The Molecular Epidemiology of Enteric Fever in South and Southeast Asia

Thesis

THE MOLECULAR EPIDEMIOLOGY OF ENTERIC FEVER IN SOUTH AND SOUTHEAST ASIA

Pham Thanh Duy

A thesis submitted to the Open University U.K

For the degree of Doctor of Philosophy in the field of Life Sciences

Oxford University Clinical Research Unit

Hospital for Tropical Diseases

Ho Chi Minh City, Viet Nam

April, 2018

Abstract

Typhoid fever is a life-threating systemic infection caused by *Salmonella enterica* sub-species *enterica* serovars Typhi (*S.* Typhi) and Paratyphi A (*S.* Paratyphi A). While the disease is mainly travel-associated in developed countries, it still causes significant burden in the poorest areas in developing countries where safe water and adequate sanitation and food hygiene remain limited. Typhoid management largely relies on antimicrobial therapy; however, antimicrobial resistance (AMR) in these causative pathogens has become a global threat, compromising the effectiveness of the treatment therapy and signifying the burden of this disease. Understanding the epidemiology of typhoid fever in different endemic settings as well as the impact of AMR on the disease outcome is crucial for disease control and management.

First, this thesis utilized whole genome sequences of *S.* Typhi combined with clinical data from a randomized controlled trial to investigate the impact of AMR and bacterial genotype on the disease outcome. A novel subclade of ciprofloxacin-resistant H58 *S.* Typhi associated with increased treatment failure was identified and these organisms were likely widespread in Indian subcontinent. Subsequently, this study combined bacterial genomics with conventional epidemiological tools to reveal the population structure and spatiotemporal dynamics of *S.* Paratyphi A isolates in Nepal. The Nepalese *S.* Paratyphi A population was highly dynamic with evidences of regular inter-country transmission, clonal expansion and replacement of distinct genotypes during the study period. A number of localized spatiotemporal clusters of

S. Paratyphi A cases were also identified. A molecular epidemiological investigation was also performed to provide insights into the AMR, epidemiological features and population structure and dynamics of *S.* Typhi in rural areas of Siem Reap, Cambodia. A substantial burden of pediatric typhoid fever was revealed and communes with high-risk of infection were identified. Multidrug resistant H58 *S.* Typhi with reduced susceptibility to fluoroquinolones was dominant in this setting. This study also investigated the phylogenetic relationship between acute and carriage *S.* Typhi isolates in Nepal and deciphered the genetic characteristics associated with carriage isolates. My study suggested that typhoid carriage was likely not an important source of new infections in endemic area.

Acknowledgements

First and foremost, I would like to express my deep gratitude to my PhD supervisor, Professor Stephen Baker, who has been supervising, teaching and guiding my scientific career since the first day I came into the science world. He has broadened my mind, helped me to expand my horizons, and constantly pushed me forward to become a better scientist. I have received timely and valuable encouragement, support, advice, and guidance from him during my PhD journey.

I have also received lots of guidance and support from my supervisor team, Dr Maia Rabaa, Professor Christopher Parry and Professor Gordon Dougan. Particularly, I would like to express my gratitude to Maia, who has taught me many useful things about phylogenetics and spent time chasing me up and editing my PhD thesis.

During my PhD, there are many colleagues and collaborators who have provided support for my research in terms of data analysis, project delivery, bioinformatics training, advice and guidance. I would like to extend my thanks to Dr Corrine Thompson, Dr Nhan Ho Thi (OUCRU-VN), Dr Abhilasha Karkey, Dr Sabina Dongol, Dr Buddah Basnyat (OUCRU-Nepal), Professor Nicholas Thompson (Wellcome Trust Sanger Institute), Dr Kathryn Holt (University of Melbourne) and Dr Paul Turner (COMRU).

I also really appreciate all the assistance from my colleagues in the enteric infections group (OUCRU-VN) during my work in the lab. Thank you for listening to me and supporting me with your encouraging messages throughout writing this thesis.

Finally, my deepest gratitude goes to my parents, who have wholeheartedly and unconditionally supported me. I would like to thank my dad for guiding me with his strength and diligence and my mom for her wisdom.

Abbreviations

µg	Microgram
µm	Micromole
µl	Microliter
AIDS	Acquired Immune Deficiency Syndrome
AHC	Angkor Hospital for Children
AMR	Antimicrobial resistance
ATCC	American Type Culture Collection
BiP	Biallelic Polymorphisms
bp	Base pairs
CDC	Centers for Disease Control and Prevention
CLSI	Clinical and Laboratory Standards Institute
cfu	Colony Forming Unit
DNA	Deoxyribose nucleic acid
ESBL	Extended Spectrum Beta Lactamase
FCT	Fever Clearance Time
FQ	Fluoroquinolone
GPS	Global Positioning System
GTR	General Time Reversible
HLA	Human Leukocyte Antigen
ICD	International Classification of Diseases
IQR	Interquartile range
IVI	International Vaccine Institute
Kb	Kilobase pairs
LPS	Lipopolysaccharides
LSMC	Lalitpur Sub-Metropolitan City
MDa	Megadalton
MDR	Multidrug resistant
MIC	Minimum Inhibitory Concentration
MLEE	Multilocus enzyme electrophoresis
MLST	Multilocus sequence typing
PCR	Polymerase chain reaction
SNP	Single Nucleotide Polymorphism
SPI	Salmonella Pathogenicity Island
WGS	Whole genome sequencing
WHO	World Health Organization

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The genus *Salmonella*

The genus *Salmonella* belongs to the family *Enterobacteriaceae*, which is a large and diverse group of bacteria found in soil, water, waste, plants, and the gastrointestinal tracts of animals [1]. This large family is composed of genetically and phenotypically closely related bacteria.

Salmonella are facultatively anaerobic, non-spore forming, Gram-negative bacilli, 2 to 5 μm long and 0.8 to 1.5 μm wide, usually motile by peritrichous flagella. The bacteria are catalase positive, oxidase negative, ferment glucose, mannitol and sorbitol to produce acid and gas, and can use citrate as a sole carbon source. They also do not hydrolyze urea or deaminate phenylalanine, but usually produce hydrogen sulfide on triple sugar iron agar. *Salmonella* yield negative Voges-Proskauer and positive methyl red tests and do not produce cytochrome oxide. Most *Salmonella* are positive for lysine decarboxylase and ornithine decarboxylase, except for *Salmonella* Typhi and *Salmonella* Paratyphi A. *Salmonella* is most closely related to *Escherichia coli* (*E. coli*); the geneses are thought to have shared a common ancestor more than 100 million years ago [2]. During their evolution, *E. coli* have become lactose positive while *Salmonella* are generally identified as being non-lactose fermenting [3]. *Salmonella* can be identified from biochemical tests based on lactose fermentation,

acid and gas production from glucose, mannitol, maltose, sorbitol, and the production of hydrogen sulfide [4].

1.2 *Salmonella* serology

The *Salmonella* are traditionally classified into different serotypes (also referred to as serovars) by the Kauffmann-White scheme, the system was established to categorise all identified *Salmonella* serotypes [5]. The documentation of this scheme is updated annually by the World Heath Organization's (WHO) Collaborating Centre for Reference and Research on *Salmonella* at the Pasteur Institute, Paris, France. Up until 2002, The U.S. Centers for Disease Control and Prevention (CDC) used a marginally different version of the scheme, but have since also adopted the Kauffmann-White Scheme [6].

Salmonella serotyping is based on antibody agglutination reactions with the bacterial surface structures, the O antigen (somatic) and H antigen (flagella). The O antigen is a polysaccharide polymerized from O subunits, with each subunit typically comprised of four to six sugars depending on the O antigen. It is the outermost component of the lipopolysaccharides (LPS) found in the outer membrane of Gram-negative bacteria. Variation in the sugar components of the subunit, the covalent bond between the sugars of the subunit or the linkage between O subunits that form the O antigen polymer results in O antigen diversity. O antigens are classified as primary O factors that define O serogroups (O groups) and additional O factors that are often variably present or expressed in some O groups. For instance, *S.* Typhi possess type 9 and 12

O antigens and belong to serogroup O:9 or D_1, whereas *S.* Paratyphi A carry type 1, 2 and 12 O antigens and belong to serogroup O:2 or A.

The H antigen is made up of protein subunits called flagellin and is the filamentous portion of the bacterial peritrichous flagellar. *Salmonella* is unique among enteric bacteria as it can express two forms (phases) of H antigens, which are encoded by different genes. The expression of these two genes is regulated so that only one flagellar antigen is expressed at a time in a single bacterial cell, a phenomenon known as phase variation [7]. Most *Salmonella* are diphasic and can express both phase 1 and phase 2 H antigen; however, some *Salmonella* serovars like *S.* Typhi, *S.* Paratyphi A and *S.* Enteritidis are naturally monophasic and can only express a single flagellin type due to disruption of the gene encoding phase 1 or phase 2 antigen.

Identification can also be performed by agglutination with another surface antigen, such as the virulence (Vi) polysaccharide capsular antigen that can be found on some *Salmonella* serovars, including *S.* Typhi, *S.* Paratyphi C and *S.* Dublin. The Vi capsule typically blankets the O antigen and blocks O agglutination but can be selectively removed by heat treatment prior to the O agglutination assay. Vi agglutination is used widely to identify *S.* Typhi isolates. As of 2002, the Kauffmann-White scheme was comprised of 2,541 serotypes [8]. This classification scheme is essential for laboratory and clinical identification of the *Salmonella* as well as for international communication between scientists, health officials, and the public. Although more advanced molecular typing techniques have been introduced, *Salmonella* serotyping

remains important in epidemiological surveillance and outbreak investigation and is widely used in most microbiology laboratories in the world [9,10].

1.3 *Salmonella* nomenclature

The nomenclature of *Salmonella* has been controversial as the original classification of the genus was not based on DNA relatedness; rather, names were originally given according to the particular disease caused, the animal from which the organism was isolated (e.g., *S.* Typhi causing human typhoid and *S.* Typhimurium causing typhoid-like in murine), or by the geographical location where the strain was first identified (e.g., *S.* Montevideo, *S.* Newport). Classification of *Salmonella* evolved over time beginning from the one serotype-one species concept that was originally proposed by Kauffmann in 1966 [11]. However if this concept were used today, it would result in the distinction of over 2,500 species.

Several classical *Salmonella* nomenclatural systems were subsequently proposed and inconsistently and confusingly divided the genus *Salmonella* into species, subspecies, subgenera, groups, subgroups and serotypes [12]. The turning point for *Salmonella* nomenclature occurred in the early 1970s when DNA-DNA hybridization experiments demonstrated that all *Salmonella* serotypes share greater than 85 percent of their genetic information and thus form a single species [13]. In 1986, Le Minor and Popoff proposed to designate this single species as *Salmonella enterica*, which became widely accepted [14]. Based on DNA relatedness, *Salmonella enterica* species were further divided into seven subgroups corresponding to seven subspecies (I, II, IIIa, IIIb, IV, V, VI), with subspecies V being the most distantly related of the

subgroups. In 1989, Reeves *et al.* published and classified subspecies V (*Salmonella enterica* subsp *bongori*) as a separate species (*Salmonella bongori*)[15].

According to the current *Salmonella* nomenclature used by CDC, the genus *Salmonella* comprises of two species, *Salmonella enterica* (*S. enterica*) and *Salmonella bongori*. *Salmonella enterica* is further divided into six subspecies, which are referred to by a Roman numeral and a name: I, *S. enterica* subsp. *enterica*; II, *S. enterica* subsp. *salamae*; IIIa, *S. enterica* subsp. *arizonae*; IIIb, *S. enterica* subsp. *diarizonae*; IV, *S. enterica* subsp. *houtenae*; and VI, *S. enterica* subsp. *indica* [12]. The majority (59%) of the *Salmonella* serotypes belong to subspecies I (*S. enterica* subsp. *enterica*) [16]. Within this subspecies I, the most common O serogroups are A, B, C1, C2, D and E, which cause approximately 99% of *Salmonella* infections in humans and warm-blooded animals [16]. Serotypes of other subspecies and *Salmonella bongori* are usually found in cold-blooded animals and the environment, and are rarely associated with disease in humans [17].

1.4 *Salmonella*, the host and disease

Salmonella are widely distributed in nature and can colonize the gastrointestinal tracts of both cold- and warm-blooded animals, including humans. Even though there are more than 2,500 *Salmonella* serovars, most serotypes are not pathogenic in their natural hosts and the majority of infections in humans and animals are caused by a small number of serotypes within subspecies I [18]. The manifestation of disease largely depends on both host susceptibility and the infecting *Salmonella enterica* serotype, and is generally associated with one of the four major syndromes: typhoid fever,

enterocolitis/diarrhea, bacteremia and chronic asymptomatic carriage [19]. For instance, in humans, most *Salmonella* serotypes are associated with acute and self-limiting gastroenteritis, whereas other serotypes (e.g, *S.* Typhi, *S.* Paratyphi and *S.* Sendai) cause enteric fever and a few serotypes (e.g, *S.* Choleraesuis and *S.* Dublin) are more likely to cause bacteremia than diarrhea.

Regarding host adaptability, the majority of *Salmonella* serotypes (e.g, *S.* Typhimurium, *S.* Enteritidis) have a wide host range including animals and humans; however, other serotypes are largely adapted to specific animals and are infrequently found in humans (e.g, *S.* Dublin in cattle, *S.* Gallinarum in poultry, *S.* Abortusequi in horses, *S.* Abortusovis in sheep and *S.* Choleraesuis in pigs). Furthermore, a small number of serotypes such as *S.* Typhi, *S.* Paratyphi A, B, C and *S.* Sendai are fully adapted to humans and higher primates and are unable to cause disease in other hosts [20]. Host adaptation in *Salmonella* can be defined as the ability of a pathogen to circulate and cause disease in a particular host population [21]. Host-adapted *Salmonella* serovars (e.g., *S.* Typhi, *S.* Choleraesuis) tend to be more virulent and cause systemic infections with a higher mortality rates in their hosts compared to broad-host-range serovars (e.g., *S.* Typhimurium or *S.* Enteritidis), which are often associated with non-invasive infections [22].

1.5 Typhoid fever

1.5.1 Clinical features of typhoid fever

Typhoid fever is a life-threatening systemic infection predominantly caused by *Salmonella enterica* subsp *enterica* serovar Typhi (*S.* Typhi) and *Salmonella enterica*

subsp *enterica* serovar Paratyphi A (*S*. Paratyphi A). The onset of symptoms is marked with prolonged fever, headache, malaise, anorexia, nausea, dry cough and disturbances of bowel function (constipation or diarrhea), which typically occurs after the end of first week. A coated tongue, hepatomegaly and splenomegaly are also common. The fever is low grade during the first week but rises gradually; by the second week, the temperature increases, reaching a plateau of 39 to 40^{o}C. Rose spots on the chest, abdomen and back are reported in 5 to 30 percent of cases. The clinical manifestations and severity of typhoid fever are highly variable due to factors such as duration of illness before appropriate therapy, virulence and antimicrobial resistance of the causative agents, selection of antimicrobial therapy, patient age, previous exposure or vaccination, inoculum size, host factors (HLA type, AIDS or other immune suppression) and antacid consumption [23–32]. Depending on the clinical setting and quality of available heathcare, serious complications can occur in 10 to 15 percent of typhoid patients, particularly in those who have been sick for more than two weeks. A wide range of complications are described, of which gastrointestinal bleeding, intestinal perforation, and typhoid encephalopathy are the most common and likely associated with risk of death [33,34].

In the pre-antimicrobial era, case fatality rates were as high as 10 to 30 percent [35,36]. Currently, with effective antimicrobial treatment, the average case fatality rate is usually less than 1 percent. However, case fatality rates vary significantly among different regions of the world, ranging from less than 2 percent in Pakistan and Vietnam to 30 to 50 percent in some areas of Indonesia and Papua New Guinea [37–40]. Poor outcomes are highest among children less than one year of age and the elderly,

and often result from delayed treatment with effective antimicrobials [41]. Relapse can occur 1 to 3 weeks after the patient recovers from the first episode, with milder symptoms than those experienced during the initial illness [34]. The relapse rate is 5 to 10 percent in untreated cases but still occurs after antimicrobial treatment, especially in those treated with first line drugs (chloramphenicol, trimethoprim-sulfamethoxazole, and ampicillin). Relapse rates in patients treated with newer antimicrobials including fluoroquinolones (1.5%) or broad-spectrum cephalosporins (5%) are normally lower than those treated with first line drugs [42–48]. Typically, relapses are caused by the same isolate as the original episode and can be distinguished from reinfection by molecular typing [47]. Clinical features of *S.* Paratyphi A infections are commonly considered milder, with fewer complications than infections caused by *S.* Typhi. However, in the largest comparison to date there were no significant differences between clinical presentation, duration or outcome of typhoid caused by these organisms [15].

1.5.2 Pathogenesis of typhoid fever

As *S.* Typhi and *S.* Paratyphi A only cause disease in humans, pathogenesis studies of typhoid fever are hindered by the lack of a suitable animal model. Much of the understanding of pathogenic features of typhoid fever has arisen from *Salmonella enterica* serovar Typhimurium infection in a susceptible murine model that is thought to mimic human typhoid. The infectious dose has been determined in human challenge studies and varies between 1,000 and 1 million organisms, depending on the individual and the settings [26,49–52]. After ingestion, the bacteria have to survive the gastric acid before reaching the small intestine. In the small intestine, bacteria can

adhere to the intestinal mucosa and subsequently invade the gut mucosa, probably through specialized epithelial cells overlaying the Peyer's patches known as the M cells [53]. After penetration, the invading bacteria translocate to the intestinal lymphoid follicles and are drained into mesenteric lymph nodes and the thoracic duct, and eventually into the bloodstream. In this primary bacteremia (within 24 hours of their ingestion), the organisms reach the liver, spleen, bone marrow and other parts of the mononuclear phagocyte systems where they can survive and multiply within the cells of monocytic lineage [54]. After an incubation period of 7-14 days (depending on bacterial load, virulence and host response), bacteria are shed back into the bloodstream causing secondary, sustained bacteremia and marking the onset of the clinical symptoms. The most common sites of secondary infection are the spleen, liver, bone marrow, gallbladder and Peyer's patches of the terminal ileum. Gallbladder colonisation occurs either directly from blood or by retrograde spread from the bile. Organisms are shed via the bile duct into the small intestine resulting in fecal shedding.

There are some notable differences in the pathogenesis of typhoid fever compared to infections caused by nontyphoidal *Salmonella* serovars. *S.* Typhi can translocate across the intestinal mucosa during an early phase of infection without causing any physical cellular damage and therefore without triggering a rapid acute inflammatory response as normally seen in gastrointestinal infections caused by nontyphoidal *Salmonella*. After infection, the incubation period may not always be followed by clinical symptoms [55]. Furthermore, *S.* Typhi can survive and multiply in monocytes, which is essential for dissemination and persistence within the host, whereas

nontyphoidal *Salmonella* serovars, like Typhimurium, are effectively cleared by human monocytes [56]. It is also noteworthy that the number of recoverable *S.* Typhi from patients with typhoid fever is low, with a median of 1 cfu/ml of blood and 10 cfu/ml of bone marrow, which has a negative impact on diagnostics [57–59].

