)
EA2	18 (64.3%)	9 (34.6%)
EA3	0	2 (7.69%)
non-H58	0	5 (19.2%)

Supplementary table 7 - Climatic predictors of elevated case and carrier counts for all samples

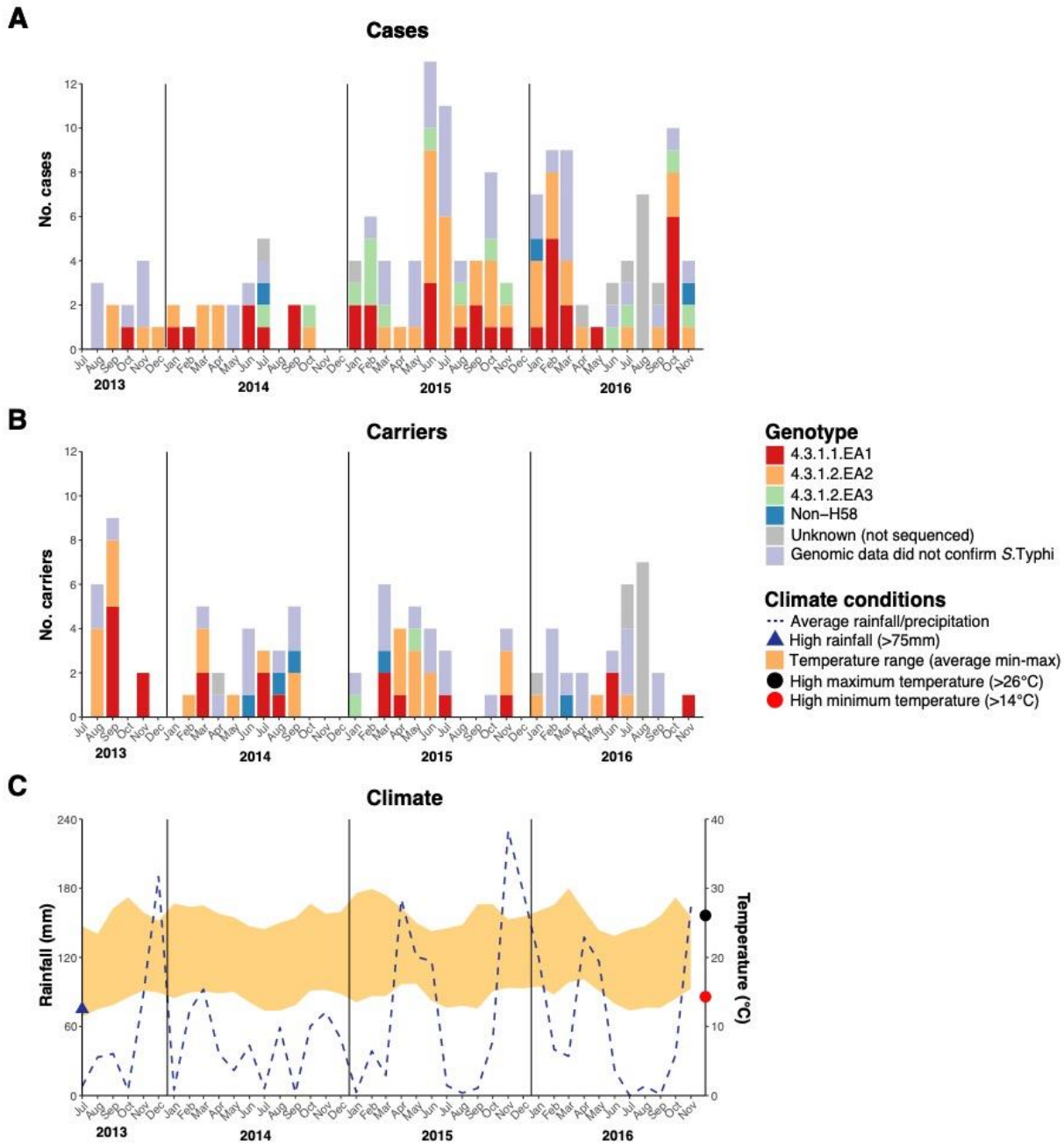
Typhoid Cases						
Month	**Same month**		**Previous month**		**2 months prior**	
	OR (95% CI)	**p-value**	**OR (95% CI)**	**p-value**	**OR (95% CI)**	**p-value**
Rainfall (precipitation) > 75 mm	0.65 (0.13-3.2)	0.72	0.85 (0.17-4.4)	1	2.4 (0.46-17.1)	0.29
Minimum temperature >14°C	0.17 (0.025-0.86)	0.022*	0.39 (0.080-1.7)	0.20	0.60 (0.13-2.6)	0.53
Maximum temperature >26°C	0.43 (0.094-1.8)	0.22	0.43 (0.094-1.8)	0.22	0.55 (0.12-2.3)	0.52
Asymptomatic Carriers						
Month	**Same month**		**Previous month**		**2 months prior**	
	OR (95% CI)	**p-value**	**OR (95% CI)**	**p-value**	**OR (95% CI)**	**p-value**
Rainfall (precipitation) > 75 mm	0.95 (0.19-4.5)	1	1.18 (0.23-5.9)	1	1.18 (0.23-5.9)	1
Minimum temperature >14°C	0.16 (0.029-0.74)	0.0095*	0.72 (0.17-3.1)	0.75	1.1 (0.26-4.7)	1
Maximum temperature >26°C	0.19 (0.037-0.86)	0.024*	0.30 (0.063-1.3)	0.11	0.82 (0.19-3.4)	1

Values in cells are odds ratios and p-values for Fisher's exact test between high case/control count and high rainfall/temperature. * highlights p-value <0.05.

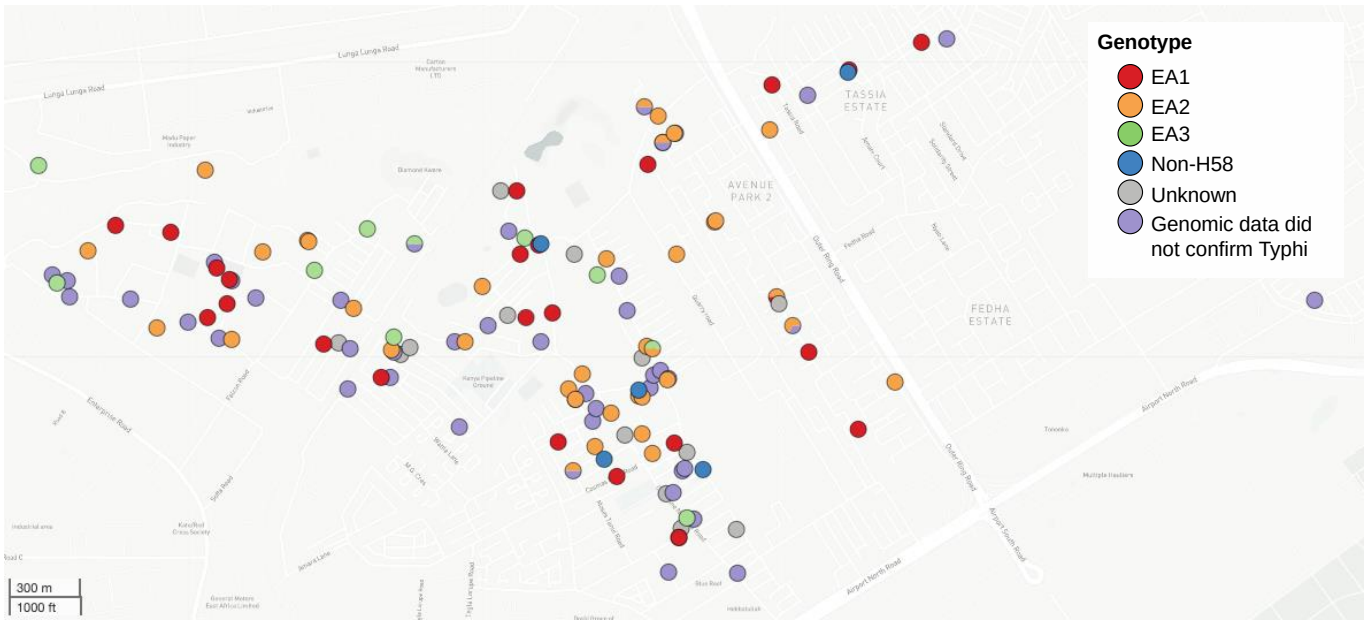
Supplementary Table 8: nonsynonymous (NS) Mutations among S. Typhi isolates