1.5.3 Typhoid diagnosis

The clinical diagnosis of typhoid fever is challenging in endemic regions as it is difficult to distinguish typhoid from other acute febrile illnesses such as malaria or dengue. Currently, blood culture followed by microbiological identification is still the gold standard diagnostic method, with a sensitivity of approximately 80 percent [33]. Even though blood culture is the most reliable method and can be standardized, it remains costly and requires specialist facilities and personnel and therefore is generally only utilised in major hospitals in developing countries. The sensitivity of blood culture largely relies on the volume of blood taken from typhoid patients due to a low number of circulating organisms. To achieve the highest recovery rates, 10-15 ml of blood from school children and adults and 2-4 ml from toddlers and preschool children are required. In some areas, it can be challenging to obtain such large volumes of blood, especially in children, which undermines diagnosis. Bone marrow culture is more sensitive than blood culture because the number of organisms in bone marrow is comparatively high [58,59]. The culturing of bone marrow can be valuable for patients who have been treated with antimicrobials regardless of duration of illness [60–62]. However, this is an invasive procedure and thus not widely accepted, particularly in children, and is rarely performed outside of specialist hospitals. Stool culture has a positive rate of 30 percent in acute typhoid patients and its sensitivity depends on the

amount of feces cultured as well as the duration of the illness [61]. Stool culture can be used to detect typhoid carriers but requires that multiple samples be examined, as shedding is sporadic and may persist at low levels. Instead, a Vi agglutination assay has been used to identify typhoid carriers because these individuals may produce high levels of Vi antibodies over a longer period compared to acutely infected patients [63,64].

The Felix-Widal test was developed in 1896 and is the first serological method used to diagnose typhoid fever. This method is based on the measurement of agglutinating antibody levels against the O and H antigens of *S.* Typhi. The sensitivity and specificity of this test are moderate as *S.* Typhi share O and H antigens with other *Salmonella* serovars and also have cross-reacting epitopes with members of Enterobacteriaceae, which can result in a high false positive rate [65–68]. Additionally, typhoid patients might not develop detectable antibodies or show demonstrable increases in antibody titers. This problem also presents in commercial serological tests such as Tubex and Typhidot, which demonstrated moderate sensitivity and specificity for the diagnosis of typhoid fever in Papua New Guinea, India and Bangladesh. In these studies, the sensitivity and specificity of these commercial kits varied from 51.1-60 % and 58-88.3 %, respectively (Tubex) and 56-70.0 % and 54-88 %, respectively (Typhidot) [69–71]. Despite these limitations, the use of the Felix-Widal test with suitable local cut-off values for positive diagnosis can be helpful in areas where access to alternative, more expensive methods is limited [72–74]. PCR-based assays have also been developed for typhoid diagnosis. However, they are either unreliable or have poor sensitivity when performed on DNA extracted from blood samples [57,75]. As a result, these methods are not widely used and also considered impractical in many typhoid

endemic areas. Serological assays based on novel *S.* Typhi-specific immunogenic proteins are under development which could potentially be converted into cheap and rapid diagnostic kits [76,77]. Alternative approaches based on host-specific biomarkers such as metabolomics are also promising and warrant further investigation [78].

1.5.4 Typhoid treatment

In endemic areas, up to 90 percent of typhoid cases are managed at home with oral antimicrobials, supportive care and regular follow-up for complications [31,79]. For hospitalized patients with severe symptoms, effective antimicrobials, good nursing care, maintenance of appropriate nutrition and hydration, and timely recognition and treatment are crucial to prevent serious complications and deaths. Appropriate antimicrobial therapy is critical to cure typhoid and avoid complications. The selection of antimicrobial drugs depends on a number of important criteria such as availability, efficacy and cost. An ideal drug should have fast time to defervescence and clinical improvement, render blood and stool cultures sterile during and after treatment, prevent relapse, and be available through oral and intravenous routes for both adults and children with low cost and minimal adverse effects [80].

Prior to the 1990s, chloramphenicol, ampicillin and trimethoprim-sulphamethoxazole were the first drugs of choice for typhoid treatment as these drugs are inexpensive, widely available and rarely associated with side effects. However, the emergence and global dissemination of plasmid-mediated multidrug resistant *S.* Typhi in the 1990s rendered these drugs ineffective and resulted in the widespread use of fluoroquinolones as the treatment of choice [33]. Data from published clinical trials

demonstrated that fluoroquinolones were the most effective drugs for typhoid treatment. They were rapidly effective even with short courses of treatment (3-7 days), clearing fever and symptoms in 3 to 5 days with a cure rate exceeding 96 percent and very low rates of post-treatment carriage (less than 2 percent) [81–86]. However, nalidixic acid resistant *S.* Typhi and *S.* Paratyphi A exhibiting reduced susceptibility to fluoroquinolones have since emerged and become endemic in many South and Southeast Asian countries, and are associated with increased rates of fluoroquinolone treatment failure [87–90]. For patients infected with nalidixic acid resistant organisms, treatment with the maximal recommended dose of fluoroquinolones should be given for a minimum of 10 to 14 days and patients should be followed up carefully to assess the shedding of bacteria in their stools. Third generation cephalosporins (ceftriaxone, cefixime, cefotaxime, and cefoperazone) and the macrolide, azithromycin, have been successfully used for typhoid treatment and can be used as effective alternative drugs for treating *S.* Typhi and *S.* Paratyphi A with reduced susceptibility to fluoroquinolones [87,91–100]. In areas where fluoroquinolones are not available or unaffordable, first line drugs (chloramphenicol, ampicillin and trimethoprim-sulphamethoxazole) remain appropriate for the treatment of typhoid if bacteria are still susceptible to these drugs. For severe typhoid, parenteral fluoroquinolones are often the drugs of choice and are given for a minimum of 10 days [33]. A high dose of intravenous dexamethasone and antimicrobials should be given promptly to adults and children with severe symptoms to reduce mortality [37]. Relapse should be treated in the same manner as primary infection. The majority of chronic carriers can be cured with a prolonged course of antimicrobials [101–103].

1.6 The epidemiology of typhoid fever

1.6.1 Epidemiological features of typhoid fever

S. Typhi and *S.* Paratyphi A are human-restricted pathogens and humans are the only known natural hosts and reservoir of infection [34]. The disease is transmitted via the fecal-oral route. The most common mode of typhoid transmission is by ingestion of water or food contaminated with human feces. Direct person-to-person transmission through contact with patients or chronic carriers who are shedding the organisms is uncommon [33]. In endemic areas, water contaminated with human feces and food contaminated with contaminated water or by food handlers who are typhoid carriers are the main sources of infection [104]. The incubation period and attack rate of typhoid fever are affected by the inoculum size and the vehicle of transmission. Waterborne transmission usually requires smaller inoculum, whereas foodborne transmission is associated with a larger inoculum and a higher attack rate over short periods [105,106]. Bacteria causing typhoid fever can survive for weeks or months in the environment, including in seawater, sewage, pond, stream and lake water [107,108]. Shellfish grown in polluted water, fruits and vegetables fertilized with sewage, milk and milk products have been documented as potential sources of infection [109]. Often, large typhoid outbreaks occur when the source of drinking water serving large populations is contaminated or the water supply is disrupted [110,111]. In developed countries, typhoid occurs sporadically and is mainly associated with travellers returning from endemic areas [112–118].

1.6.2 Typhoid carriage

Asymptomatic carriage and the shedding of bacteria in feces can occur after patients recover from acute infection. Carriage can be divided into different periods depending on the duration of shedding: convalescent (three weeks to three months), temporary (three to twelve month) and chronic (more than one year) [34]. In endemic regions, the chronic carriage rate is estimated to be 2-5 percent of the population, most of which are asymptomatic; up to 25 percent of chronic carriers have no clinical history of infection [34,119]. The relative importance of short-term and convalescent fecal carriers versus chronic carriers in the transmission dynamics of typhoid in endemic regions is largely unknown, probably due to a lack of follow-up and the absence of a robust method for detecting *S.* Typhi carriers. Risk factors associated with persistent carriage are not extensively studied, as this population is very challenging to identify prospectively. Previous studies have found that risk of becoming a chronic carrier following acute infection increases with age, and is higher in women and patients with cholelithiasis and cholecystitis [120,121]. Data from a murine model and electron micrographic observation of gallstones retrieved from *S.* Typhi human carriers have suggested that colonization of the gallbladder epithelial cells and biofilm formation on the gallstones are the primary mechanisms by which *Salmonella* survives and persists in the gallbladder environment [122–126].

Chronic carriage and fecal shedding have been long considered as central dogma determining the transmission and persistence of typhoid fever, as chronic carriers intermittently shed the bacteria into the local environment and may spread the disease in the community; additionally, these people can act as reservoirs for maintaining

specific genotypes [127,128]. However, epidemiological studies have shown that direct transmission between household members primarily occurs through close contact with convalescent carriers rather than chronically infected individuals [129,130]. Furthermore, recent advanced molecular epidemiological investigations in endemic regions such as Vietnam, Indonesia, and Nepal have demonstrated that acute typhoid is generally caused by a wide diversity of genotypes rather than singular local genotypes [131–134]. Therefore, environmental transmission pathway appears to be most important transmission route rather than direct person-to-person contact in endemic regions Conversely, in areas where public health interventions have been successful in reducing typhoid burden, chronic carriage might play an important role in maintaining the disease in the population, probably via foodborne transmission from asymptomatic chronic carriers who are food preparers or handlers as reported during outbreaks of typhoid fever in the United States [135]. As a result, the detection and treatment of chronic carriers is thought to be essential for eliminating typhoid from a population.

1.6.3 The global burden of typhoid fever

Typhoid fever remains a significant public health problem in many low and middle-income countries where there may be a lack of safe water and adequate sanitation. Global estimates suggested that 21.7 million new cases and 217 000 deaths due to typhoid fever caused by *S.* Typhi occurred in the year 2000 alone, whilst *S.* Paratyphi A also caused 5.4 million cases with an unknown case fatality rate [118]. The majority of illness occurred among infants, children and adolescents, with south-central and south-eastern Asia exhibiting the highest disease burden (>100 cases per 100 000 per

year) [118]. More recently, Buckle and colleagues estimated that 13.5-26.9 million episodes of typhoid occurred worldwide in 2010 [136]. In the context of informing vaccine policy, Vittal-Mogasale and colleagues re-estimated the burden of typhoid fever with a focus on low- and middle-income countries. After adjusting for blood culture sensitivity and water-related risk, they estimated that 11.9 million typhoid fever illnesses and 129 000 deaths occurred in low and middle-income countries during 2010, and also suggested that typhoid burden is higher in Africa than previously thought [137]. Between 2010 and 2014, multi-country population-based standardized surveillance for invasive *Salmonella* infection was conducted in 13 sentinel sites across 10 countries in sub-Saharan Africa and identified a number of sites with remarkably high burden of typhoid fever, especially in children less than 15 years of age (overall adjusted incidence rate >100 per 100 000 person-years of observation) [138]. Additionally, data from this study also demonstrated that many rural populations exhibited similar or even higher typhoid incidences than urban populations.

Simultaneously, typhoid fever caused by *S.* Paratyphi A has emerged at an unprecedented rate in many Asian countries, including Nepal, China, Pakistan and India [139]. In some regions, the isolation rate of *S.* Paratyphi A from typhoid patients is similar or even higher than that of *S.* Typhi [140,141]. The rapid emergence of *S.* Paratyphi A infections poses a significant public health concern as *S.* Paratyphi A display differences in epidemiology compared to *S.* Typhi and, thus, effective control and preventive measures for *S.* Typhi may not protect against *S.* Paratyphi A [129,142]. Currently, there is no vaccine for *S.* Paratyphi A and the licensed typhoid vaccines do

not provide protection against *S.* Paratyphi A infections. Additionally, it has been reported that typhoid fever caused by *S.* Paratyphi A increased after typhoid vaccine implementation in Guangxi, China, which suggests that the role of *S.* Paratyphi A in the epidemiology of typhoid fever may impede the progress in typhoid management, and calls into question the impact of vaccines in areas where these two pathogens co-circulate [143].

1.6.4 Typhoid control and prevention

As contaminated water and food are important vehicles for typhoid transmission, improvements in water, sanitation, and hygienic food preparation represent the ultimate solutions for reducing the burden of disease. Historical surveillance data suggest that rate of typhoid fever in Western Europe and North America substantially declined in parallel with the introduction of water treatment, pasteurization of dairy products and exclusion of human feces from food production [144]. Recent reductions in typhoid fever incidence have also been reported in Latin America and some Asian countries, in parallel with water and sanitation improvements and economic transition [118,145]. However, such structural improvements remain challenging in low and middle-income countries given the huge economic costs and long timelines that are often required to improve water quality and sanitation. In view of the continued burden of typhoid fever and the increasing antimicrobial resistance of the organisms, in 2008 the World Health Organization (WHO) recommended the programmatic use of licensed typhoid vaccines for endemic and epidemic disease control [146]. There are currently two internationally licensed typhoid vaccines, the parenteral Vi-based polysaccharide vaccine and the live oral Ty21a vaccine, both of which have been proven to be safe

and efficacious in children aged > 2 years. Many Asian countries such as Vietnam and China have successfully used the Vi polysaccharide vaccine to reduce the burden of typhoid fever, and several other countries also have national policies targeting high-risk groups such as food handlers [147]. Additionally, a mass vaccination campaign with the Vi polysaccharide vaccine was also conducted in Fiji in cyclone-affected and high-risk areas [148]. Despite the recommendation of WHO, strong evidence for the disease burden and the low cost, availability, and efficacy of typhoid vaccines, neither of these vaccines has been widely used in endemic low-resource settings. This might be due to a lack of epidemiological data in endemic settings as well as the short duration of vaccine-induced protection; further, none of these vaccines are licensed for children less than 2 years of age [149]. A number of typhoid conjugate vaccines are under development, including Vi-rEPA (US National Institutes of Health), Typbar-TCV (Bharat Biotech International Ltd) and Vi-CRM197 (Scla Behring Vaccines for Global Health) [150–152]. Indeed, Typbar-TCV has been prequalified by the WHO and is currently being evaluated in Nepal, Bangladesh and Malawi. These new generation typhoid vaccines are expected to provide higher efficacy and longer duration of protection, and can be given to infants and young children. Efforts are also being made to develop a vaccine for *S*. Paratyphi A as well as a bivalent vaccine that can protect against both *S*. Typhi and *S*. Paratyphi A infections [147].

1.7 Antimicrobial resistance in Typhoidal *Salmonella*

1.7.1 Global dissemination of multidrug resistant Typhoidal *Salmonella*

Antimicrobial therapy is crucial for the management of typhoid fever and for reducing mortality. In the pre-antibiotic era, reported case-fatality ratios ranged from 10 to 37

percent [153]. In 1948, chloramphenicol was first used to treat typhoid fever and became the standard drug of choice [154]. Fifteen years after the introduction of chloramphenicol, the mortality rate decreased from 12 to 1 percent [155]. This drug was the mainstay for typhoid treatment until mid-1970, when antimicrobial resistant organisms started to emerge and subsequently caused epidemics in Mexico, India, Vietnam and South Korea [156–161]. The outbreak strains carried a transferable resistant factor (R factor) located on a self-transmissible plasmid of the HI1 incompatibility group (IncHI1), which was responsible for resistance to chloramphenicol and also to streptomycin, sulphonamides and tetracyclines [162]. The spread of chloramphenicol resistant *S.* Typhi lead to the increased use of the other two first line drugs, amoxicillin and trimethoprim-sulfamethoxazole. However, toward the end of the 1980s and the 1990s, *S.* Typhi developed multidrug resistance (MDR) to all first line drugs of treatment, including chloramphenicol, trimethoprim-sulfamethoxazole and ampicillin [163]. Outbreaks of multidrug resistant *S.* Typhi infections were subsequently reported in many countries in South Asia, Southeast Asia, Middle East and Africa [40,164–171]. These MDR *S.* Typhi carried self-transmissible plasmids of the HI1 incompatibility type with typical size from 100 to 120 MDa [166,167,172–174]. Recently, there have also been reports about chromosome-mediated MDR in *S.* Typhi isolates in Bangladesh, India, Pakistan, Iraq, and Fiji [175,176]. MDR *S.* Typhi are currently considered to be endemic in many areas of South, Southeast Asia and Africa. Compared to *S.* Typhi, *S.* Paratyphi A have predominantly been found to be more susceptible to antimicrobials; however, MDR *S.* Paratyphi A isolates have also commonly reported since the 1990s in many settings in Asia [177–179]. These MDR *S.* Paratyphi A isolates also harbored IncHI1 plasmids that share a common backbone

with other IncHI1 plasmids found in *S.* Typhi and acquired similar mobile elements, conferring a multidrug resistant phenotype [180].

1.7.2 Emergence of fluoroquinolone resistance

The global distribution and dominance of MDR *S.* Typhi led to the extensive use of fluoroquinolones (ciprofloxacin, ofloxacin) as alternative drugs of choice in the 1990s. Again, the rampant use of ciprofloxacin not only for typhoid but also for other infections resulted in the emergence of *S.* Typhi isolates that were resistant to nalidixic acid (a quinolone) and reduced susceptibility to ciprofloxacin with MICs ranging from 0.125 to 1 μg/ml, ten times higher than the usual values for fully susceptible strains. Since the early 1990s, *S.* Typhi isolates with reduced susceptibility to ciprofloxacin have become a major problem in many developing countries in Asia [89,164,181–183]. Further, patients infected with *S.* Typhi with reduced susceptibility to ciprofloxacin are prone to exhibiting prolonged fever clearance times and increased rates of treatment failure [89,90,183]. In 1997, a massive outbreak caused by such isolates occurred in Tajikistan, infecting 8000 people and causing 150 deaths [111]. The first highly ciprofloxacin resistant *S.* Typhi isolate was identified in 2005 in India with an MIC of ciprofloxacin > 32 μg/ml [184]. Subsequent reports of increased incidence of infections caused by fully ciprofloxacin resistant *S.* Typhi later emerged in South Asia and travellers from developed countries who returned from this region [185–192]. It is unclear if the emergence of highly ciprofloxacin resistant *S.* Typhi isolates was due to clonal expansion of a particular genotype or independent acquisition of resistance in multiple genotypes. Similarly, *S.* Paratyphi A isolates with reduced susceptibility to ciprofloxacin also emerged and subsequently disseminated throughout South Asia

over the same period and were associated with infections among travellers returning home from travel to these endemic areas [179,193–199]. Alarmingly, *S.* Paratyphi A have shown a tendency to be more resistant to ciprofloxacin compared to *S.* Typhi [197,200]. Highly ciprofloxacin resistant *S.* Paratyphi A isolates have also been sporadically reported in India and Japan [201,202]. Together with the widespread use of ciprofloxacin, there have been reports of a decline in chloramphenicol resistance and MDR *S.* Typhi isolates in some areas; however, the reintroduction of first line antibiotics is likely to result in the reemergence of MDR *S.* Typhi [203–207].