Lane_and_tree_ID	Accession_number	Status	Study_area	Position_in_	Ancestor_All	Derived_Alle	CT18_Gene_	Classificatior	Gene_produ	Functional_c	Codon	Ancestor_Co	Derived_Cod	Ancestor_An	Derived_Am	Mutation
22204_8_178	ERS3403506	Control	Yes	4626124	C	A	ytfM/STY476	4.1.0	Putative exp	Periplasmic/	322	AAC	AAA	N	K	ytfM-N322K
24276_3_116	ERS1909313	Case	No	4021517	G	A	yhjW/STY41(	4.1.1	putative mer	Inner membr	168	GGC	AGC	G	S	yhjW-G168S
24276_3_116	ERS1909313	Case	No	3385249	C	T	yhcP/STY354	4.1.1	Putative mer	Inner membr	106	TGT	TAT	C	Y	yhcP-C106Y
22204_7_236	ERS1573056	Control	Yes	2171000	G	A	yegD/STY233	3.0.2	conserved hy	Unknown fur	183	GCC	ACC	A	T	yegD-A183T
22306_3_66	ERS3403489	Case	No	962807	T	C	ycaM/STY09(	1.5.0	Probable tra	Transport/bi	440	TTC	CTC	F	L	ycaM-F440L
22204_7_145	ERS3403339	Case	Yes	868216	T	G	ybiR/STY087	4.1.1	Putative mer	Inner membr	322	TTT	GTT	F	V	ybiR-F322V
24276_3_146	ERS3403282	Case	No	662248	A	G	ybdQ/STY06(	0.0.2	Conserved hy	Unknown fur	128	TCA	CCA	S	P	ybdQ-S128P
22204_7_236	ERS1573056	Control	Yes	2143349	T	G	wzxC/STY231	4.1.4	putative tran	Surface polys	26	TTA	TTC	L	F	wzxC-L26F
22204_7_130	ERS3403329	Control	Yes	2160205	A	C	wza/STY2331	4.1.4	Putative poly	Surface polys	137	GTC	GGC	V	G	wza-V137G
24276_3_129	ERS1909314	Case	No	4725276	A	T	uxuR/STY486	6.1.1	uxu operon t	Global regula	71	ATC	TTC	I	F	uxuR-I71F
24276_3_159	ERS1909348	Control	No	4517487	G	A	tviE/STY4656	5.1.5	Vi polysacch	Surface polys	263	CCG	TCG	P	S	tviE-P263S
24276_3_194	ERS1909377	Control	Yes	4521071	G	A	tviD/STY4655	5.1.5	Vi polysacch	Surface polys	159	CGT	TGT	R	C	tviD-R159C
24276_3_135	ERS1909320	Case	No	4523895	C	T	tviB/STY4661	5.1.5	Vi polysacch	Surface polys	1	GTG	ATG	V	M	tviB-V1M
24276_3_149	ERS3403283	Case	No	1282526	G	A	trpB/STY132	3.1.19	tryptophan s	Amino acid b	35	GCG	GTG	A	V	trpB-A35V
24276_3_161	ERS1909350	Case	No	4653330	G	A	treB/STY479	1.5.37	PTS system, t	Transport/bi	323	ACC	ATC	T	I	treB-T323I
22204_7_111	ERS3403322	Control	Yes	4318042	C	T	STY4449	4.1.0	putative lipo	Periplasmic/	55	TGG	TAG	W	*	STY4449-W55*
24276_3_150	ERS3403284	Control	No	4104133	C	T	STY4236	4.1.1	putative mer	Inner membr	43	TGG	TAG	W	*	STY4236-W43*
24276_3_194	ERS1909377	Control	Yes	4102295	C	T	STY4235	1.5.30	heavy metal-	Transport/bi	420	GGC	GAC	G	D	STY4235-G420D
24276_3_194	ERS1909377	Control	Yes	3788415	G	T	STY3925	1.5.37	probable PTS	Transport/bi	38	GTT	TTT	V	F	STY3925-V38F
24276_3_146	ERS3403282	Case	No	3695295	G	A	STY3837	4.1.0	putative lipo	Periplasmic/	36	GCG	ACG	A	T	STY3837-A36T
22204_7_216	ERS1573035	Case	Yes	3628992	G	A	STY3765	0.0.0	hypothetical	Unknown fur	377	CCG	TCG	P	S	STY3765-P377S
22306_3_185	ERS3403513	Control	Yes	3471310	A	C	STY3618	4.1.1	Putative mer	Inner membr	35	TTA	GTA	L	V	STY3618-L35V
24276_3_193	ERS1909376	Case	Yes	3471046	T	C	STY3618	4.1.1	Putative mer	Inner membr	123	ATT	GTT	I	V	STY3618-I123V
24276_3_116	ERS1909313	Case	No	3170381	G	T	STY3325	1.1.1	methyl-acce	Chemotaxis a	493	GCG	GAG	A	E	STY3325-A493E
24276_3_182	ERS1909371	Case	Yes	3119041	G	A	STY3257	3.2.06	possible oxy	Biosynthesis	82	GAC	AAC	D	N	STY3257-D82N
24276_3_194	ERS1909377	Control	Yes	2501788	C	T	STY2664	1.7.1	cell division	Cell division	299	GTC	ATC	V	I	STY2664-V299I
22204_7_130	ERS3403329	Control	Yes	2480004	G	C	STY2645	4.1.1	Putative mer	Inner membr	88	GAA	CAA	E	Q	STY2645-E88Q
22204_7_274	ERS3403567	Case	Yes	2399174	G	T	STY2563	7.0.0	putative sodi	Unclassified/	557	TCC	TAC	S	Y	STY2563-S557Y
22204_7_216	ERS1573035	Case	Yes	2333750	T	C	STY2499	2.2.03	DNA gyrase s	DNA - replica	87	GAC	GGC	D	G	gyrA-D87G
22306_3_207	ERS3403522	Case	Yes	2333750	T	C	STY2499	2.2.03	DNA gyrase s	DNA - replica	87	GAC	GGC	D	G	gyrA-D87G
22204_7_232	ERS3403299	Control	Yes	2194420	G	A	STY2361	4.1.0	putative exp	Periplasmic/	100	CCA	CTA	P	L	STY2361-P100L
22204_7_207	ERS3403295	Case	Yes	2051377	G	A	STY2216	4.1.1	putative inne	Inner membr	397	GCC	GTC	A	V	STY2216-A397V
22306_3_66	ERS3403489	Case	No	1749811	G	T	STY1831	0.0.2	Conserved hy	Unknown fur	110	GGA	TGA	G	*	STY1831-G110*
22204_7_274	ERS3403567	Case	Yes	1670239	G	A	STY1750	3.3.15	putative ami	Central inter	9	CGG	TGG	R	W	STY1750-R9W
22306_3_66	ERS3403489	Case	No	1533338	T	C	STY1587	4.1.1	putative mer	Inner membr	77	GTA	GCA	V	A	STY1587-V77A
22204_7_105	ERS3403315	Case	No	1491452	A	G	STY1537	6.1.1	putative reg	Global regula	42	GAC	GGC	D	G	STY1537-D42G
22204_7_208	ERS3403296	Case	No	1456243	G	A	STY1499	0.0.0	hypothetical	Unknown fur	187	CAA	TAA	Q	*	STY1499-Q187*
22306_3_66	ERS3403489	Case	No	1410841	T	C	STY1460	2.1.4	putative pep	Degradation	250	TAT	CAT	Y	H	STY1460-Y250H
22204_7_237	ERS3403300	Case	Yes	1396606	C	T	STY1443	4.1.0	putative exp	Periplasmic/	185	CAG	TAG	Q	*	STY1443-Q185*
22306_3_66	ERS3403489	Case	No	1385997	T	C	STY1434	4.1.1	putative mer	Inner membr	274	AAC	AGC	N	S	STY1434-N274S
24276_3_116	ERS1909313	Case	No	1369705	C	T	STY1419	7.0.0	probable pyr	Unclassified/	613	GCG	ACG	A	T	STY1419-A613T
22204_7_216	ERS1573035	Case	Yes	1239385	T	C	STY1284	5.1.5	putative inva	Pathogenicit	466	TGG	CGG	W	R	STY1284-W466R
24276_3_135	ERS1909320	Case	No	1061235	A	G	STY1083	1.5.01	ABC transpor	Transport/bi	64	CAG	CGG	Q	R	STY1083-Q64R
22204_7_130	ERS3403329	Control	Yes	339227	G	A	STY0326	5.1.5	conserved hy	Pathogenicit	84	TGG	TGA	W	*	STY0326-W84*
24276_3_150	ERS3403284	Control	No	337744	C	T	STY0324	5.1.5	Rhs-family pr	Pathogenicit	427	CAG	TAG	Q	*	STY0324-Q427*
22204_7_216	ERS1573035	Case	Yes	319124	C	T	STY0306	5.1.5	Putative mer	Pathogenicit	36	CGT	TGT	R	C	STY0306-R36C
22204_7_216	ERS1573035	Case	Yes	306487	C	T	STY0290	5.1.5	conserved hy	Pathogenicit	151	CGC	CAC	R	H	STY0290-R151H
22204_7_130	ERS3403329	Control	Yes	306208	C	T	STY0289	5.1.5	hypothetical	Pathogenicit	74	CGG	CAG	R	Q	STY0289-R74Q
22204_7_216	ERS1573035	Case	Yes	1641786	G	A	sscA/STY172	5.1.5	putative Typ	Pathogenicit	142	CTC	TTC	L	F	sscA-L142F
24276_3_194	ERS1909377	Control	Yes	1874307	G	A	sopE2/STY19	5.1.5	putative inva	Pathogenicit	9	CAG	TAG	Q	*	sopE2-Q9*
22204_7_232	ERS3403299	Control	Yes	4603536	C	T	sgaH/STY474	3.3.15	putative hex	Central inter	186	CAT	TAT	H	Y	sgaH-H186Y