Fluoroquinolone resistance is chromosome-mediated via point mutations occurring in various chromosomal genes (*gyrA, gyrB, parC* and *parE*). These single point mutations alter the enzymes, DNA gyrase and topoisomerase IV, all of which are targets for the quinolones. A single point mutation in the quinolone resistance determining region in the *gyrA* gene (frequently at codon 83 or codon 87) can result in decreased susceptibility to fluoroquinolones [181,183]. Double mutation in the *gyrA* gene (two single point mutations at codon 83 and codon 87) results in reduced susceptibility to ciprofloxacin, whilst an additional mutation in the *parC* gene (typically at codon 80) is required for high level resistance to ciprofloxacin [196,208,209]. Plasmid-mediated quinolone resistance was also reported in the late 1990s and was predominantly associated with qnr proteins that can protect DNA gyrase from ciprofloxacin and also by aminoglycoside-modifying enzyme (aac(6')-Ib-cr), which acetylates ciprofloxacin [210]. Resistance genes associated with these plasmids, such as *qnrS1, qnrB, aac(6')-Ib-cr,* have been sporadically detected in *S.* Typhi isolates from India and from travellers returning from South Asian countries [211,212].

1.7.3 Resistance to third generation cephalosporins and macrolides

In areas where MDR *S.* Typhi and *S.* Typhi with reduced susceptibility to ciprofloxacin are highly prevalent, azithromycin and third generation cephalosporins (ceftriaxone, cefixime, cefotaxime and cefoperazone) tend to be used as alternative drugs of treatment for uncomplicated typhoid fever. These drugs have become important for typhoid treatment as they have been proven effective during clinical trials and the prevalence of resistance to these agents is very low [45,93,213]. However, since 2008, *S.* Typhi isolates gaining resistance to third generation cephalosporins have been reported in many countries, including India, Bangladesh, Pakistan, the Philippines, Germany, Congo, and Guatemala [214–220]. There have been a number of extended-spectrum beta-lactamase (ESBL) enzymes identified among those isolates, such as SHV-12, CTX-M types, and AmpC of the ACC-1 type. In 2013, an MDR *S.* Paratyphi A isolate harboring a CTX-M-15 beta-lactamase was also reported from a Japanese traveler returning from India [221]. Recently, outbreaks of typhoid fever caused by ESBL-producing *S.* Typhi have drawn significant public and scientific attention in India and Pakistan [222]. For azithromycin, *S.* Typhi isolates frequently have an MIC ranging from 4 to 16 μg/ml and are found to respond well to a short-course of azithromycin treatment [223]. The high effectiveness of azithromycin in clearing the infection is probably due to azithromycin's remarkable property of high intracellular concentration (50 times more than in blood) [224]. However, treatment failures even with low azithromycin MIC of *S.* Typhi and *S.* Paratyphi A as well as high MIC of azithromycin (64 μg/ml) *S.* Paratyphi A have been sporadically reported in India and

travellers returning from Pakistan and India [225–227]. The mechanism for azithromycin resistance in typhoidal *Salmonella* is not well described.

1.8 The convergent evolution of *Salmonella* Typhi and *Salmonella* Paratyphi A

While most of the *Salmonella enterica* serovars can infect a broad range of host species and are usually associated with self-limiting gastrointestinal infections, Typhi and Paratyphi A are human-restricted serovars and cause life-threatening systemic disease [228]. There is a high level of similarity in pathological and epidemiological characteristics between these two typhoidal serovars. For instance, Typhi and Paratyphi A display similar pathogenic features during the course of infection, clearly distinguishing them from non-typhoidal S*almonella enterica* serovars [153]. This likely results in the indistinguishable clinical features and severity of infections caused by these two serovars as shown in the largest comparison up to date [197]. Further, both can establish chronic infections in the human gall bladder, which could be a strategy to increase their transmissibility [130]. Regarding their transmission patterns, Typhi and Paratyphi A generally infect via similar transmission routes, namely contaminated water or food, and both can also be transmitted directly from person to person [229].

In the early 1990s, the population structure of *Salmonella enterica* was described based on multilocus enzyme electrophoresis (MLEE), through which Typhi and Paratyphi A were shown to belong to distinct lineages, suggesting that these two *Salmonella* serovars have different evolutionary histories and may have evolved independently to become human-adapted with identical disease phenotype [15,230]. Multiple complete genomes of Typhi and Paratyphi A were subsequently generated,

followed by genome-wide comparisons between these two serovars as well as to other host-generalized serovars like Typhimurium; these studies provided unprecedented insights into the evolutionary processes of these organisms. Typhi and Paratyphi A are much more closely related to each other at the DNA level than Typhimurium [231]. This is a consequence of extensive homologous recombination within a quarter of their genomes, which results in similar gene content and much lower nucleotide divergence (0.18%) compared to the rest of their genomes (1.2%) [232]. These recombination events might play a role in the host restriction/pathogenesis of these two serovars, although this remains unproven. Further, the genomes of Typhi and Paratyphi A contain an exceptionally high level of pseudogenes (about 4% of coding sequences in each genome), much higher than the pseudogene contents of Typhimurium (0.9%) and *E.* coli (0.7%) [231,233–235]. Pseudogenes are coding sequences that are putatively inactivated by mutations (such as nonsense substitution, frameshifts, gene truncation by deletion or rearrangement), causing the loss of their gene functions. Pseudogenes are not unique to the *Salmonella*, and are abundant in other host-adapted bacteria such as *Shigella flexneri, Yersinia pestis, Rickettsia prowazekii,* and *Mycobacterium leprae* [236–239] . Even though there are only a limited number of pseudogenes shared between Typhi and Paratyphi A, these genes are known to be involved in adaptation to their comparable niches and the interaction between *Salmonella* and host. For examples, loss of gene functions related to intestinal colonization and persistence (typhoidal *Salmonella* are less efficient in intestinal colonization and more favorable of systemic sites [55]); pathogenicity (disruption in genes encoding for *Salmonella*-translocating effector proteins of *Salmonella* pathogenicity island I and II); chemotaxis receptor; iron metabolism and

surface-exposed proteins [231,233,240]. Similar loss of protein functions also occurs in genes belonging to the same pathway, which often results in phenotypic convergence. For instance, disruption of fimbrial clusters in both Typhi and Paratyphi A can have important implications for host interaction or disruption in the *hin* gene but through different mechanisms, resulting in the loss of phase variation in both Typhi and Paratyphi A (both serovars are monophasic for phase 1 flagella whilst most *Salmonella enterica* subspecies 1 have a switching mechanism leading to diphasic flagella) [231]. Gene degradation is likely the most important evolutionary force that alters the pathogenesis and narrows the host range of both Typhi and Paratyphi A [233,234,240]. There are also some specific gene clusters gained by lateral gene transfer that are present in both Typhi and Paratyphi A; however, most of their functions are unknown and unlikely to have a role in phenotypic differences from other *Salmonella enterica* serovars [231].

1.9 Genomics and molecular epidemiology of typhoid fever

1.9.1 Genomics and its role in molecular epidemiological studies in Asia and Africa

In the pre-genomic era, it was very challenging to study the population structures and genetic characteristics of both *S.* Typhi and *S.* Paratyphi A as they showed very little genetic variation by conventional typing methods such as MLEE, PFGE and MLST [15,241–243]. At the turn of the 21st century, the first complete genome of *S.* Typhi multidrug resistant strain, CT18, was published, with the first *S.* Paratyphi A genome strain, ATCC9150, released a few years later [231,233]. The availability of these genomes provided novel insights into the genetic traits, virulence, resistant determinants and

host adaptation of these pathogens. The *S.* Typhi CT18 genome comprises of a chromosome (4,809,037 bp), a multidrug resistant plasmid of incompatibility type HI1 (pHCM1- 218,150 bp) and a cryptic plasmid (pHCM2-106,516 bp), whilst the *S.* Paratyphi A ATCC9150 genome is about 200 kb smaller (4,585,299 bp) and does not harbor any plasmid. The difference in chromosome size between these two serovars is primarily reflected in the prophage content and other mobile elements; for instance, *S.* Typhi carry SPI-7 (a 134 kb region containing the Vi polysaccharide biosynthetic operon, sopE prophage and type IVB pili operon), which is absent from *S.* Paratyphi A. On the other hand, *S.* Paratyphi A carries three prophage regions that are not present in *S.* Typhi, including a sopE prophage (SPA-2-SopE) [231,233]. Subsequently, additional genomes of *S.* Typhi (strain Ty2) and *S.* Paratyphi A (strain AKU12601) became available and allowed extensive genomic comparisons among *S.* Typhi isolates (CT18 versus Ty2) and *S.* Paratyphi A isolates (ATCC9150 versus AKU12601). Comparative genomics exhibited an exceptionally high level of pseudogenes and distinctly conserved genomic backbones in both serovars [235,240]. More than 98 percent of the two Typhi genomes are shared and only 282 single nucleotide polymorphisms were identified [235]. Similarly, the two Paratyphi A genomes are collinear and highly similar, with only 188 single nucleotide polymorphisms detected [240]. The clonal natures of both Typhi and Paratyphi A have imposed significant challenges in understanding their bacterial population structures and molecular epidemiological characteristics.

In 2006, Roumagnac and co-workers were the first to describe the global population structure and evolutionary history of *S.* Typhi by using a mutation discovery method

to characterize SNPs within 200 gene fragments (1.85% of the Typhi genome) from a globally representative strain collection. The selection of gene fragments was based on the available complete genomes of strains CT18 and Ty2. Eighty-eight SNPs were detected and used to resolve Typhi into a rooted, maximally parsimonious phylogenetic tree defining 59 genetically distinct haplotypes (H1-H59) [128]. This study also gave an early warning of the emergence and global spread of H58 *S.* Typhi, which was particularly associated with nalidixic acid resistant mutations in the *gyrA* gene [128]. Latterly, Holt *et al*. paved the way for using a whole-genome approach to study the population structure and evolutionary traits of *S.* Typhi [244]. Their study identified 1,787 SNPs among 17 Typhi isolates belonging to diverse haplotypes, providing much better resolution for the Typhi phylogenetic tree, which also established an important framework for further molecular investigations.

Between 2008 and 2012, molecular epidemiological investigations were performed in highly endemic settings including Indonesia, Nepal, Kenya, Vietnam, Cambodia, and India by using SNP-based approaches to characterize the local *S.* Typhi population structure and investigate the transmission patterns of typhoid fever [131–134,245–247]. These studies revealed that local bacterial populations are usually highly diverse and involve multiple co-circulating haplotypes, and showed that environmental transmission is dominant in urban settings. The findings also highlighted the increasing frequency of H58 *S.* Typhi in many parts of the world (except for Indonesia); in many locations, it was observed that the H58 lineage might be replacing locally existing haplotypes. This particularly successful clone of *S.* Typhi is predominantly non-susceptible to fluoroquinolones and associated with an MDR

phenotype, suggesting that the heavy use of antimicrobials for typhoid treatment has exerted significant selective pressure for the maintenance and expansion of the H58 *S.* Typhi population [128].

1.9.2 The origin and global dissemination of H58 *Salmonella* Typhi

The rapid emergence and subsequent global dissemination of antimicrobial resistant H58 *S.* Typhi throughout South Asian, Southeast Asian and African countries has become an important public health threat, which urgently requires further investigation on the origin, population structure, transmission and antimicrobial resistance determinants of this particular lineage. Additionally, the global dominance of H58 *S.* Typhi has also revealed the limitations of SNP-based typing methods, which show a lack of resolution within H58 isolates and have thus been recognised as less effective methods for use in molecular epidemiological studies of typhoid fever. Consequently, the whole genome approach is preferred as an ultimate solution to better understand the origin and evolution of H58 Typhi, as well as to track their global transmission. Over the last decade, the advances of next generation sequencing technology have allowed large-scale bacterial genome sequencing, marking the dawn of the genomic era and revolutionizing the molecular surveillance of infectious diseases. Using an integrative approach of genomics and conventional epidemiology, Wong and colleagues performed whole genome sequencing and phylogeographical analyses of the largest ever collection of *S.* Typhi (1,832 isolates from 63 countries, including 832 H58 isolates collected from 1992 to 2013) to investigate the evolutionary history, population structure and global transmission of H58 *S.* Typhi [248]. This study confirmed that the global H58 *S.* Typhi population is highly clonal

compared to non-H58 populations, an indicator of recent clonal expansion. Their findings also provided strong evidence that H58 lineage emerged in South Asia and expanded dramatically from the early 1990s. There were numerous international transmission events reconstructed from phylogeographical analyses, which demonstrated the propagation of H58 *S.* Typhi from South Asia to many countries in Southeast Asia, Western Asia, East and South Africa and Fiji, followed by local and region-wide transmission in different endemic areas. In particular, this study underlined the silent ongoing epidemic of MDR H58 *S.* Typhi across many countries in Eastern and Southern Africa, probably driven by antimicrobial use considering most of the non-H58 isolates in these regions were drug susceptible. The H58 lineage is associated with an MDR phenotype and reduced susceptibility to fluoroquinolones; however, this study showed that the pattern clearly differs between geographical regions, likely due to differences in regional antimicrobial usage. H58 *S.* Typhi isolates from Southeast Asia were commonly multidrug resistant and possessed *gyrA* mutations, yielding simultaneous factors contributing to reduced susceptibility to fluoroquinolones, whereas most of the South Asian H58 isolates from recent years were non-MDR and harbored *gyrA* mutations. In Africa, however, the majority of H58 isolates were multidrug resistant but did not possess gyrA mutations [248]. Importantly, the MDR gene cassette (Tn2670-like) is largely found in plasmids but can also be inserted into the chromosome at different positions, which can affect the bacterial fitness while still maintain the MDR phenotype [248].

The global expansion of H58 *S.* Typhi population has been reshaping the global population structure of *S.* Typhi and changing the epidemiology of the disease; for

instance, African countries have recently experienced a surge of MDR typhoid fever which had previously been uncommon [249–252]. Wong and her colleagues' work set milestones in understanding the molecular epidemiology of *S.* Typhi, providing a global population framework and valid methodology for further local investigations into the transmission of antimicrobial resistant lineages of *S.* Typhi in different endemic areas. Their study also demonstrated the invaluable role of genomics in understanding the *S.* Typhi population structure and tracking the emergence and global transmission of antimicrobial resistant organisms.

1.10 Aims of this study

The emergence of fluoroquinolone resistant *S.* Typhi in South Asia, particularly Nepal, has caused significant challenges in typhoid management and has been sporadically reported to be associated with treatment failure. However, the bacterial population structure, antimicrobial resistance determinants and the association between fluoroquinolone resistant phenotype/bacterial genotype with clinical outcomes of typhoid fever have not yet been well characterized. Concurrently, *S.* Paratyphi A infections have also increased rapidly in Nepal and little is known about the bacterial population dynamics, antimicrobial resistance and epidemiological features in this settings. Previous epidemiological surveillance studies of typhoid fever have primarily focused on urban populations, where high incidences of typhoid fever have been reported. There has been a large gap in knowledge about the burden of typhoid fever as well as risk factors, transmission patterns and antimicrobial resistance in rural settings such as described in this thesis for Siem Reap, Cambodia. One of the most long-lasting and elusive research questions related to typhoid fever is

about the role of chronic carriage in disease transmission and bacterial persistence. Genomic characterization and comparison between chronic and acute *S.* Typhi isolates is necessary to provide novel understanding into the relative importance of chronic carriage.

In order to fill in these current gaps in typhoid research, the aims of my study are as follows:

- To use a combined genomic and clinical approach to characterize the population structure and dynamics of *S.* Typhi in Nepal, identify fluoroquinolone resistant determinants and understand the clinical features of typhoid fever.
- To characterize the spatiotemporal dynamics, antimicrobial resistance patterns/determinants and phylogenetics of *S.* Paratyphi A in Kathmandu, Nepal.
- Combine genomics and conventional epidemiological approaches to understand the molecular and spatial epidemiology of typhoid fever in rural areas in Siem Reap, Cambodia.
- Describe the genomic traits of carrier *S.* Typhi isolates and the genetic relationship between carrier and acute *S.* Typhi to provide insights into the role of typhoid carriage in the disease transmission

Chapter 2

Methods

2.1 Study sites and settings

2.1.1 Patan Hospital in Kathmandu, Nepal

The majority of bacterial isolates and associated metadata in my study originated from typhoid studies conducted in Patan Hospital in Kathmandu, Nepal. Kathmandu, the capital city of Nepal, has a population of 1.5 million and is situated at an altitude of 1300m. The climate varies from cold winter months (December to February) to spring (March to May), the hot monsoon season (June to August) and autumn (September to November). Traditionally, the monsoon is characterized by a heavy burden of enteric infections. Patan Hospital, a 318-bed government hospital located in the Lalitpur Sub-Metropolitan City (LSMC) in the Kathmandu Valley providing both emergency and elective inpatient services. Each year, Patan Hospital has 300,000 outpatients and 16,000 inpatients, and the bed occupancy is approximately 85%. Ninety per cent of patients are from the immediate Kathmandu Valley and Lalitpur area. Enteric fever is common at the outpatient clinic at Patan Hospital, which has approximately 200,000 outpatient visits annually. The population of LSMC is generally poor, with most living in overcrowded conditions and obtaining their water from stone spouts or sunken wells. Antimicrobials are available without prescription in the community in a variety of public and private outlets and there are numerous private physician clinics where patients may seek advice and clinical diagnosis for febrile disease. There has been no widespread implementation of a typhoid vaccine in this area, yet a generic typhoid Vi vaccine is available for purchase in some health care settings.

2.1.2 Angkor Hospital for Children in Siem Reap, Cambodia

Angkor Hospital for Children (AHC) is one of two pediatric hospitals in Siem Reap City and has approximately 125,000 attendees and 4,000 admissions per year. The patients attending AHC are <16 years of age and come from a wide geographical radius and attend the hospital for various conditions. The majority of patients reside in the province of Siem Reap, which is located in northwest Cambodia and is bordered in the south by the Tonle Sap Lake, the largest freshwater lake in Southeast Asia. According to available census data, the province had a population of 896,443 people living in an area of 10,299 km^2 in 2008; the province is subdivided administratively into 12 districts, 100 communes (which are within districts) and 907 villages [253]. Cambodia has a tropical climate with a dry and wet season each year. During the wet season (April–October) the area of the Tonle Sap Lake can expand dramatically, increasing from 3,500 km^2 up to approximately 14,500 km^2, with the depth increasing from 0.5m up to 6-9m [254].

2.2 Bacterial identification and antimicrobial susceptibility testing

2.2.1 Blood culture

Routine diagnosis of typhoid fever was performed by blood culture. Blood (1-3 ml for children; 5-10 ml for adults) was taken from all patients for bacterial culture on enrolment. Adult blood samples were inoculated into media containing tryptone soya broth and sodium polyanethol sulphonate, up to a total volume of 50 ml. Bactec Peds Plus culture bottles (Becton Dickinson, New Jersey, USA) were used for paediatric blood samples. The bottles were incubated at 37^{o}C and examined daily for bacterial growth up to 7 days. Positive bottles were subcultured onto blood, chocolate and

MacConkey agar and presumptive Salmonella colonies were identified using standard biochemical tests and serotype-specific antisera (Murex Biotech, Dartford, England)

2.2.2 Bile and stool culture

Bile and stool were collected for culture from all patients undergoing cholecystectomy or laparotomy surgery between June 2007 and October 2010. Bile was inoculated into equal volumes of Selenite F broth and Peptone broth and incubated at 37ºC overnight. Broths were subcultured onto MacConkey agar and Xylene Lysine Deoxycholate (XLD) agar. After overnight incubation at 37ºC the plates were examined for the growth of Gram-negative bacteria and colonies were identified by standard microbiological methods and identified by API20E manufactured by bioMerieux, Inc. *S.* Typhi and *S.* Paratyphi A isolates were confirmed by slide agglutination by specific antisera (Murex Biotech, Biotech, England).