24276_3_168	ERS1909357	Case	Yes	4227390	G	A	rpsG/STY435	4.2.2	30S ribosom	Ribosome co	43	GTA	ATA	V	I	rpsG-V43I
22204_7_274	ERS3403567	Case	Yes	4606083	G	A	rpsF/STY474	4.2.2	30s ribosom	Ribosome co	9	ATG	ATA	M	I	rpsF-M9I
24276_3_194	ERS1909377	Control	Yes	4794595	C	T	rob/STY4933	1.2.1	right origin-b	Chromosome	185	CGC	CAC	R	H	STY4933-R185H
22204_7_274	ERS3403567	Case	Yes	219180	C	T	pncB/STY02(	2.2.11	poly(A) polyr	RNA synthes	174	AGC	AAC	S	N	pcnB-S174N
22204_7_130	ERS3403329	Control	Yes	4080652	C	T	pitA/STY421	1.5.23	putative low-	Transport/bi	340	CGT	CAT	R	H	STY4214-R340H
22306_3_68	ERS3403491	Control	Yes	687449	C	T	pbpA/STY06(	4.1.2	penicillin-bin	Murein sacc	409	GGT	AGT	G	S	mrdA-G409S
24276_3_150	ERS3403284	Control	No	3892927	G	T	misL/STY403	5.1.5	putative viru	Pathogenicit	402	CGC	AGC	R	S	misL-R402S
22204_7_274	ERS3403567	Case	Yes	2216980	G	A	metG/STY23	2.2.01	methionyl-tR	Amino acyl t	554	GCA	ACA	A	T	metG-A554T
24276_3_168	ERS1909357	Case	Yes	4294622	T	G	malE/STY442	1.5.03	periplasmic r	Transport/bi	340	GAC	GCC	D	A	malE-D340A
22306_3_66	ERS3403489	Case	No	742865	T	C	kdpE/STY074	6.1.1	KDP operon t	Global regula	3	AAC	AGC	N	S	kdpE-N3S
22204_7_232	ERS3403299	Control	Yes	2841954	G	A	hypf/STY296	3.5.2	hydrogenase	Anaerobic re	305	CCA	CTA	P	L	hydA-P305L
24276_3_150	ERS3403284	Control	No	1103229	C	A	hpcG/STY113	3.4.3	2-oxo-hepta-	Degradation	261	GCG	GAG	A	E	hpcG-A261E
22204_7_236	ERS1573056	Control	Yes	2420019	G	A	hisJ/STY2584	1.5.03	histidine-bin	Transport/bi	92	TCG	TTG	S	L	hisJ-S92L
22204_7_130	ERS3403329	Control	Yes	3810322	C	T	gyrB/STY394	2.2.03	DNA gyrase s	DNA - replica	464	TCC	TTC	S	F	gyrB-S464F
24276_3_135	ERS1909320	Case	No	4142147	G	A	glgC/STY427	2.2.08	glucose-1-ph	Polysacchari	110	GGC	AGC	G	S	glgC-G110S
22306_3_66	ERS3403489	Case	No	2984469	T	C	fucI/STY311	3.4.3	L-fuculose is	Degradation	238	TTC	CTC	F	L	fucI-F238L
24276_3_146	ERS3403282	Case	No	231921	G	A	fhuD/STY022	1.5.23	ferrichrome-	Transport/bi	246	ATG	ATA	M	I	fhuD-M246I
24276_3_135	ERS1909320	Case	No	628273	G	A	fes/STY0629	1.5.23	enterochelin	Transport/bi	51	TGG	TGA	W	*	fes-W51*
22204_7_236	ERS1573056	Control	Yes	3176266	G	A	exbB/STY333	1.5.23	biopolymer t	Transport/bi	54	CGC	TGC	R	C	exbB-R54C
24276_3_150	ERS3403284	Control	No	2038114	C	T	dcm/STY220	2.2.03	DNA-cytosin	DNA - replica	419	GCG	ACG	A	T	dcm-A419T
22204_8_178	ERS3403506	Control	Yes	1692237	C	A	btuC/STY177	1.5.02	vitamin B12 t	Transport/bi	294	GCC	TCC	A	S	btuC-A294S
24276_3_135	ERS1909320	Case	No	4018331	C	A	bisC/STY415	7.0.0	biotin sulfoxi	Unclassified/	570	CCG	CAG	P	Q	bisC-P570Q
24276_3_146	ERS3403282	Case	No	3346023	G	A	arcB/STY350	6.1.1	aerobic respi	Global regula	443	CCG	CTG	P	L	arcB-P443L
24276_3_135	ERS1909320	Case	No	2169799	C	T	alkA/STY233	2.2.03	DNA-3-meth	DNA - replica	174	ATG	ATA	M	I	alkA-M174I
22204_7_274	ERS3403567	Case	Yes	2551336	C	T	aegA/STY271	7.0.0	putative oxid	Unclassified/	116	GCC	ACC	A	T	aegA-A116T
24276_3_197	ERS1909380	Control	Yes	2551116	G	T	aegA/STY271	7.0.0	putative oxid	Unclassified/	189	GCG	GAG	A	E	aegA-A189E

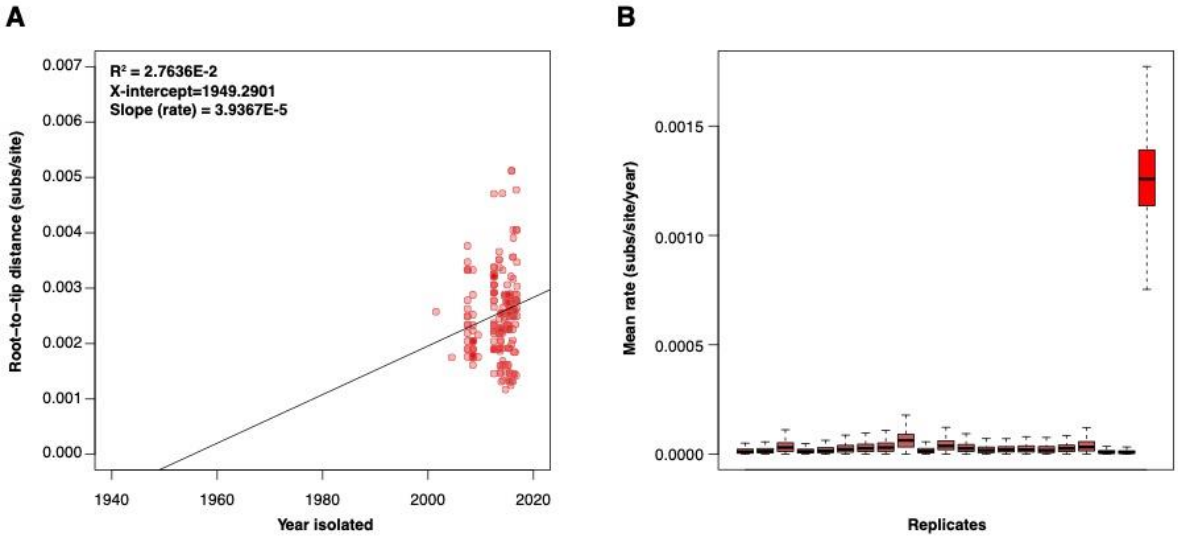
Supplementary Figure 1



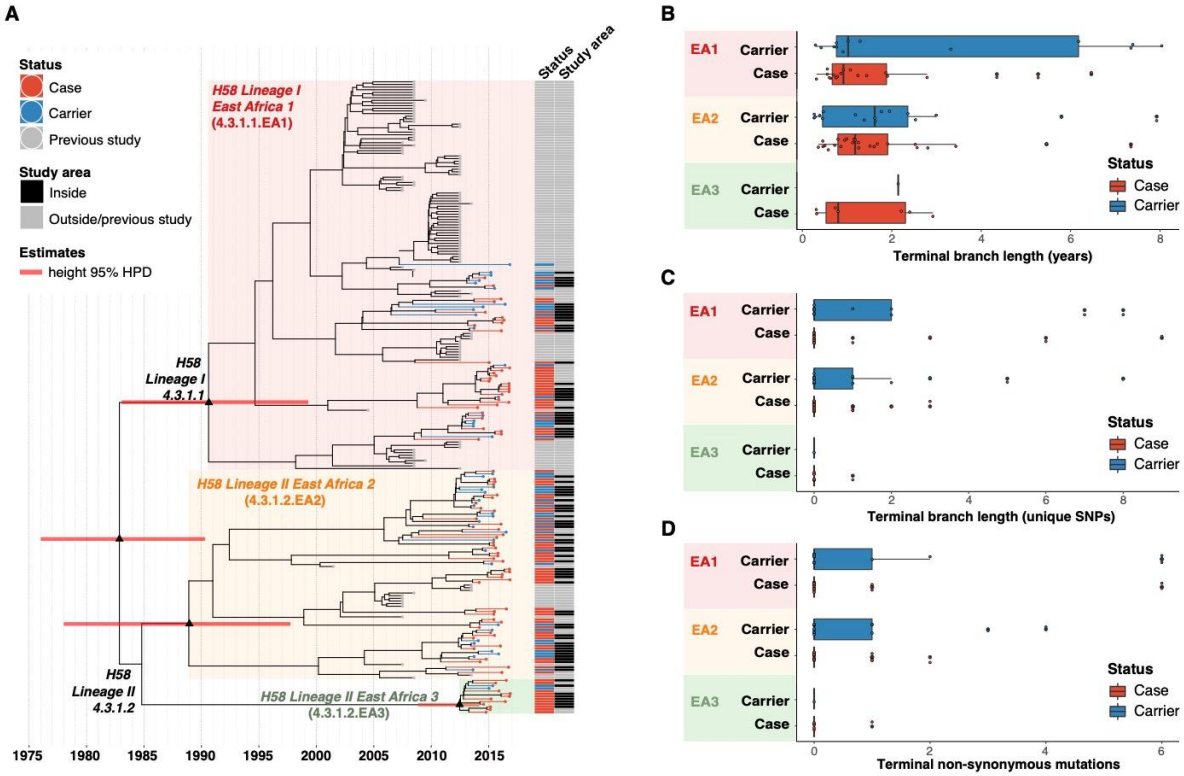
Supplementary Figure 2



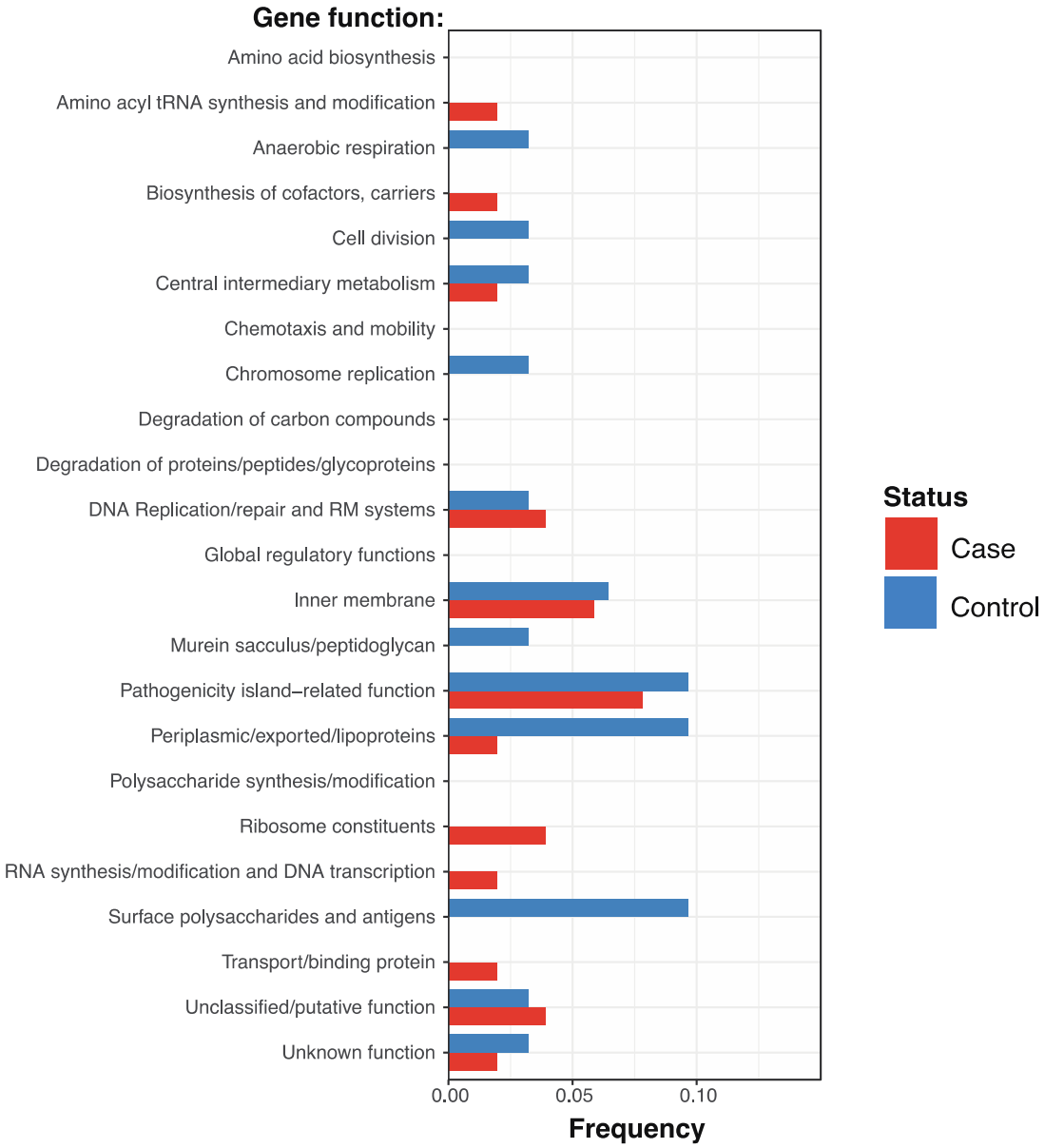
Supplementary Figure 3



Supplementary Figure 4



Supplementary Figure 5



Supplementary references:

1 Parkhill J, Dougan G, James KD, *et al.* Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18. *Nature* 2001; **413**: 848–52.

2 Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Meth* 2012; **9**: 357–9.

3 Li H, Handsaker B, Wysoker A, *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; **25**: 2078–9.

4 Wong VK, Baker S, Connor TR, *et al.* An extended genotyping framework for Salmonella enterica serovar Typhi, the cause of human typhoid. *Nat Commun* 2016; **7**: 12827.