2.2.3 Antimicrobial susceptibility testing

Antimicrobial susceptibility testing was performed by the modified Bauer-Kirby disc diffusion method [255] with zone size interpretation based on Clinical and Laboratory Standards Institute guidelines (CLSI, 2012) [256]. The control strains used for all susceptibility tests were *E. coli* ATCC 25922, *Pseudomonas aeruginosa* ATCC 27853 and *Staphylococcus aureus* ATCC 29213. Etests were used to determine the minimum inhibitory concentrations (MICs), following the manufacturer's recommendations (bioMérieux, France). Ciprofloxacin MICs were used to categorize *S.* Typhi and *S.*

Paratyphi A isolates as susceptible (≤0.06 μg/mL), intermediate (0.12–0.5 μg/mL) and resistant (≥1 μg/mL) following CLSI guidelines (CLSI, 2012) [256].

2.3 Vi agglutination assay

The two carriage isolates lacking ViaB operon (GB003 and GB428) were grown in LB agar plates with different NaCl concentrations (1mM, 85mM and 170mM). Agglutination tests were performed on glass microscope slides by mixing 10μl of suspensions of single colonies with 50μl of antisera against Vi (Murex Biotech, Biotech, England). Agglutination was read after rocking the slide for 1 minute. Two other carriage isolates (GB125 and GB169) were used as controls for this experiment.

2.4 Data sources and bacterial isolates

2.4.1 Data sources and bacterial isolates in chapter 3

The *S.* Typhi isolates and corresponding clinical data for chapter 3 originated from an open-label, randomized, controlled, superiority trial conducted at Patan Hospital and the Civil Services Hospital in the Kathmandu valley, Nepal between 2011 and 2014. The trial was registered at www.clinicaltrials.gov (ISRCTN63006567). Briefly, patients were randomly assigned to seven days of treatment with either oral gatifloxacin (400 mg tablets, Square Pharmaceuticals Limited, Bangladesh) at a dose of 10 mg/kg once daily or intravenous ceftriaxone (Powercef, 1000mg injection vial, Wock- hardt Ltd, India), injected over 10 min at a dose of 60 mg/kg up to a maximum of two grams (aged 2 to 13 years) or two grams (≥ 14 years) once daily. The trained community medical auxiliaries (CMAs) visited each patient assigned to treatment twice per day for at least 10 days or until the patient was asymptomatic. The CMAs

gave the drugs, and recorded drug doses, administration times, oral temperatures, symptoms, and potential adverse effects in a standard case-record form. Complete blood count, serum creatinine, liver-function parameters and serum glucose at enrolment and on day 8 of treatment were measured. Blood from all patients was subjected to bacterial culture at enrolment and on day 8 after randomization if *S.* Typhi or *S.* Paratyphi were isolated at enrolment, or if their symptoms suggested a clinical relapse.

The primary endpoint was a composite of treatment failure, defined as the occurrence of at least one of the following events: fever clearance time (FCT) (time from the first dose of a study drug until the temperature dropped to 37.5˚C and remained there for at least two days) more than seven days post-treatment initiation; requirement for rescue treatment as judged by the treating physician; blood culture positivity for *S.* Typhi or *S.* Paratyphi on day eight of treatment (microbiological failure); culture-confirmed or syndromic enteric fever relapse within 28 days of initiation of treatment; and the development of any enteric fever-related complication (e.g. clinically significant bleeding, fall in the Glasgow Coma Score, perforation of the gastrointestinal tract and hospital admission) within 28 days after the initiation of treatment. Time to treatment failure was defined as the time from the first dose of treatment until the date of the earliest failure event. FCTs were calculated electronically using twice-daily recorded temperatures and treated as interval-censored outcomes. Patients without fever clearance or relapse, respectively, were censored at the time of their last follow-up visit.

There was a total of 78 *S.* Typhi isolates identified from this trial. Additionally, 58 *S.* Typhi isolates from previous studies conducted between 2008 and 2013 in Patan Hospital and genome sequences of 19 *S.* Typhi isolates from a recent international study of the H58 lineage were also included for phylogenetic analysis. Details about these *S.* Typhi isolates and their corresponding sequencing metadata are shown in appendix A.

2.4.2 Data sources and bacterial isolates in chapter 4

223 *S.* Paratyphi A isolates were collected from several studies conducted at Patan Hospital, Nepal between 2005 and 2014. Of which, 206 *S.* Paratyphi A isolates were collected from acutely infected people (defined as acute *S.* Paratyphi A isolate) from four randomized control trials: gatifloxacin vs cefixime (2005) [257], gatifloxacin versus chloramphenicol (2006-2008) [258], gatifloxacin versus ofloxacin (2008-2011) [259], gatifloxacin versus ceftriaxone (2011-2014) [260] and a matched case-control study (2011) [129]. 17 *S.* Paratyphi A isolates were recovered from gallbladder bile of people who underwent cholecystectomy or laparotomy surgery for symptomatic cholelithiasis from 2007 to 2010 [130]. These 17 *S.* Paratyphi A isolates were defined as carrier *S.* Paratyphi A isolates given the fact that they were isolated from the gallbladder of asymptomatic carriers and the duration of carriage was unknown. Additionally, epidemiological and clinical information was respectively collected from the above studies where available. Major reported variables included age, sex, main water sources, water treatment methods, clinical symptoms at presentation, antimicrobial resistance information and individual GPS location. Details about the *S.* Paratyphi A isolates and their associated metadata are shown in appendix B.

2.4.3 Data sources and bacterial isolates in chapter 5

The *S.* Typhi isolates and associated metadata described in chapter 5 originated from a retrospective study of invasive salmonellosis conducted at Angkor Hospital for Children (AHC) in Siem Reap City in Cambodia between January 2007 and December 2014 (appendix C). Case and control populations were identified from the electronic hospital and laboratory information system of AHC. For this investigation, the case population was defined as the population of hospital inpatients from whom *S.* Typhi was isolated from a blood culture. The control population was defined as the patient population admitted to AHC who did not have typhoid fever based on the recorded discharge diagnosis (International Classification of Disease (ICD)-10 code). Patients with a discharge diagnosis of typhoid fever but without blood culture confirmation (n=410) were not included in the risk factor analysis. Additionally, for the mapping and population risk factor analyses, cases that lived outside of a 100km radius from AHC were excluded. Data on age, sex, home location (commune level), admission and discharge dates for cases and controls were extracted from the electronic hospital information system. If a case or control was readmitted to the hospital with the same discharge diagnosis within a seven-day period, only the initial admission was included in the analysis.

Commune-level census data were obtained from the Cambodian National Report on General Population Census of 2008 [253]. The extracted information included details regarding demographic indicators, age structure, literacy and education, housing and household characteristics, and access to toilet facilities and drinking water. Based on this report, a commune was classified as urban if the population density exceeded

200/km^2, less than half of men were employed in agriculture and the total population exceeded 2,000. Monthly average precipitation was collected from Siem Reap Weather Station and MRCS (Mekong River Commission Secretariat) [261]. Shuttle Radar Topography Mission (SRTM) elevation data were obtained from the CGIAR Consortium for Spatial Information (CGIAR-CSI) [262]. Shapefile layers containing 2008 commune-level population census data were accessed from Open Development Cambodia, an open-access data website providing data on Cambodia and its economic and social development (http://www.opendevelopmentcambodia.net).

2.4.4 Data sources and bacterial isolates in chapter 6

Between June 2007 and October 2010, a study of microbiology and epidemiology of invasive *Salmonella* carriage was conducted at Patan Hospital in Kathmandu [130]. Bile samples from 1377 individuals who underwent cholecystectomy for acute or chronic cholecystitis were subjected to microbiological examination. 24 *S.* Typhi isolates identified from the positive bile samples and their associated metadata were retrospectively collected. These *S.* Typhi isolates were defined as carrier *S.* Typhi isolates as they were isolated from gallbladder of asymptomatic carriers and the duration of carriage was unknown. Additionally, 96 *S.* Typhi isolates recovered from acute typhoid patients (defined as acute *S.* Typhi isolates) during a randomized controlled trial conducted in Patan Hospital between July 2008 and August 2011 were also included [259]. Details about these *S.* Typhi are shown in appendix D.

2.5 Whole genome sequencing

2.5.1 DNA extraction

Bacterial isolates grown overnight on Nutrient Agar plates were subjected to DNA extraction using the Wizard Genomic DNA Purification Kit (Promega, Wisconsin, USA). This DNA extraction kit includes nuclei lysis solution, RNase solution, protein precipitation solution and DNA rehydration solution.

There are four main steps:

- Cell lysis
- RNA digestion
- Removal of proteins by salt precipitation (DNA left in solution)
- Isopropanol precipitation to concentrate and desalt the DNA

Procedure:

– Scrape the surface the overnight culture plate and add into 1 ml of sterile saline to the appropriately labelled 1.5ml microcentrifuge tube

– Centrifuge the tubes at 13,000 rpm for 2 minutes to pellet the cells

– Remove and discard the supernatants from the tubes

– Add 600 µl nuclei lysis solution to the tubes and gently pipette until the cells are fully re-suspended

– Incubate at 80^{o}C for 5 min to lyse the cells; then cool to room temperature

– Add 3 µl of RNase solution to the cell lysates. Invert the tubes 2-5 times to mix

– Incubate at 37^{o}C for 5 min. Cool the samples to room temperature

– Add 200 µl of protein precipitation solution to each lysate. Vortex vigorously for 20 seconds to mix the protein precipitation solution with the cell lysates

– Incubate the samples on ice for 5 min

– Centrifuge at 13,000 rpm for 3 min

– Pipette 600 µl isopropanol into the newly labeled tubes

– Transfer the supernatants containing the DNA to the appropriately labeled microcentrifuge tubes containing 600 µl of isopropanol

– Gently mix by inversion until the thread like strands of DNA form a visible mass

– Centrifuge at 13 000 rpm for 2 min

– Carefully pour off the supernatants and drain the tubes on clean absorbent paper

– Add 600 µl of room temperature 70% ethanol to each tube and gently invert the tubes several times to wash the DNA pellet

– Centrifuge at 13,000 rpm for 2 minutes and carefully aspirate the ethanol

– Drain the tubes on clean absorbent paper and allow the pellet to air-dry (will take approximately 10-15 minutes)

– Add 200 µl of DNA rehydration solution to the tubes and incubate at 65^{o}C for 1 hour. Periodically mix the solution gently. Alternatively the DNA can rehydrate by incubating overnight at room temperature or at 4°C

– Store the DNA at -20^{o}C until required

2.5.2 DNA quantification

Qubit dsDNA high sensitivity kit (Thermo Fisher Scientific, Waltham, MA, USA) was used to accurately quantify the DNA concentration. The kit includes concentrated assay reagent, dilution buffer, and prediluted DNA standards.

Procedure:

- Prepare thin-wall, clear, 0.5 ml PCR tubes
- Prepare the Qubit working solution by diluting Qubit HS reagent 1:200 with Qubit HS buffer in a plastic tube
- For 2 standards (positive and negative):
    - 10 µl each standard
    - 190 µl working solution
- For DNA samples
    - 2 µl each DNA sample
    - 198 µl working solution
- Incubate all tubes at room temperature for 2 minutes
- Proceed to “reading standards and samples using Qubit 3.0 fluorometer

2.5.3 DNA library preparation

DNA samples were subjected to library construction using Nextera DNA sample prep kit (Illumina, USA). Nextera DNA library prep protocol can fragment and add adapter sequences onto template DNA with a single tube Nextera reaction to generate multiplexed pair-end sequencing libraries.

2.5.3.1 Tagment genomic DNA

Nextera library preparation procedure combines DNA fragmentation, end-polishing, and adaptor-ligation steps into one, termed tagmentation.

Procedure:

- Add 20 µl of genomic DNA at 2.5 ng/µl (50 ng total) to each 0.2 ml tube
- Add 25 µl of TD Buffer to the tubes containing genomic DNA
- Add 5 µl of TDE1 to the tubes containing genomic DNA and TD Buffer
- Pipette up and down 10 times (or more) to mix thoroughly. Quick spin
- Place on the thermal cycler with a heated lid and run the program

  55°C for 5 minutes

  Hold at 10°C

2.5.3.2 Clean up Tagmented DNA

The tagmented DNA is purified from the Nextera transposome as the transposome can bind tightly to DNA ends and interfere with downstream reactions.

Procedure**:**

- Transfer 50 µl from each tube to a new eppendorf, add 300 µl DNA binding buffer (Zymo), mix and transfer the mixture to Zymo-Spin column in a collection tube
- Centrifuge 13,200 rpm for 1 minute
- Discard the flow-through
- Wash twice as follow
  - Add 500 µl wash buffer (Zymo) to each Zymo-Spin column
  - Centrifuge 13,200 rpm for 30 seconds
  - Discard the flow-through
- Centrifuge at 13,200 rpm for 1 minute to make sure there is no residual wash buffer
- Transfer the Zymo-spin column to a new eppendorf

- Add 25 µl of RSB (Resuspension Buffer) directly to the Zymo-Spin column
- Incubate for 2 minutes at room temperature
- Centrifuge at 13,200 rpm for 2 minutes

2.5.3.3 Amplify Tagmented DNA

Purified tagmented DNA is amplified using a 5-cycle PCR program. This step adds index 1 (i7) and index 2 (i5) and adapters (P5, P7) required for cluster generation and sequencing. Arrange index 1 (i7) and index 2 (i5) in an Index Plate as follow:

Columns 1-12: Index 1 (i7) adapter (orange caps)

Rows A-H: Index 2 (i5) adapter (white caps)

Procedure:

- Prepare a new 0.2ml tubes
- Add 5 µl each index 1 adapter down each column
- Add 5 µl each index 2 adapter across each row
- Add 15 µl NPM (Nextera PCR Master Mix)
- Add 5 µl PPC (PCR Primer Cocktail)
- Add 20 µl each purified tagmented DNA (from step 3)
- Mix by pipetting up and down or vortex briefly. Quick spin
- Place on a thermal cycle with a heated lid and run the PCR program:

  72°C for 3 minutes

  98°C for 30 seconds

  5 cycles of:

  98°C for 10 seconds

  63°C for 30 seconds

72°C for 3 minutes

Hold at 10°C

2.5.3.4 Clean up libraries

This step uses AMPure XP beads to purify the library DNA and perform size selection to remove short library fragments.

- Prepare fresh 80% ethanol from absolute ethanol.
- Bring the AMPure XP beads to room temperature.

Procedure:

- Vortex AMPure beads rigorously to ensure beads are homogenous.
- Add 30 µl AMPure beads to each tube containing amplified library
- Mix by pipetting up and down
- Incubate at room temperature for 5 minutes
- Place on a magnetic stand and wait until the liquid is clear. Keep the tubes on the magnetic stand for the following steps.
- Discard supernatant
- Wash twice as follows:
    - Add 200 µl fresh 80% Ethanol. Avoid disturbing the beads
    - Incubate on the magnetic stand until the supernatant is clear
    - Remove and discard the supernatant from each tube
- Remove residual Ethanol using white tip with pipette P20
- Air-dry beads for 10 minutes
- Remove the tubes from the magnetic stand

- Add 32.5 µl RSB (Resuspension Buffer) and pipette up and down to mix
- Incubate at room temperature for 5 minutes
- Place the tubes on the magnetic stand and wait until the supernatant is clear
- Transfer 30 µl supernatant to a new tube, avoid bead carryover

2.5.3.5 Library quantification

Library quantification was performed using KAPA SYBR® FAST qPCR Kit (Kapa Biosystems, Wilmington, USA). KAPA Library Quantification Kits enable accurate and reproducible qPCR-based quantification of libraries prepared for Illumina sequencing. The kit contains six DNA standards (a 10-fold dilution series of a linear, 452 bp template ranging from 0.0002 pM up to 20 pM), Library Quantification Primer Premix (10X) and KAPA SYBR® FAST qPCR Master Mix (2X).

Procedure**:**

- Make 1:1000 dilution of dsDNA library using Molecular Grade (MG) Water (or 10mM Tris-HCl, pH 8 + 0.05% Tween 20)
- Prepare qPCR mix as follows:

KAPA SYBR FAST qPCR Master Mix containing Primer Premix 6 µl

Diluted library DNA or DNA Standard (1-6) 4 µl

- Run qPCR protocol

  Initial activation 95^{o}C 5 min

  35 cycles:

Denaturation	95^{o}C	30 sec
Annealing/extension/data acquisition	60^{o}C	45 sec

Melting curve analysis (optional)

- Analyze the data and calculate the library concentration
  - Confirm 90-100% reaction efficiency for standards
  - Calculate the concentration of diluted library as determined by qPCR in relation to concentrations of DNA Standards 1-6
  - Perform a size adjusted calculation for the difference in the size between average fragment length of library and the DNA Standard (452 bp)
  - Calculate the undiluted library by taking into account the dilution factor and volume used per reaction (4 µl)

2.5.3.6 Normalize and Pool libraries

DNA Libraries were normalized to 4 nM and diluted to 20 pM before loading into the flow cell for Illumina sequencing on Miseq System.

Procedure:

Normalize and pooling

- Dilute each library to 4 nM using MG water (or 10mM Tris-HCl, pH 8 + 0.05% Tween 20)
- Transfer 5 µl each normalized library to a new eppendorf
- Vortex and spin down.

These are pooled libraries, next step is to denature with 0.2 N NaOH and dilute to 20 pM

Denaturing and Diluting

- Prepare 0.2 N NaOH (to denature the pooled libraries) as follows:

- MG water: 800 µl
- 1.0 N NaOH: 200 µl
- Vortex and spin down (use only within 12 hours)

- Thaw HT1 (hybridization buffer) and store on ice until you are ready to dilute denatured library
- Combine the following volumes in a 0.2 ml tube

  4 nM pooled libraries (5 µl)

  0.2 N NaOH (5 µl)
- Vortex briefly and spin down
- Incubate at room temperature for 5 minutes
- Denature at 95°C for 2 minutes (optional)
- Add 990 µl prechilled HT1 to the tube containing the denatured library. Keep on ice. This step results in 1 ml of a 20 pM denatured library ready for loading into the flow cell.

2.5.4 Perform a run on Miseq

Prepared pooled libraries were sequenced using Miseq Reagent v2 (300/500 cycles). The sequencing kit contains: Reagent Cartridge, buffer HT1, PR2 Bottle and Miseq Flow Cell.

Procedure:

- Clean the flow cell thoroughly using ELGA water prior to use
- Remove the reagent cartridge from -20°C, thaw in a water bath at room temperature (take about an hour)

- When thawed, remove the cartridge from water bath, gently tap it to the bench and dry the base.
- Invert the cartridge ten times to check if all the reagents are thawed
- Gently tap the cartridge on the bench to reduce air bubbles
- Load 600 µl of your prepared library onto the reagent cartridge in the reservoir labeled “Load Samples” (position 17)
- Using the illumina Experiment Manager to set up sample sheet
- Using the Miseq Control Sofware (MCS) interface, follow the run setup steps to load the flow cell and reagents and then start the run

2.6 Single Nucleotide Polymorphism (SNP) detection and analysis

2.6.1 SNP detection and analysis for *Salmonella* Typhi genomes

2.6.1.1 SNP detection and annotation

All raw Illumina reads were mapped to the reference sequence of *S.* Typhi strain CT18 (Accession no: AL513382), plasmid pHCM1 (AL513383) and pHCM2 (AL513384) using SMALT (version 0.7.4). Candidate single nucleotide polymorphisms (SNPs) were called against the reference sequence using SAMtools [263] and filtered with a minimal Phred quality of 30 and a quality cut-off of 0.75. The allele at each locus in each isolate was determined by reference to the consensus base in that genome, using *samtools mpileup* and removing low confidence alleles with consensus base quality ≤20, read depth ≤5 or a heterozygous base call. SNPs called in phage regions, repetitive sequences or recombinant regions identified previously were excluded [244]. Gubbins was also used to identify recombinant regions from the whole

genome alignment produced by SNP-calling isolates [264] and SNPs detected within these regions were also removed. SNPs were annotated using the parseSNPTable.py script in the RedDog pipeline (https://github. com/katholt/RedDog). Strains belonging to haplotype H58 were defined by the SNP *glpA*-C1047T (position 2,348,902 in *S.* Typhi CT18, BiP33, as previously described [128,244]).

2.6.1.2 *Salmonella* Typhi genotyping

From the detected SNPs in *Salmonella* Typhi genomes, a subset of 68 SNPs was used to assign *Salmonella* Typhi isolates to previously defined lineages according to an extended *S.* Typhi genotyping framework (Table 2.1). This SNP-based genotyping framework included four primary clusters, 16 clades and 49 subclades, developed from whole genome data of more than 1800 globally representative *S.* Typhi isolates [265]. Under this genotype nomenclature, the globally disseminated haplotype H58 belongs to subclade 4.3.1.

Table 2.1 Canonical SNPs for genotyping *Salmonella* Typhi into Clades and Subclades [265]

Group	N	In	Out	CT18 position	Gene	Product	Nt	Codon	ns	s	%Div
0.1	11	T	C	655112	STY0653	alkyl hydroperoxide reductase c22 protein	547	183	0	2	0.35
0.0.1	2	A	G	773487	STY0776	succinate dehydrogenase hydrophobic membrane anchor protein	333	111	0	1	0.29
0.0.2	2	C	T	1804415	STY1910	putative ATP/GTP-binding protein	996	332	0	3	0.27
0.0.3	2	A	G	1840727	STY1951	putative ATP-dependent helicase	1302	434	1	6	0.37
0.1.1	2	A	G	3640678	STY3775	primosomal protein replication factor	825	275	0	10	0.46
0.1.2	1	T	C	270120	STY0255	acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha	196	66	0	1	0.10
0.1.3	1	A	G	102135	STY0104	CorD protein	330	110	0	1	0.26
1	7	C	T	316489	STY0304	conserved hypothetical protein	117	39	14	6	1.49
1.1	6	T	C	4105384	STY4239	conserved hypothetical protein	267	89	0	1	0.17
1.1.1	1	A	C	555826	STY0548	putative ABC transporter ATP-binding protein	117	39	0	2	0.30
1.1.2	2	T	C	2360997	STY2526	Ais protein	273	91	2	5	1.15
1.1.3	2	T	C	4664137	STY4803	ornithine carbamoyltransferase	213	71	11	13	2.39
1.1.4	1	A	G	2166082	STY2335	uridine kinase	81	27	0	1	0.16
1.2	1	T	C	30192	STY0031	fimbrial subunit	195	65	0	1	0.19
1.2.1	1	T	C	4288272	STY4419	putative lipoprotein	588	196	0	2	0.31
2	319	A	G	2737027	STY2867	hypothetical protein	171	57	1	3	0.72
2.0.1	9	G	T	1215983	STY1260	putative ROK-family protein	651	217	2	2	0.44
2.0.2	4	A	G	4132985	STY4268	gluconate utilization operon repressor	237	79	3	3	0.60
2.1		A	G	146673	STY0144	UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diaminopimelate--D-alanalanyl ligase	297	99	0	4	0.29
2.1.1	3	A	G	2517324	STY2682	putative oxidoreductase	283	95	0	3	0.38
2.1.2	2	T	C	3920009	STY4059	conserved hypothetical protein	768	256	2	5	0.80
2.1.3	3	A	G	3276735	STY3432	D-galactarate dehydratase	909	303	0	1	0.06
2.1.4	1	T	C	1173984	STY1220	flagellar P-ring protein precursor	360	120	2	4	0.54
2.1.5	2	A	G	2683312	STY2819	putative oxidoreductase	168	56	4	4	0.87
2.1.6	9	A	G	4013386	STY4154	putative DNA-binding protein	73	25	1	3	0.98
2.1.7	49	A	G	4094437	STY4227	nickel responsive regulator	222	74	0	1	0.25
2.1.8	6	A	G	827036	STY0829	biotin synthesis protein BioC	93	31	1	1	0.26
2.1.9	12	T	C	3476114	STY3622	uroporphyrinogen III synthase	195	65	3	3	0.81
2.2		C	T	4355243	STY4468	lysR family regulatory protein	156	52	0	3	0.34
2.2.1	11	A	G	2723847	STY2853	putative sigma(54) modulation protein	300	100	0	2	0.59
2.2.2	9	C	T	4388609	STY4497	alpha-galactosidase	1092	364	3	2	0.37
2.2.3	7	A	G	703762	STY0707	glutamate/aspartate transport ATP-binding protein GltL	255	85	0	4	0.55
2.2.4	4	A	G	431216	STY0419	pyrroline-5-carboxylate reductase	351	117	0	2	0.25
2.3	72	A	G	3095443	STY3228	D-erythrose 4-phosphate dehydrogenase	717	239	2	7	0.86
2.3.1[a]	21	T	C	316186	STY0303	probable lipoprotein	354	118	0	2	0.37
2.3.2	15	T	C	1934711	STY2080	putative cation transporter	468	156	1	2	0.34
2.3.3	9	T	C	2811222	STY2937	glycine betaine-binding periplasmic protein precursor	48	16	2	1	0.30

2.3.4	12	T	C	3092900	STY3226	fructose 1,6-bisphosphate aldolase	927	309	1	3	0.37
2.3.5[b]	15	G	A	2723724	STY2853	putative sigma(54) modulation protein	177	59	0	2	0.59
2.4	38	C	A	3437570	STY3584	sec-independent protein translocase protein	621	207	0	3	0.389
2.4.1	24	A	G	1780319	STY1876	hypothetical protein	126	42	0	1	0.57
2.5	42	T	C	1535365	STY1588	pyridine nucleotide transhydrogenase subunit-beta	342	114	4	5	0.65
2.5.1	21	T	C	1792810	STY1897	2-dehydro-3-deoxyphosphooctonate aldolase	54	18	0	2	0.23
3[c]	399	T	C	3062270	STY3196	lysyl tRNA synthetase (LysRS)	989	323	3	4	0.46
3.0.1	2	C	T	1799842	STY1906	ribose-phosphate pyrophosphokinase	480	160	0	4	0.42
3.0.2	2	T	C	432732	STY0421	shikimate kinase II	387	129	0	2	0.37
3.1	66	G	A	3069182	STY3203	conserved hypothetical protein	9	3	0	1	0.37
3.1.1	25	A	G	2732615	STY2863	30S ribosomal subunit protein S16	154	52	1	3	1.60
3.1.2	27	T	C	3770391	STY3909	ATP synthase subunit B	177	59	0	2	0.42
3.2	58	A	G	2269835	STY2438	endonuclease IV	819	273	2	3	0.58
3.2.1	44	T	C	4215341	STY4333	phosphoribulokinase	294	98	1	3	0.46
3.2.2	12	T	C	4602946	STY4741	probable sugar phosphotransferase	444	148	1	2	0.65
3.3	62	T	C	3368641	STY3527	conserved hypothetical protein	177	59	2	2	0.99
3.3.1	32	A	G	2245432	STY2413	cytidine deaminase	684	228	1	2	0.34
3.4	56	A	C	3164162	STY3316	hydrogenase-2 component protein	30	10	1	2	0.61
3.5	122	A	G	3923165	STY4063	deoxyuridine 5'-triphosphate nucleotidohydrolase	411	137	0	1	0.22
3.5.1	1	T	C	1811809	STY1918	hydrogenase-1 operon protein HyaF	738	246	1	3	0.47
3.5.2	5	T	G	3729635	STY3874	glutamine synthetase	1317	439	1	2	0.21
3.5.3[d]	6	T	C	3817752	STY3949	hypothetical protein	586	196	4	2	0.53
3.5.4	31	T	C	183033	STY0176	dihydrolipoamide acetyltransferase component (E2) of pyruvate dehydrogenase	969	323	3	7	0.53
4	1089	T	C	1615350	STY1689	conserved hypothetical protein	105	35	1	3	1.15
4.1	96	A	G	2342045	STY2507	ribonucleoside-diphosphate reductase 1 beta chain	588	196	1	6	0.62
4.1.1	18	T	C	3996717	STY4134	alpha-amylase	1623	541	2	6	0.39
4.2	140	A	G	2640029	STY2781	SseB protein	210	70	2	2	0.51
4.2.1	40	T	C	989024	STY0994	killing factor KicB	312	104	2	3	0.38
4.2.2	77	A	G	3806278	STY3940	chromosomal replication initiator protein	1152	384	6	3	0.64
4.2.3	12	A	G	1611156	STY1683	putative oxidoreductase	502	168	1	4	0.56
4.3.1	853	T	C	2348902	STY2513	anaerobic glycerol-3-phosphate dehydrogenase subunit A	1047	349	2	8	0.61

Key: Group = Primary cluster/Clade/Subclade; N = number of *S*. Typhi isolates in the Group; In = allele within the group; Out = allele outside of group; Nt = nucleotide position in the gene; Codon = codon position in the gene; ns = number of non-synonymous SNPs in the gene; s = number of synonymous SNPs in the gene; %Div = % of diversity in the gene; [a] nested in 2.3.2; [b] nested in 2.3.3;[c] BiP48; [d] nested in 3.5.4; [e] BiP33

2.6.1.3 Functional analysis

First, SNPs occurring exclusively in carrier isolates were identified. Genes containing these SNPs were subsequently grouped by their predicted function based on the *Salmonella* Typhi functional classification scheme used by the Sanger Institute (www.sanger.ac.uk). This functional classification scheme is based on the genome annotation of *S.* Typhi strain CT18 [233]. A summary of main functional classes in this scheme was shown in Table 2.2. Similar analyses were performed to classify genes containing SNPs that occurred exclusively in acute isolates to different functional classes for comparisons with that of carrier isolates.

2.6.1.4 Pairwise SNP distance

Pairwise genetic distances (defined as difference in the number of SNPs) within and between acute and carrier isolates were estimated from their SNP alignment using packages ape (v4.1) and adegenet (v2.0.1) in R (v3.3.2). Pairwise SNP distances were extracted and plotted using the function pairDistPlot in adegenet package. Subsequently, the Wilcoxon rank sum test was used for testing the difference in the average pairwise SNP distances between groups.

Table 2.2 *Salmonella* Typhi functional classification scheme [233]

1.A Degradation	1.G.1 Biotin	3.C.4 Murein sacculus, peptidoglycan
1.A.1 Degradation of carbohydrate	1.G.10 Thioredoxin	4.A Transport/binding proteins
1.A.2 Degradation of amino acids	1.G.11 Menaquinone	4.A.1 Transport amino acid and amines
1.B.1 Glycolysis	1.G.12 Heme and porphyrin	4.A.2 Transport Cations
1.B.10 Glyoxylate Bypass	1.G.13 Cobalamin	4.A.3 Transport Carbohydrates, organic acids and alcohols
1.B.2 Pyruvate dehydrogenase	1.G.14 Iron uptake and storage	4.A.5 Transport Anions
1.B.3 Tricarboxylic acid cycle	1.G.2 Folic Acid	4.A.6 Transport Other
1.B.5.b Non-oxydative branch	1.G.3 Lipoate	4.B Chaperones
1.B.6 Entner-Doudoroff pathway	1.G.4 Molybdopterin	4.C Cell division
1.B.7.a Aerobic Respiration	1.G.5 Pentothenate	4.D Chemotaxis and mobility
1.B.7.b Anaerobic Respiration	1.G.6 Pyridoxine	4.G Detoxification
1.B.7.c Electron Transport	1.G.7 Pyridine nucleotide	4.H Cell Killing
1.B.8 Fermentation	1.G.8 Thiamine	4.I Pathogenicity
1.B.9 ATP-proton motive force	1.G.9 Riboflavin	5.A IS element, Phage related
1.C Central intermediary metabolism	1.H Fatty acid biosynthesis	5.B Colicin-related function
1.C.2 Gluconeogenesis	2 Broad regulatory function	5.D Drug/Analogue sensitivity
1.C.3 Sugar-nucleotide biosynthesis, conversions	3.A.10 Polysaccharides - (cytoplasmic)	5.F Adaptions and atypical conditions
1.C.4 Amino sugars	3.A.11 Phospholipids	5.H.a Hypothetical protein
1.C.5 Sulphur Metabolism	3.A.2 Ribosomal proteins - synthesis, modification	5.H.b Hypothetical protein
1.D Amino Acid Biosynthesis	3.A.3 Ribosomes - maturation and modification	5.I Unknown
1.D.1 Glutamate Family	3.A.5 Amino acyl tRNA synthesis; tRNA modification	SPI-1
1.D.2 Aspartate Family	3.A.7 DNA - replication, repair, restriction./modification	SPI-10
1.D.3 Serine Family	3.A.8 Protein translation and modification	SPI-2
1.D.4 Aromatic Amino Acid	3.A.9 RNA synthesis, modification	SPI-3
1.D.5 Histidine	3.B.1 Degradation of RNA	SPI-4
1.D.6 Pyruvate Family	3.B.2 Degradation of DNA	SPI-5
1.E Polyamine synthesis	3.B.3 Degradation of proteins, peptides, glycoproteins	SPI-6
1.F.1 Purine ribonucleotide biosynthesis	3.B.4 Degradations of polysaccharides	SPI-7
1.F.2 Pyrimidine ribonucle	3.C Cell envelope	SPI-8
1.F.3 2'-Deoxyribonucleotide metabolism	3.C.1 Membranes lipoprotein	SPI-9
1.F.4 Salvage of nucleosides and nucleotides	3.C.2 Surface polysaccharides & antigens	
1.F.5 Miscellaneous	3.C.3 Surface structure	

2.6.2 SNP detection and analysis for *Salmonella* Paratyphi A genomes

Raw Illumina reads were mapped to the reference sequence of *S.* Paratyphi A strain AKU_12601 (accession no: FM200053) and plasmid pAKU_12601 (accession no: AM412236) using SMALT version 0.7.4 (http://www.sanger.ac.uk/resources/software/smalt/). Candidate single nucleotide polymorphisms (SNPs) were identified against the reference sequence using SAMtools [263] and filtered with a minimal Phred quality of 30 and a quality cut-off of 0.75. The allele at each locus in each isolate was determined by reference to the consensus base in that genome, using *samtools mpileup* and removing low confidence alleles with consensus base quality ≤20, read depth ≤5 or a heterozygous base call. SNPs called in prophage regions, repetitive sequences (IS elements, transposases, etc.) identified previously were excluded [231,240]. Subsequently, Gubbins software was used to identify potential recombinant regions from whole genome alignment produced by SNP-calling isolates [264]. Detected regions were manually checked and SNPs found in these regions were also removed.

2.7 Phylogenetic analysis

2.7.1 Phylogenetic analysis for chapter 3

A maximum likelihood (ML) phylogeny was estimated using a 1440 SNP alignment of the 78 RCT isolates in RAxML (version 7.8.6) with the generalized time-reversible substitution model (GTR) and a gamma distribution, with support for the phylogeny assessed via 1000 bootstrap replicates. The alignment was then compared to a global *S.* Typhi sequence database, with a particular focus on identifying sequences with a mutational profile suggestive of shared ancestry with a divergent H58 clade identified

in the previous phylogeny. A secondary ML phylogenetic tree was then inferred from the SNP alignment (1642 SNPs) of the 136 Nepalese *S.* Typhi along with 19 recently described *S.* Typhi H58 with the aforementioned mutational profile, using the same parameters as above.

2.7.2 Phylogenetic analysis for chapter 4

The best-fit evolutionary model for the SNP alignment of 223 *S.* Paratyphi A isolates was identified based on the Bayesian Information Criterion in jModelTest implemented in IQ-TREE software [266] and maximum likelihood phylogenetic trees were subsequently reconstructed under the selected model (TVM). Clade support for this maximum likelihood tree was assessed via 1000 bootstrap pseudo-analyses. Phylogenetic subgrouping of major lineages in the phylogenetic tree was performed on the basis of monophyletic groups (at least three isolates) with moderately to highly supported bootstrap values (>60%). In order to investigate the population structure of Nepalese *S.* Paratyphi A in the global context, a secondary maximum likelihood phylogenetic tree was inferred from a separate alignment of 3958 SNPs of the 223 Nepalese *S.* Paratyphi A along with 111 *S.* Paratyphi A isolates from a global sequence database described previously [267]. Bootstrap support for this maximum likelihood tree was estimated from 100 pseudo-replicates.

2.7.3 Phylogenetic analysis for chapter 5

A maximum likelihood phylogenetic tree was constructed from a 188 chromosomal SNP alignment of H58 isolates with RAxML (version 7.8.6) using the generalized time-reversible model (GTR) and a gamma distribution to model site-specific rate

variation (GTR+Γ nucleotide substitution model in RAxML). Support for the ML phylogeny was assessed via 1,000 bootstrap pseudo-analyses of the alignment data. Phylogenetic subgrouping was defined based on monophyletic groups (lineages) with bootstrap values indicative of strong support (≥85%). To investigate the short-term divergence within the bacterial population and the transmission within the local population, a minimum spanning tree was reconstructed from the SNP alignment of lineage III and lineage IV identified in the ML tree (accounting for 95% of isolates) using the goeBURST algorithm in Phyloviz software (version 1.1) [268]. This algorithm identified seven sublineages based on similarity among allelic profiles and frequency of isolation within the population. Sequences with identical SNP profiles and isolated at the highest frequency within each sublineage were assigned as founder genotypes (viewed as the central nodes within each of the sublineages), with descendant genotypes (represented by terminal nodes surrounding the founder genotype) assigned based on similarity to founder SNP profiles. These descendent genotypes can differ from the parental genotype by a single or multiple SNPs.

2.7.4 Phylogenetic analysis for chapter 6

A maximum likelihood phylogenetic tree was reconstructed from the SNP alignment of 120 *S.* Typhi isolates, plus a *S.* Paratyphi A isolate included as an outgroup to root the tree, using RAxML version 8.2.8 with the generalized time-reversible model and a Gamma distribution to model the site-specific rate variation (GTR+Γ). Support for the maximum likelihood tree was assessed via bootstrap analysis with 1000 pseudoreplicates. To investigate the phylogenetic structure of Nepalese H58 *S.* Typhi in the global context, a secondary ML phylogenetic tree was inferred from a separate

SNP alignment of 78 Nepalese H58 *S.* Typhi along with 836 globally representative H58 *S.* Typhi described previously [248]. Support for this ML tree was assessed via 100 bootstrap replicates.