5 Britto CD, Dyson ZA, Duchene S, *et al.* Laboratory and molecular surveillance of paediatric typhoidal Salmonella in Nepal: Antimicrobial resistance and implications for vaccine policy. *PLoS Negl Trop Dis* 2018; **12**: e0006408.

6 Rahman SIA, Dyson ZA, Klemm EJ, *et al.* Population structure and antimicrobial resistance patterns of Salmonella Typhi isolates in urban Dhaka, Bangladesh from 2004 to 2016. *PLoS Negl Trop Dis* 2020; **14**: e0008036.

7 Wong VK, Baker S, Pickard DJ, *et al.* Phylogeographical analysis of the dominant multidrug-resistant H58 clade of Salmonella Typhi identifies inter- and intracontinental transmission events. *Nat Genet* 2015; **47**: 632–9.

8 Pham Thanh D, Karkey A, Dongol S, *et al.* A novel ciprofloxacin-resistant subclade of H58 Salmonella Typhi is associated with fluoroquinolone treatment failure. *Elife* 2016; **5**: e14003.

9 Park SE, Pham DT, Boinett C, *et al.* The phylogeography and incidence of multi-drug resistant typhoid fever in sub-Saharan Africa. *Nat Commun* 2018; **9**: 5094.

10 Holt KE, Thomson NR, Wain J, *et al.* Pseudogene accumulation in the evolutionary histories of Salmonella enterica serovars Paratyphi A and Typhi. *BMC Genomics* 2009; **10**: 36.

11 Holt KE, Parkhill J, Mazzoni CJ, *et al.* High-throughput sequencing provides insights into genome variation and evolution in Salmonella Typhi. *Nat Genet* 2008; **40**: 987–93.

12 Ingle DJ, Nair S, Hartman H, *et al.* Informal genomic surveillance of regional distribution of Salmonella Typhi genotypes and antimicrobial resistance via returning travellers. *PLoS Negl Trop Dis* 2019; **13**: e0007620.

13 Croucher NJ, Page AJ, Connor TR, *et al.* Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* 2015; **43**: e15–5.

14 Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014; **30**: 1312–3.

15 Argimon S, Abudahab K, Goater RJE, *et al.* Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microbial Genomics* 2016; **2**: e000093.

16 Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. ggtree: an rpackage for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* 2016; **8**: 28–36.

17 Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 2004; **20**: 289–90.

18 Edwards DJ, Duchêne S, Pope B, Holt KE. SNPPar: identifying convergent evolution and other homoplasies from microbial whole-genome alignments. *bioRxiv* DOI:10.1101/2020.07.08.194480.

19 Thanh DP, Thieu NTV, Nguyen Thi Nguyen T, *et al.* Gallbladder carriage generates genetic variation and genome degradation in Salmonella Typhi. *PLoS Pathog* 2020; **16**: e1008998.

20 Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* 2016; **2**: vew007.

21 Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* 2018; **4**: vey016.

22 Duchene S, Holt KE, Weill F-X, *et al.* Genome-scale rates of evolutionary change in bacteria. *Microb Genom* 2016; **2**: e000094.

23 Firth C, Kitchen A, Shapiro B, Suchard MA, Holmes EC, Rambaut A. Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. *Mol Biol and Evol* 2010; **27**: 2038–51.

24 Kariuki S, Revathi G, Kiiru J, *et al.* Typhoid in Kenya is associated with a dominant multidrug-resistant Salmonella enterica serovar Typhi haplotype that is also widespread in Southeast Asia. *J Clin Microbiol* 2010; **48**: 2171–6.

25 Inouye M, Dashnow H, Raven L-A, *et al.* SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med* 2014; **6**: 90.

26 Carattoli A, Zankari E, García-Fernández A, *et al.* In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 2014; **58**: 3895–903.

27 Gupta SK, Padmanabhan BR, Diene SM, *et al.* ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob Agents Chemother* 2014; **58**: 212–20.

28 Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017; **13**: e1005595.

29 Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* 2015; **31**: 3350–2.

30 Hawkey J, Hamidian M, Wick RR, *et al.* ISMapper: identifying transposase insertion sites in bacterial genomes from short read sequence data. *BMC Genomics* 2015; **16**: 667–11.

31 Jaillard M, Lima L, Tournoud M, *et al.* A fast and agnostic method for bacterial genome-wide association studies: Bridging the gap between k-mers and genetic events. *PLoS Genet* 2018; **14**: e1007758.

32 Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci* 2003; **14**: 927–30.

33 Accou-Demartin M, Gaborieau V, Song Y, *et al.* Salmonella enterica Serotype Typhi with nonclassical quinolone resistance phenotype. *Emerg Infect Dis* 2011; **17**: 1091–4.

34 Feasey NA, Gaskell K, Wong V, *et al.* Rapid emergence of multidrug resistant, H58-lineage Salmonella typhi in Blantyre, Malawi. *PLoS Negl Trop Dis* 2015; **9**: e0003748.