2.8 Resistance gene and plasmid analysis

Investigation of antimicrobial resistance genes and plasmids of *S.* Typhi and *S.* Paratyphi A isolates was performed using a local assembly approach with ARIBA (Antimicrobial Resistance Identifier by Assembly) [269]. Resfinder [270] and Plasmidfinder [271] were used as databases of antimicrobial resistant genes and plasmid replicons, respectively. Briefly, reference sequences in the database are clustered by similarity (minimum 90% sequence identity) and paired sequence reads are mapped to the reference sequences to generate a set of reads for each cluster. The reads for each cluster are assembled independently and the closest reference sequence to the resulting contigs is identified. The reads for the cluster are subsequently mapped back to the assembled contig to identify variants. ARIBA not only reports the acquired antimicrobial resistance genes, but also the quality of assemblies and any variants detected between the sequencing reads and the reference sequences including known resistance SNPs.

2.9 Pan-genome analysis

Raw reads from each isolate were *de novo* assembled using the short-read assembler Velvet with parameters optimized by Velvet Optimizer [272]. Contigs that were less than 300 bp long were excluded and the assembled contigs were annotated using Prokka [273]. The pan-genome was subsequently reconstructed using Roary (rapid

large-scale prokaryote pan-genome analysis) [274] with a threshold set to 95% sequence similarity at the amino-acid level. Unique and shared gene content between groups of isolates was identified based on set operations on the pan-genome.

2.10 Spatiotemporal cluster analysis

2.10.1 Cluster analysis for chapter 4

Spatial and spatiotemporal clustering analyses were performed using SaTScan v9.4 [275]. A Bernoulli model was used to examine the spatial clusters of each identified genotype, using all other genotypes as the background distribution of *S.* Paratyphi A cases in Lalitpur, Kathmandu. The upper limit for cluster detection was specified as 25% of the study population over 25% of the study duration (for spatiotemporal clustering). The significance of the detected clusters was assessed by a likelihood ratio test, with a *p*-value obtained by 999 Monte Carlo simulations generated under the null hypothesis of a random spatial and spatiotemporal distribution. All maps were created in ArcGIS 10.2 (ESRI, Redlands, CA, USA).

2.10.2 Spatiotemporal clustering detection for chapter 5

Spatiotemporal clustering analysis of *S.* Typhi in Siem Reap, Cambodia was performed using Moran's I and SaTScan methodologies. First, Moran's I test was used to evaluate global autocorrelation amongst communes that reported at least one case (n=78) of typhoid fever in GeoDa software (v1.6.7, https://geodacenter.asu.edu/). This test statistic provides an evaluation of whether the rates across the area of interest are spatially random (Moran's I=0), over-dispersed (Moran's I<0) or clustered (Moran's I>0) [276]. Next, Kulldorff's scan statistic in SaTScan (v9.1.1,

http://www.satscan.org/) was used to identify the location of clusters of communes with high rates of typhoid fever over space and time [275,277]. A cylindrical window was used to scan the area for clusters, with the size of the circle corresponding to the spatial scan and the height of the cylinder corresponding to time. The significance of the detected clusters was assessed by a likelihood ratio test, with a *p*-value obtained by 999 Monte Carlo simulations generated under the null hypothesis of random spatiotemporal distribution. In this analysis, scan windows were used to fit discrete Poisson models. For the sublineage-specific analyses, all case communes were included and those without cases of a specific sublineage were classified as having 0 cases. The upper limit for cluster detection was specified as 25% of the study population over each year. All maps were created in ArcGIS 10.2 (ESRI, Redlands, CA, USA).

2.11 Statistical analysis

2.11.1 Statistical analysis for chapter 3

Comparison of baseline characteristics patient groups, stratified by the H58 status or susceptibility category of their corresponding *S*. Typhi isolates was performed using the Kruskal Wallis test for continuous variables and Fisher's exact test for categorical variables. Time to treatment failure was analysed using Firth's penalized maximum likelihood bias reduction method for Cox regression as a solution for the non-convergence of likelihood function in the case of zero event counts in subgroups [278]. For comparisons between treatment arms, H58 status, or ciprofloxacin susceptibility group, the model included treatment arm, H58 status, or susceptibility group as a single covariate. Confidence intervals (CI) and *p*-values were calculated by profile-

penalized likelihood. FCT was analyzed as an interval-censored outcome, i.e. as the time interval from the last febrile temperature assessment until the first afebrile assessment, using parametric Weibull accelerated failure time models [279]. Median and inter-quartile range (IQR) FCT calculations for subgroups were based on models for each subgroup separately. Acceleration factors were based on models that included treatment arm as the only covariate. The non-parametric maximum likelihood estimator (NPMLE) was used to visualize the distribution of FCT between groups. Heterogeneity between subgroups was tested with models that included an interaction between treatment arm and the sub-grouping variable. All analyses were performed using R software version 3.2.2 [280].

2.11.2 Statistical analysis for chapter 5

Rates of hospitalized typhoid fever were calculated at the commune level using the population under the age of 15 years from 2008. Multivariable negative binomial regression was used to identify commune-level risk factors associated with the rate of cases per 1,000 population under the age of 15 years. Interaction between commune level factors was evaluated using the likelihood ratio test. Variables included in the evaluation of the final model included those with significant associations ($p<0.10$) in the univariate analysis and *a priori* sanitation and water source variables. Variables that did not add significantly to the fit of the final model (determined by the likelihood ratio test) were not included. All analyses were performed in STATA (v13, College Station, TX, USA) and plots were created in R v3.2.2 (R Foundation for Statistical Computing, Vienna, Austria, https://cran.r-project.org/) using ggplot2 [281].

Chapter 3

Emergence of a novel ciprofloxacin-resistant subclade of H58 *Salmonella* Typhi associated with fluoroquinolone treatment failure in Nepal

3.1 Introduction

Antimicrobial treatment is essential to effectively manage typhoid fever and avoid serious complications. However, antimicrobial resistance in *S.* Typhi has now become a major global health problem, limiting the treatment options and increasing the treatment failure rates. Over the last 30 years, the extensive use of antimicrobials for typhoid treatment has successively driven the emergence and global dissemination of resistant organisms [282]. First line drugs of choice such as chloramphenicol, amoxicillin and trimethoprim-sulfamethoxazole have been rendered ineffective due to the emergence of MDR *S.* Typhi since the 1980s. Fluoroquinolones were subsequently used in the 1990s and became standard drugs of choice after officially recommended by WHO for typhoid treatment in 2003 [33]. However, the effectiveness of this group of compounds has been diminished since *S.* Typhi isolates with reduced susceptibility to fluoroquinolones started to emerge [283]. Reduced susceptibility to fluoroquinolones in *S.* Typhi isolates is induced by point mutation(s) in quinolone resistance determining region in the *gyrA* gene and such organisms appear to have a fitness advantage, even in the absence of antimicrobial exposure [284]. It is now known that the widespread of *S.* Typhi isolates with reduced susceptibility to

fluoroquinolones has been partly facilitated by the dissemination of a specific MDR lineage (H58) across Asia and Africa [248]. H58 lineage has been replacing other genotypes rapidly and reshaping the global population structure of *S.* Typhi.

Previous studies have shown that patients infected with *S.* Typhi isolates with elevated Minimum Inhibitory Concentrations (MIC) against ciprofloxacin and ofloxacin are more likely to have prolonged fever clearance time (FCTs) and more frequent treatment failure [89,90,283,285]. Whilst older fluoroquinolones have showed a reduced efficacy in curing the disease, the fourth-generation fluoroquinolone, gatifloxacin, has been demonstrated as highly efficacious for uncomplicated typhoid, even in patients infected with *S.* Typhi isolates displaying reduced susceptibility to ciprofloxacin (MIC ≥0.125 μg/ml) [257,259,286,287]. However, during a randomised controlled trial (RCT) comparing gatifloxacin and ceftriaxone conducted recently in Nepal, there was an increased number of treatment failures associated with *S.* Typhi strains with high level of ciprofloxacin and gatifloxacin resistance (ciprofloxacin MIC >32μg/ml, gatifloxacin MIC >1μg/ml), prompting the data safety and monitoring board to stop the trial [288]. In this study, I aimed to investigate the molecular epidemiology of the *S.* Typhi isolates from this trial and understand how bacterial genotype may affect the treatment outcome. All *S.* Typhi isolates were genome-sequenced and stratified by genotype and the association between genotype and clinical presentation and outcome was assessed. The findings from my research were published in eLife journal in 2016 (appendix G) [289].

3.2 Results

3.2.1 *Salmonella* Typhi whole genome sequencing

I performed whole genome sequencing (WGS) on 78 *S.* Typhi isolates from patients in both RCT treatment arms (gatifloxacin and ceftriaxone) (Appendix A). The resulting phylogeny indicated that the majority of isolates (65/78; 83.3%) fell within the H58 lineage, while the remaining 13 (16.7%) represented eight different lineages (Figure 3.1). All but four of the H58 strains contained the common DNA gyrase (*gyrA*) mutation in codon 83 (S83F), which confers reduced susceptibility against fluoroquinolones (FQs) (ciprofloxacin MIC; 0.125- 0.5μg/ml) [290]. Nested within the S83F H58 group but separated from the rest of the group by a branch defined by 30 SNPs, was a H58 subclade comprised of 12 isolates containing the S83F *gyrA* mutation, a mutation in *gyrA* at codon 87 (D87N), and an additional mutation in the topoisomerase gene, *parC* (S80I) (H58 triple mutant). Notably, these H58 triple mutants shared high MICs against ciprofloxacin (≥24μg/ml). Further, an additional two non-H58 RCT isolates with ciprofloxacin MIC≥24μg/ml had the S83F *gyrA* mutation, an alternative mutation at codon 87 (D87V), the S80I *parC* mutation, and an A364V mutation in *parE* (Figure 3.1, Appendix A).

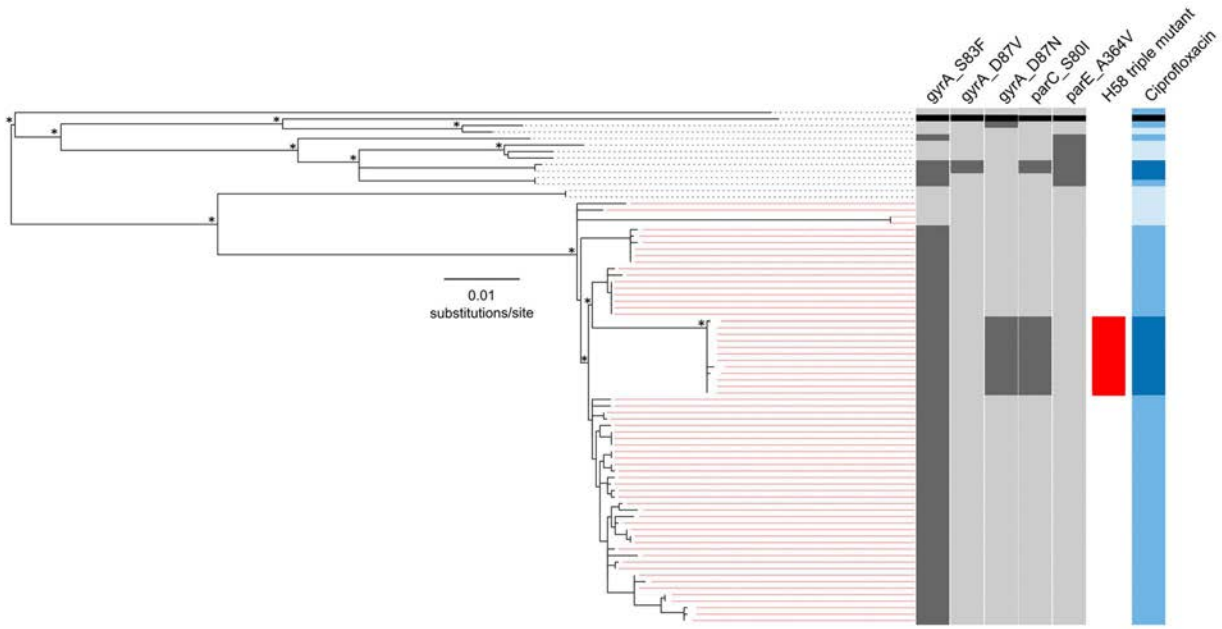

Figure 3.1 The phylogenetic structure of 78 Nepali *Salmonella* Typhi isolated during a gatifloxacin versus ceftriaxone randomised controlled trial

Maximum likelihood phylogeny based on core-genome SNPs of 78 *Salmonella* Typhi isolates with the corresponding metadata, including the presence of mutations (dark grey) in *gyrA* (S83F, D87V and D87N), *parC* (S80I) and *parC* (A364V) and susceptibility to ciprofloxacin (susceptible, light blue; intermediate, mid-blue and non-susceptible, dark blue) by Minimum Inhibitory Concentration (MIC). Missing metadata is indicated by black boxes. Red lines linking to metadata show isolates belonging to the *Salmonella* Typhi H58 lineage (with H58 triple mutants highlighted), other lineages (non-H58) are shown with black lines. The scale bar indicates the number of substitutions per variable site (see methods). Asterisks indicate ≥85% bootstrap support at nodes of interest.

3.2.2 Clinical presentation of *Salmonella* Typhi infections

I stratified clinical data from the RCT by H58 status of the corresponding *S*. Typhi isolates (H58; N=65, non-H58; N=13) and compared baseline characteristics between these groups. I found no significant differences in demographics and no association between disease severity at presentation between those infected with an H58 *S*. Typhi isolate or a non-H58 isolate (Table 3.1). Next, I compared the baseline characteristics of patients stratified by ciprofloxacin susceptibility (susceptible, intermediate and resistant), and found no differences in disease severity or demographics on presentation; the only exception being FQ resistant *S*. Typhi were more frequently isolated from adults (Table 3.2). A significantly lower proportion of H58 *S*. Typhi (4/65; 6.2%) were susceptible to FQs compared to non-H58 isolates (6/13; 46%) (p=0.001) (Table 3.3) and, overall, H58 isolates had significantly higher MICs against the majority of tested antimicrobials than non-H58 isolates (Table 3.3).

Table 3.1 Baseline characteristics by *Salmonella* Typhi lineage

Baseline characteristic	**n**	**Non-H58 (N=13)**	**n**	**H58 (N=65)**	***p* value***
Age (years) – median (IQR)	13	18.0 (13.0,21.0)	65	18.0 (13.0,22.0)	0.75
Sex (male)	13	12 (92.3%)	65	46 (70.8%)	0.17
Temperature (°C) - median (IQR)	13	39.0 (38.3,39.4)	62	39.0 (38.3,39.4)	0.77
Days of illness before enrolment - median (IQR)	13	5.0 (4.0,7.0)	65	5.0 (4.0,7.0)	0.39
Antimicrobials in last two weeks	13	1 (7.7%)	65	9 (13.8%)	1.00
Previous history of typhoid	13	1 (7.7%)	65	5 (7.7%)	1.00
Family history of typhoid	13	1 (7.7%)	65	8 (12.3%)	1.00
Typhoid vaccination	13	1 (7.7%)	65	0 (0%)	0.17
Fever	13	13 (100%)	64	64 (100%)	NA
Cough	12	4 (33.3%)	62	15 (24.2%)	0.49
Constipation	12	1 (8.3%)	63	4 (6.3%)	1.00
Headache	13	12 (92.3%)	64	59 (92.2%)	1.00
Diarrhoea	12	5 (41.7%)	62	24 (38.7%)	1.00
Vomiting	12	4 (33.3%)	63	16 (25.4%)	0.72
Abdominal pain	12	2 (16.7%)	62	18 (29.0%)	0.49
Anorexia	12	9 (75.0%)	64	53 (82.8%)	0.68
Nausea	12	9 (75.0%)	61	38 (62.3%)	0.52
Leucocyte count ($\times 10^9$/L) - median (IQR)	13	5.9 (4.5,6.6)	65	6.2 (4.8,7.3)	0.54
Neutrophils (%) - median (IQR)	13	65.0 (60,73.0)	65	70 (65.0,75.0)	0.17
Lymphocytes (%) - median (IQR)	13	33.0 (22.0,37.0)	65	28.0 (22.0,34.0)	0.32
Haematocrit (%) - median (IQR)	12	38.7 (37.1,41.2)	65	38.0 (35.0,41.4)	0.39
Platelet count ($\times 10^9$/L) - median (IQR)	13	177.0 (165.0,189.0)	65	158.0 (140,210)	0.24
AST (U/L) - median (IQR)	12	52.5 (38.2,64.8)	62	59.0 (44.0,83.8)	0.21
ALT (U/L) - median (IQR)	13	46.0 (37.0,57.0)	63	50 (35.5,66.5)	0.18

N refer to the number of patients in each group
n refers to the number of observations with non-missing data for the respective characteristic
*Comparisons between the two were done using Fisher's exact test for categorical variables and the Wilcoxon rank sum test for continuous variables

Table 3.2 Baseline characteristics by *Salmonella* Typhi ciprofloxacin susceptibility

Baseline characteristic	**n**	**Susceptible (N=10)**	**n**	**Intermediate (N=52)**	**n**	**Resistant (N=16)**	***p* value***
Age (years) – median (IQR)	10	14.0 (9.2,17.0)	52	18.0 (13.0,22.0)	16	20 (17.8,21.2)	0.05
Sex (male)	10	8 (80%)	52	39 (75.0%)	16	11 (68.8%)	0.86
Temperature (°C) - median (IQR)	10	39.0 (38.6,39.4)	49	39.0 (38.3,39.4)	16	38.8 (38.3,39.3)	0.42
Days of illness before enrolment - median (IQR)	10	4.0 (4.0,6.0)	52	5.0 (4.0,7.0)	16	5.0 (4.0,7.0)	0.88
Antimicrobials in last two weeks	10	1 (10%)	52	7 (13.5%)	16	2 (12.5%)	1.00
Previous history of typhoid	10	0 (0%)	52	6 (11.5%)	16	0 (0%)	0.34
Family history of typhoid	10	1 (10%)	52	6 (11.5%)	16	2 (12.5%)	1.00
Typhoid vaccination	10	1 (10%)	52	0 (0%)	16	0 (0%)	0.13
Fever	10	10 (100%)	51	51 (100%)	16	16 (100%)	1.00
Cough	10	2 (20%)	49	13 (26.5%)	15	4 (26.7%)	1.00
Constipation	10	0 (0%)	50	4 (8.0%)	15	1 (6.7%)	1.00
Headache	10	10 (100%)	51	45 (88.2%)	16	16 (100%)	0.34
Diarrhoea	10	3 (30%)	50	21 (42.0%)	14	5 (35.7%)	0.82
Vomiting	10	2 (20%)	50	14 (28.0%)	15	4 (26.7%)	0.93
Abdominal pain	10	1 (10%)	50	16 (32.0%)	14	3 (21.4%)	0.38
Anorexia	10	8 (80%)	51	41 (80.4%)	15	13 (86.7%)	0.91
Nausea	10	6 (60%)	48	31 (64.6%)	15	10 (66.7%)	1.00

N refers to the number of patients in each group
n refers to the number of observations with non-missing data for the respective characteristic
* Comparison between the three groups were done using Fisher's exact test for categorical variables and the Kruskal-Wallis test for continuous variables

Table 3.3 Comparison of antimicrobial susceptibility by *Salmonella* Typhi lineage

	Non-H58 (N=13)			H58 (N=65)			p value*
E test	**MIC50**	**MIC90**	**GM (range)**	**MIC50**	**MIC90**	**GM (range)**	
Amoxicillin	0.5	1	0.77 (0.38–38)	0.75	>256	1.43 (0.38–>256)	0.0412
Chloramphenicol	3	4	2.7 (1.5–8)	4	12	5.7 (2–>256)	0.0147
Ceftriaxone	0.06	0.06	0.06 (0.05–0.13)	0.09	0.19	0.11 (0.03–0.64)	0.0004
Gatifloxacin	0.13	0.25	0.06 (0.01–2)	0.13	2	0.21 (0.01–3)	0.1197
Nalidixic acid	>256	>256	21.6 (1–>256)	>256	>256	346.8 (1–>256)	0.0004
Ofloxacin	0.25	0.75	0.24 (0.03–>32)	0.5	>32	1.09 (0.03–>32)	0.0240
Trimethoprim sulphate	0.02	0.05	0.03 (0.02–0.05)	0.05	0.32	0.09 (0.01–>32)	0.0016
Ciprofloxacin	0.13	0.75	0.11 (0.01–>32)	0.38	>32	0.80 (0.02–>32)	0.0051
Ciprofloxacin susceptibility group							0.0008[#]
- Susceptible	6 (46.2%)			4 (6.2%)			
- Intermediate	4 (30.8%)			48 (73.8%)			
- Resistant	3 (23.1%)			13 (20.0%)			

*Comparisons between *Salmonella* Typhi lineage for MICs and ciprofloxacin susceptibility groups were based on the Wilcoxon rank sum test and Fisher's exact test, respectively.
MIC: minimum inhibitory concentration, measured in µg/ml
#p value for comparison of susceptible vs. intermediate/resistant combined between groups by Fisher's exact test is 0.001.
GM: geometric mean, the upper range of the values was determined by multiplying the MIC by 2 if the result was >X (for example, >256 = 256*2 = 512).

3.2.3 Treatment failure and fever clearance times

The primary endpoint of the RCT in which these data were generated was a composite for treatment failure (method section 2.4.1). Treatment failure with H58 *S.* Typhi was significantly less common in the ceftriaxone group (3/31; 9.7%) than the gatifloxacin group (15/34; 44.1%) (Hazard Ratio (HR) of time to failure 0.19, 95%CI 0.05-0.56, p=0.002) (Table 3.4). Conversely, there was no significant difference in treatment failure between those infected with non-H58 isolates treated with gatifloxacin (0/6; 0%) or ceftriaxone (2/7; 28.6%) (p=0.32). Similarly, time to fever clearance differed significantly between the two treatment groups in H58 infections, with median FCTs of 5.03 days (interquartile range (IQR): 3.18-7.21) in the gatifloxacin group and 3.07 days (IQR: 1.89-4.52) in the ceftriaxone group (p<0.0006). Again, this trend was not mirrored in the non-H58 *S.* Typhi infections, with FCTs of 2.87 (IQR: 2.08-3.7) and 3.12 (IQR: 2.2-4.12) days for gatifloxacin and ceftriaxone, respectively (p=0.61) (Table 3.4). Moreover, in the gatifloxacin arm, H58 *S.* Typhi tended to be associated with a higher risk of treatment failure (p=0.06) (Figure 3.2.A, Table 3.5) and a lengthier fever clearance time (p=0.013) (Figure 3.2.C, Table 3.4).

As I identified two non-H58 isolates that were also FQ-resistant (Figure 3.1), I additionally stratified outcome for the gatifloxacin arm (N=40 patients) by FQ susceptibility of the infecting organism. Those infected with FQ-resistant *S.* Typhi failed gatifloxacin treatment more frequently (8/10; 80%) than those infected with an intermediately resistant organism (7/25; 28%) or a susceptible organism (0/5; 0%) (p=0.007) (Figure 3.2.B, Tables 3.6 and 3.7). Furthermore, in the gatifloxacin arm,

those infected with FQ-resistant organisms had significantly higher median FCTs than those infected with *S.* Typhi with alternative FQ susceptibility profiles (median FCTs (days): susceptible, 2.96 (IQR: 2.13-3.85), intermediate, 4.01 (IQR: 2.76-5.37) and resistant 8.2 (IQR: 5.99-10.5), respectively (p<0.0001)) (Figure 3.2.D, Table 3.6). Comparatively, the median FCT for those infected with an FQ resistant organism but randomised to ceftriaxone was 3.83 days (IQR: 2.96-4.7) (p<0.0001 for the between-treatment comparison) (Table 3.6).

Table 3.4 Summary of time to treatment failure and fever clearance time by *Salmonella* Typhi lineage

Time to treatment failure	Gatifloxacin (events/N)	Ceftriaxone (events/N)	Hazard ratio of time to failure (95%CI); p value	Heterogeneity test (p value)
H58*				0.020
- H58	15/34	3/31	0.19 (0.05, 0.56); p=0.002	
- Non-H58	0/6	2/7	3.87 (0.31, 534.24); p=0.32	
Fever clearance time	**Gatifloxacin median (IQR) days**	**Ceftriaxone median (IQR) days**	**Acceleration factor (95%CI); p value**	**Heterogeneity test (p value)**
H58¥				0.07
- H58	5.03 (3.18, 7.21)	3.07 (1.89, 4.52)	1.59 (1.22, 2.09); p=0.0006	
- Non-H58	2.87 (2.08, 3.7)	3.12 (2.2, 4.12)	0.90 (0.59, 1.36); p=0.61	

*Likelihood ratio test p=0.06 and 0.40 for comparison of time to treatment failure between H58 vs. non-H58 groups in gatifloxacin arm only and in all patients, respectively

¥p=0.013 and p=0.029 for comparison of interval censored time to fever clearance between H58 vs. non-H58 groups in gatifloxacin arm only and in all patients, respectively

Table 3.5 Treatment failure in detail by *Salmonella* Typhi lineage in the gatifloxacin treatment group

Characteristic	Non-H58 (N=6)	H58 (N=34)	*p* value*
Treatment failure	0 (0%)	15 (44.1%)	0.06
- Fever past seven days	0 (0%)	7 (21.2%)	
- Rescue treatment required	0 (0%)	9 (27.3%)	
- Microbiological failure	0 (0%)	2 (6.5%)	
- Relapse within 28 days	0 (0%)	4 (14.8%)	
- Confirmed relapse within 28 days	0 (0%)	4 (14.8%)	
Any relapse during 6 month follow-up	1 (16.7%)	4 (11.8%)	0.62
Confirmed relapse during 6 month follow-up	1 (16.7%)	4 (11.8%)	

*Time to treatment failure and time to relapse were analyzed using Firth's penalized maximum likelihood bias reduction method for Cox regression and *p* values were calculated from likelihood ratio tests

Table 3.6 Summary of time to treatment failure and fever clearance time by ciprofloxacin susceptibility

Time to treatment failure	**Gatifloxacin (events/N)**	**Ceftriaxone (events/N)**	**Hazard ratio of time to failure (95%CI); p value**	**Heterogeneity test (p value)**
Ciprofloxacin susceptibility group[†]				0.08
- Susceptible	0/5	1/5	2.40 (0.13, 350.21); p=0.57	
- Intermediate	7/25	2/27	0.27 (0.05, 0.99); p=0.049	
- Resistant	8/10	2/6	0.27 (0.05, 1.01); p=0.052	
Fever clearance time	**Gatifloxacin median (IQR) days**	**Ceftriaxone median (IQR) days**	**Acceleration factor (95%CI); p value**	**Heterogeneity test (p value)**
Ciprofloxacin susceptibility group[‡]				0.015
- Susceptible	2.96 (2.13, 3.85)	4.78 (4.01, 5.5)	0.71 (0.49, 1.02); p=0.07	
- Intermediate	4.01 (2.76, 5.37)	2.63 (1.52, 4.05)	1.31 (0.97, 1.76); p=0.07	
- Resistant	8.2 (5.99, 10.5)	3.83 (2.96, 4.7)	2.23 (1.57, 3.17); p<0.0001	

[†]Likelihood ratio test p=0.007 for comparison of time to treatment failure between MIC groups in gatifloxacin arm only
[‡]p<0.0001 for comparison of interval censored time to fever clearance between MIC groups in gatifloxacin arm only

Table 3.7 Treatment failure in detail by ciprofloxacin susceptibility in the gatifloxacin treatment group

Characteristic	**Susceptible (N=5)**	**Intermediate (N=25)**	**Resistant (N=10)**	***p* value***
Treatment failure	0 (0%)	7 (28.0%)	8 (80%)	0.007
- Fever past seven days	0 (0%)	2 (8.0%)	5 (55.6%)	
- Rescue treatment required	0 (0%)	3 (12.0%)	6 (66.7%)	
- Microbiological failure	0 (0%)	0 (0%)	2 (22.2%)	
- Relapse within 28 days	0 (0%)	2 (9.5%)	2 (28.6%)	
- Confirmed relapse within 28 days	0 (0%)	2 (9.5%)	2 (28.6%)	
Any relapse during 6 month follow-up	1 (20%)	2 (8.0%)	2 (20%)	0.31
Confirmed relapse during 6 month follow-up	1 (20%)	2 (8.0%)	2 (20%)	

*Time to treatment failure and time to relapse were analyzed using Firth's penalized maximum likelihood bias reduction method for Cox regression and *p* values were calculated from likelihood ratio tests

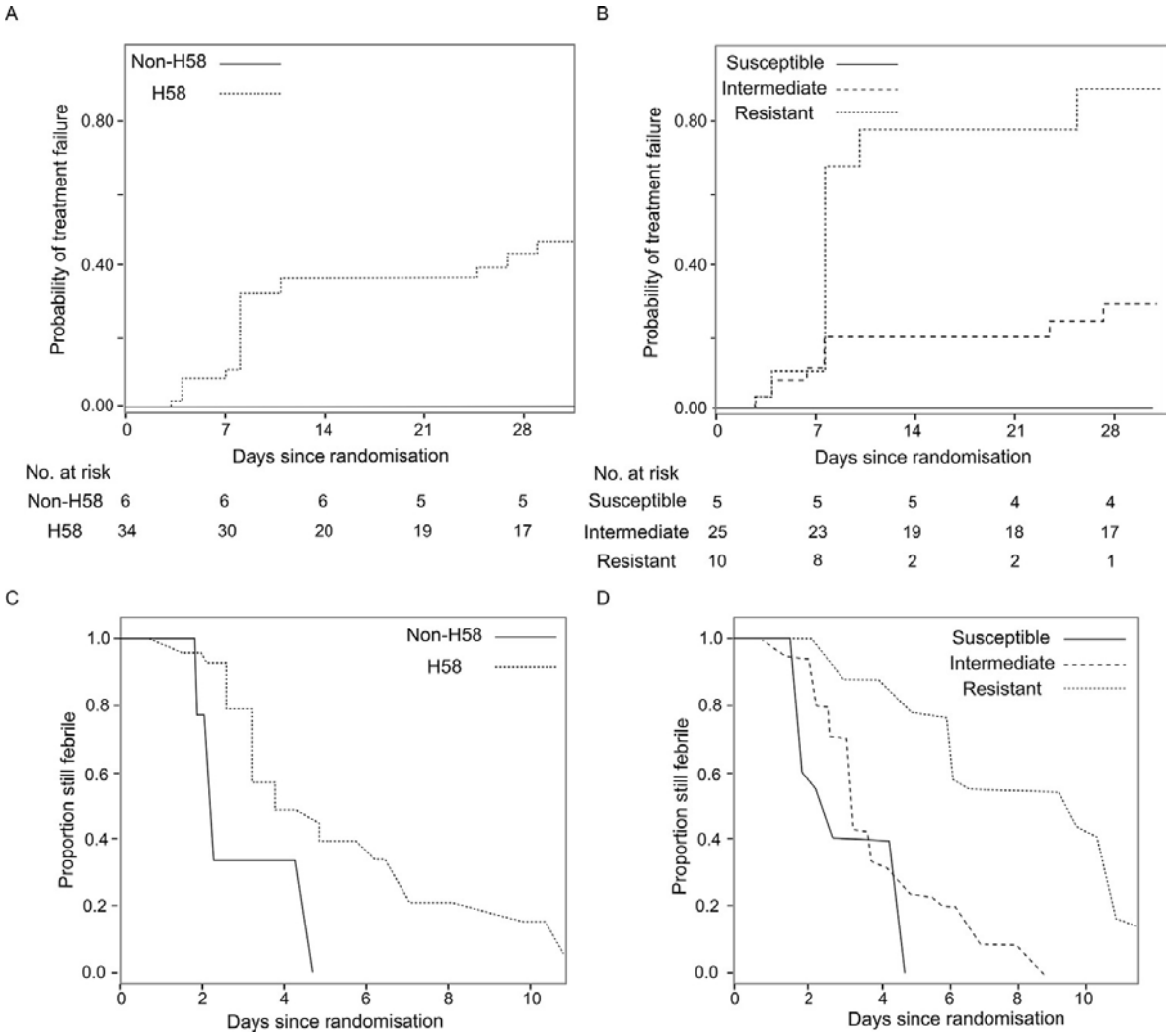

Figure 3.2 The association of *Salmonella* Typhi lineage and ciprofloxacin susceptibility with treatment failure and fever clearance time in patients randomised to gatifloxacin

A) Kaplan-Meier curve for time to treatment failure by H58 and non-H58 *Salmonella* Typhi. B) Kaplan-Meier curve for time to treatment failure by *Salmonella* Typhi susceptibility group (susceptible, intermediate, resistant) against ciprofloxacin. C) Non-parametric maximum likelihood estimators for interval-censored fever clearance time (see methods) by H58 and non-H58 *Salmonella* Typhi. D) Non-parametric maximum likelihood estimators for interval-censored fever clearance time by *Salmonella* Typhi susceptibility group (susceptible, intermediate, resistant against ciprofloxacin).

3.2.4 The emergence of fluoroquinolone-resistant *Salmonella* Typhi

Given the unusual long branch leading to the H58 triple mutant subclade on the phylogenetic tree (Figure 3.1), I hypothesised that the H58 triple mutants represented a contemporary importation into Nepal. To explore this, I compared the genomes of the 78 RCT *S.* Typhi isolates with those from 58 supplementary *S.* Typhi isolated from previous studies conducted between 2008 and 2013 in this setting (Figure 3.3, Appendix A). I found that the majority of the local H58 isolates (84/121; 69.4%) were closely related; these organisms represented "endemic" Nepali H58 clade containing a single S83F *gyrA* mutation. Additionally, I identified a further five Nepali organism isolated in 2013 that belonged to the H58 triple mutant group, and had an MIC ≥24 μg/ml against ciprofloxacin. Incorporating additional genome sequences from a recent international study of the H58 lineage [248], I found that all the Nepali H58 triple mutants were closely related (5 SNPs to nearest neighbour) to H58 triple mutants isolated in India between 2008 and 2012 (Figure 3.3).

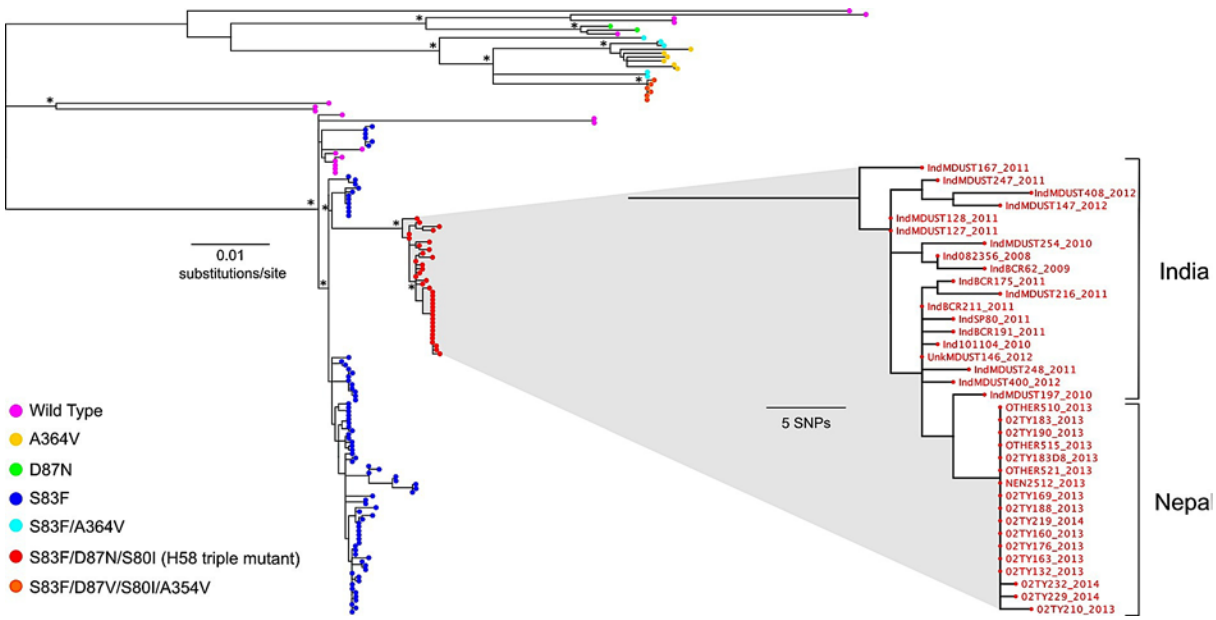

Figure 3.3 The phylogenetic structure of fluoroquinolone resistant *Salmonella* Typhi H58 in a regional context

Maximum likelihood phylogeny based on core-genome SNPs of 136 (78 from the RCT) *Salmonella* Typhi isolates from Nepal and neighbouring India (Appendix A). Main tree shows the overall phylogenetic structure and the presence of specific combinations of mutations in *gyrA* (S83F, D87V and D87N), *parC* (S80I) and *parE* (A364V). The inset shows a magnified view of the fluoroquinolone resistant *Salmonella* Typhi H58 triple mutants from Nepal and their close association with similarly fluoroquinolone resistant *Salmonella* Typhi H58 triple mutants from India. The scale bar on the primary tree indicates the number of substitutions per variable site, while that in the inset indicates genetic distance in number of SNPs. Asterisks indicate ≥85% bootstrap support at nodes of interest

3.3 Discussion

My study shows that a new FQ resistant subclade of H58 *S.* Typhi has been introduced into Nepal and is associated with a lack of FQ efficacy. This subclade was associated with longer FCTs and treatment failure in patients treated with the FQ, gatifloxacin. For the first time I can conclusively show how enteric fever patients respond to FQ treatment when infected with a specific subclade of H58, thereby linking organism genotype with a treatment phenotype. Given the international significance of FQs for the treatment of enteric fever and other bacterial infections, my findings have major global health implications for the use of this group of antimicrobials.

A single *S.* Typhi organism with the same combination of triple mutations was isolated in Nepal in 2011 [291]. This isolate was also associated with treatment failure, although, this organism was not genome sequenced and was assumed to be an isolated case. More significantly, several closely related bacteria were genome sequenced during an international study of H58 *S.* Typhi [248]. These organisms had the same combination of triple FQ resistance mutations as those described here; my analysis shows they belong to the same subclade of H58. These isolates had equivalently high MICs against ciprofloxacin and were isolated in India between 2008 and 2012. However, there were no associated outcome data for these strains and other reports from India have been limited. My data strongly suggests this lineage was recently introduced into Nepal from India (or nearby) and has since entered in an endemic

transmission cycle in Kathmandu. Given the large extent of human movement between India and Nepal, I propose this is a likely route of introduction.

Appropriate antimicrobial therapy is critical for enteric fever, as effective drugs curtail symptoms and prevent life threatening complications. My data has major repercussions for enteric fever treatment, and I advocate that FQs should no longer be used for empirical enteric fever therapy on the Indian subcontinent, as I predict these organisms are now widespread and may be associated with poor outcomes. Notably, one of the arms in the RCT used the newer generation FQ, gatifloxacin, which binds to a different location on the DNA gyrase than the older FQs and is not as susceptible to the common resistance mutations [292]. The isolates in this study were not generally resistant to gatifloxacin according to the current CLSI guidelines [293]; I suggest that these guidelines be modified to reflect these new clinical data. I additionally propose that *S.* Typhi genotyping, mapping and susceptibility testing is performed routinely and rapidly in reference laboratories outside South Asia to monitor the international spread of this lineage and to ensure the provision of alternative efficacious therapies to returning travellers [294,295]. In cases of infection with these FQ-resistant isolates, I suggest that ceftriaxone and azithromycin are used as alternatives, and do not currently recommend a return to the use of first-line drugs without contemporary data on treatment outcome. Whilst none of the isolates in this study were MDR, a rapid recrudescence of MDR strains may occur if we return to older first-line alternatives.

This study has limitations. First, the clinical data was collected from one study in a single location, thus limiting utility outside this setting. Second, the overall sample

size (and the gatifloxacin group subsampling) of those with culture-positive *S.* Typhi-associated enteric fever was relatively small. Despite these limitations, I was able to show a highly significant association between disease outcome and susceptibility profile of the infecting organism. Further, by using WGS I was able to pinpoint causative mutations, identify the subclade responsible for treatment failure and relate these strains to other isolates circulating outside Nepal. The methodologies presented here, in which clinical outcome data is combined with genome sequences and antimicrobial susceptibility data, should become the gold standard for informing empiric treatment for enteric fever and understanding the role of bacterial genotype and resistance profile on disease outcome for other bacterial infections.

In conclusion, my data show a significant association between *S.* Typhi genotype, antimicrobial susceptibility and disease outcome for those treated with gatifloxacin in a cohort of Nepali enteric fever patients. A FQ-resistant variant of Typhi H58 has emerged in Nepal and is associated with the clinical failure of FQs. My data suggest these isolates are likely widespread in the subcontinent and FQs should not be recommended for empirical enteric fever therapy in this setting.

Chapter 4

The phylogenetics and spatiotemporal dynamics of *Salmonella* Paratyphi A in Kathmandu, Nepal

4.1 Introduction

Even though typhoid fever has been studied for more than one hundred years, it still causes significant disease burden in South and Southeast Asian countries, specifically in locations with limited access to clean water and improved sanitation. In the last decade, there have been major advances in typhoid research, which have provided a better understanding of the disease epidemiology and risk factors associated with typhoid fever as well as underlining the emergence and global dissemination of antimicrobial resistant organisms [138,248]. These studies have had a major impact on the disease management and public health interventions. However, recent advances and most typhoid research have been focused mainly on *S.* Typhi, as this pathogen is considered to be more important than *S.* Paratyphi A, particularly with respect to disease magnitude and severity [296]. This gap in knowledge and a lack of public attention have rendered *S.* Paratyphi A a neglected pathogen for many years. Since the turn of the 21st century, *S.* Paratyphi A has emerged at an unprecedented rate in many Asian countries including India, Pakistan, Nepal, Bangladesh, China, and Indonesia, becoming a increasingly important causative agent of typhoid (enteric) fever [139]. Clinical and epidemiological investigations have revealed that typhoid fever caused by *S.* Paratyphi A is clinically indistinguishable from *S.* Typhi; furthermore, *S.*

Paratyphi A also displays dissimilar epidemiological features to *S.* Typhi infection, with infection more likely to arise outside the household [297,142]. These studies highlight an underestimation of the clinical relevance of *S.* Paratyphi A infection and questions if the current control and preventive measures for typhoid would also protect against this pathogen. Currently, there is no vaccine against *S.* Paratyphi A and the licensed typhoid (*S.* Typhi) vaccines do not provide cross protection against *S.* Paratyphi A and are likely to be less effective for controlling typhoid fever in areas where these pathogens are co-circulating.

In comparison to *S.* Typhi, there have been growing evidence suggesting that *S.* Paratyphi A is more likely to develop resistance to nalidixic acid, an indicator of reduced susceptibility to fluoroquinolones (the first line drug recommended by WHO) and they are generally less susceptible to various classes of antimicrobial frequently used as alternative treatments [298]. *S.* Paratyphi A cases associated with organisms that are fully resistant to ciprofloxacin (MIC of 8 μg/ml), ceftriazone (MIC >16 μg/ml) and azithromycin (MIC >64 μg/ml) are also not uncommon [299,221,225]. In endemic areas where there is a lack of safe drinking water and inadequate food hygiene, case detection and antimicrobial therapy become crucial for the management of typhoid fever. However, limited access to laboratory facilities, low sensitivity and specificity of current diagnostic methods and similar clinical manifestations caused by *S.* Paratyphi A and *S.* Typhi have imposed significant challenges for the diagnostic accuracy as well as the appropriate antimicrobial therapy [300]. As a result, the actual burden of *S.* Paratyphi A infection remains unknown in many settings and empiric antimicrobial treatment may be problematic. At present, available epidemiological

data of *S.* Paratyphi A is very limited and remains unclear; moreover, conventional epidemiological approaches only describe the demographic and epidemiological features of reported cases without any characterization of the bacterial population structure and dynamics. Understanding the structure and changing dynamics of the pathogen population in time and space, as well as the driving forces governing these in an epidemiological context, is essential for providing compelling evidence and novel insights into the source of infection and the patterns of typhoid transmission [301]. The fact that *S.* Paratyphi A isolates exhibit little genetic diversity makes whole-genome sequencing technology a fundamental approach for understanding the population of this organism. However, this approach remains a big challenge in many endemic settings.

Kathmandu, Nepal, is a highly endemic location for typhoid fever, and *S.* Paratyphi A infections have increased in recent years. During a period of sustained blood culture surveillance for typhoid fever at Patan hospital in Kathmandu between 1993 and 2003 the proportion of typhoid fever caused by *S.* Paratyphi A increased annually with the highest isolation rate up to 34% of all cases [298]. From 2005-2009, the isolation rate of *S.* Paratyphi A remained high and accounted for 31.5% of total confirmed typhoid cases; moreover, for some months, the number of *S.* Paratyphi A and *S.* Typhi infections was comparable [302]. Despite the overall decline of typhoid fever during this study period, *S.* Paratyphi A emerged rapidly and coincided with ciprofloxacin non-susceptibility. A previous case-control study demonstrated that *S.* Paratyphi A infection rate increased with age and was higher among those consuming street food,

suggesting that direct contact and consumption of contaminated food may play a significant role in the transmission of *S.* Paratyphi A in this region [129].

Approximately 1-5% of typhoid patients become chronic carriers and these people can act as reservoirs for the persistence and transmission of typhoid fever, especially if they are food handlers. Chronic carriage of *S.* Paratyphi A has been reported, and the prevalence was 1.6% among those undergoing cholecystectomy in Nepal [130]. This is comparable to the chronic carrier rate of *S.* Typhi, even though the number of acutely infected *S.* Paratyphi A patients was lower during the same period [130]. Little is known about genetic relatedness between acute and chronic *S.* Paratyphi A isolates as well as the relative role of chronic carriage in the transmission cycle of this pathogen. Additionally, all previous epidemiological investigations of *S.* Paratyphi A infections in Nepal have not included a detailed characterization of bacterial population and, as a result, have overlooked disease transmission dynamics. Here, I have, for the first time, performed a detailed molecular epidemiological study, in which I utilized whole genome sequencing to assess the phylogenetics and populations structure of *S.* Paratyphi A and combined this highly resolved genetic information with individual GPS location to interrogate the spatiotemporal dynamics of *S.* Paratyphi A infections in Nepal. I additionally reconstructed a Nepalese *S.* Paratyphi A phylogeny in a global context to further investigate into the dynamics of this bacterial population. This was also the first study characterizing the genetic relationship between carrier and acute *S.* Paratyphi A isolates to provide new insights into the role of typhoid carriage in this setting.

4.2 Results

4.2.1 Baseline characteristics

Between 2005 and 2014, 223 *S.* Paratyphi A isolates were collected and whole-genome sequenced, of which 92% (206/223) were from acutely infected patients and the remainder were from the gallbladder of carriers (8%, 17/223). The median age of acute *S.* Paratyphi A patients was 19 years (IQR: 13-24 years), while the median age of asymptomatically infected individuals was 37 years (IQR: 32-44 years) ($p<0.001$, Mann-Whitney U test). Over the observed time period, the median age of acutely infected patients did not vary substantially ($p=0.120$, Mann-Whitney U test). 71% (146/206) of the acute *S.* Paratyphi A patients were male, this contrasted with asymptomatically infected individuals, of which only 25% (4/16) were male ($p<0.001$, Fisher's exact test). There were no major differences in sex distribution between patients enrolled in acute studies ($p=0.251$, Fisher's exact test). The spatial distribution demonstrated that most *S.* Paratyphi A patients lived in the central Kathmandu Valley, specifically within Lalitpur; the carriers were more likely to live further away from the centre of the city (Figure 4.1). This is likely due to a lack of cholecystectomy facilities available outside of central Kathmandu.

A clinical history was collected for all acutely infected patients. The majority of patients reported a history of headache (93%) and anorexia (67%). Diarrhoea (13%) and constipation (12%) were rare. There were also no major differences between years were present for any symptom. Of 162 *S.* Paratyphi A patients with recorded water use information, only about half (52%, 85/162) reported using the municipal supply water as their main source. Other water sources included stone spout (14%,

23/162), sunken well (10%, 17/162), bottled water (12%, 19/162). Over a third (35%, 57/161) of patients reported using untreated water for drinking, with an additional 35% (57/161) reporting the use of a filter and 14% (23/161) reporting boiling as their primary treatment method. Additionally, a total of 52% (12/23) of patients reporting stone spout use did not treat the water prior to drinking, compared to 24% (20/85) of those using municipal supply water.

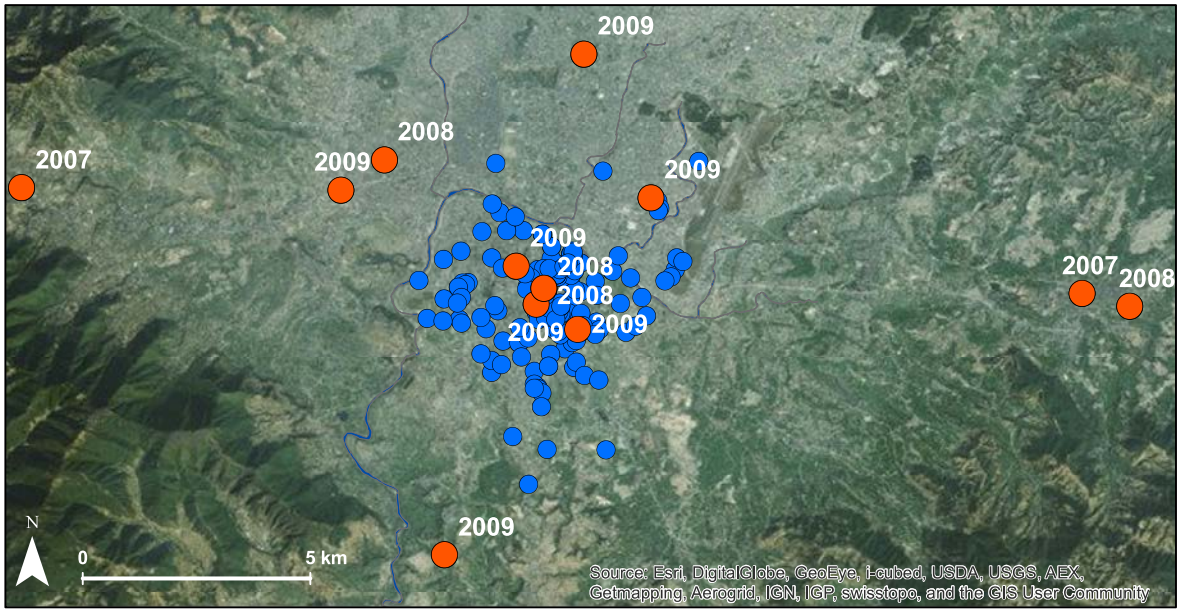

Figure 4.1 Locations of *Salmonella* Paratyphi A isolates in Kathmandu

Map shows the locations of *S*. Paratyphi A in Kathmandu between 2005 and 2014 with carrier isolates shown in red and acute isolates in blue. Note: 4 additional carrier isolates are not shown on the map due to the fact that they are 70km from central Kathmandu

4.2.2 Antimicrobial resistance

MICs of all *S.* Paratyphi A isolates were determined as described in section 2.2.3. Overall, *S.* Paratyphi A isolates were susceptible to all first-line antimicrobials including chloramphenicol, amoxicillin, and cotrimoxazole. However, 93% (208/223) of these isolates displayed resistance to nalidixic acid. Additionally, of those tested, 8% (14/178) were resistant to ciprofloxacin (MIC ≥1 μg/ml) and 39% (69/178) were resistant to ofloxacin (MIC ≥ 2 μg/ml) according to the updated CLSI breakpoints. None were resistant to ceftriaxone and gatifloxacin. As shown in Figure 4.2, MICs against ciprofloxacin, ofloxacin and gatifloxacin did not change significantly over time. In contrast, MICs to other first-line antimicrobials remained largely constant over the study period. MICs to azithromycin declined between 2006-2010 (Pearson correlation coefficient = -0.387, $p<0.001$). Additionally, the median MIC to azithromycin was significantly higher in organisms from acutely infected patients compared to carriers ($p=0.004$, Mann-Whitney U test). Nalidixic acid resistance was also more common in *S.* Paratyphi A isolates from acute patients (96%, 198/206) than the asymptomatic individuals (71%, 12/17) (Chi-square test, $p<0.001$). There were no significant differences in median MICs to ciprofloxacin and ofloxacin between these two groups. An *in silico* resistome analysis also found that none of *S.* Paratyphi A isolates in Nepal carried any plasmid or gene cassette associated with antimicrobial resistance.

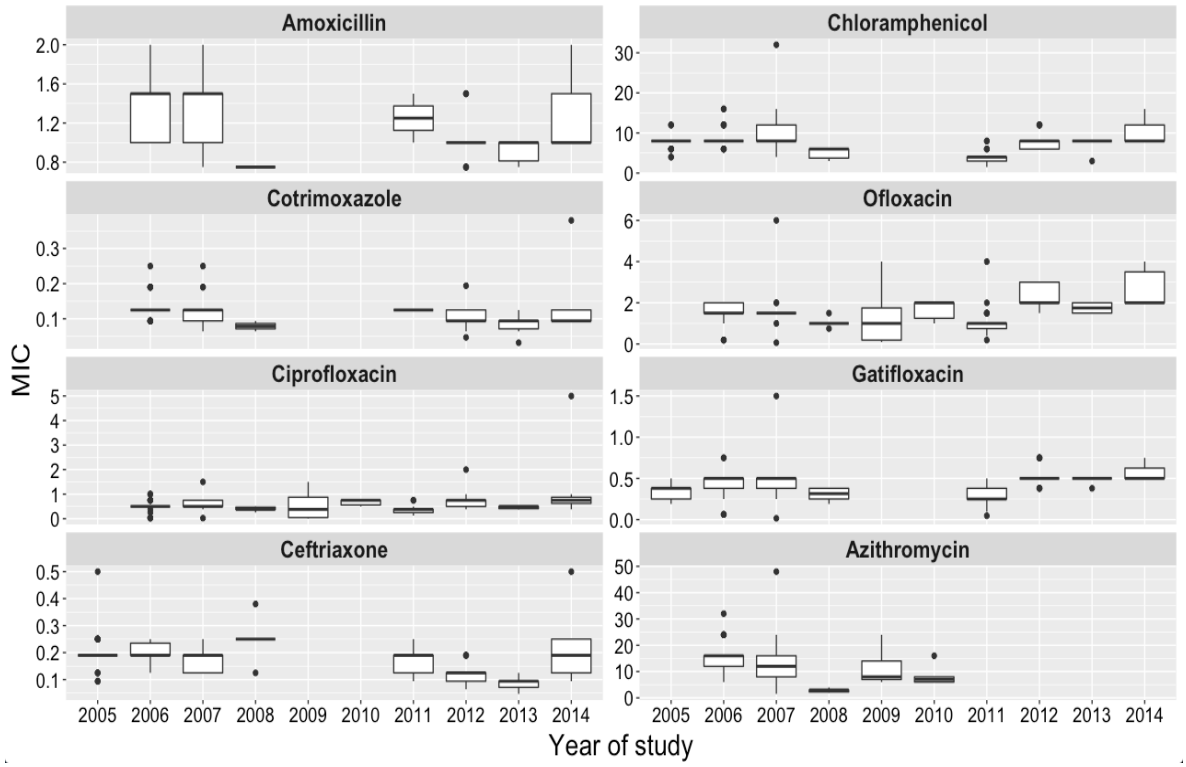

Figure 4.2 Minimal inhibitory concentrations (MICs) of *S.* Paratyphi A isolates to various antimicrobials over time

4.2.3 The population structure and dynamics of *S*. Paratyphi A isolates in Nepal

To investigate the population structure and dynamics of *S*. Paratyphi A in Nepal, I reconstructed the phylogeny of all Nepalese *S*. Paratyphi A isolates along with 111 genomes from a global collection. These data showed that the majority of Nepalese isolates (94,2%, 210/223) fell within global lineage A, which contained organisms circulating mostly within South Asian countries such as India, Nepal, and Pakistan (Figure 4.3). A small proportion of Nepalese *S*. Paratyphi A (5.4%, 12/223) also belonged to global lineage C. Unlike lineage A, lineage C has successfully spread globally and contains organisms from a wide range of geographical areas including South Asia (India, Nepal, Pakistan, Sri Lanka), Southeast Asia (Cambodia, Vietnam, Thailand, Indonesia), Middle East (Turkey), East Asia (China) and Africa (Ghana, Chad, Senegal, Guinea, Mali, Morocco, Egypt). Only 1/223 (0.4%) *S*. Paratyphi A isolate in Nepal grouped within global lineage F, which included mostly historic isolates.

Noticeably, between 2005 and 2011 I observed the phenomenon of clonal expansion of nalidixic resistant Nepalese *S*. Paratyphi A within global lineage A, which was designated as sub-lineage A1. This sub-lineage constituted 64% (135/210) of Nepalese isolates belonging to lineage A; 98.5% (133/135) had a mutation in codon 83 of *gyrA*, changing serine to phenylalanine. These organisms were resistance to nalidixic acid. From the global collection, there were one Indian and one Nepalese isolates recovered in 1999 and 2000, respectively also within sub-lineage A1; however, these two isolates were susceptible to nalidixic acid. As a result, this resistance associated mutation in *gyrA* likely developed and subsequently become

fixed in the bacterial population between 2000 and 2005. All acute *S*. Paratyphi A organisms isolated after 2005 had this mutation, suggesting a strong selective advantage for nalidixic acid resistance. Additionally, in comparison to the reference genome AKU_12601 (lineage C), isolates from sub-lineage A1 did not carry the SPA-2 sopE phage (SSPA2377 to SSPA2423) and the prophage region SPA-3 (SSPA2424 to SSPA2446). Alternatively, these organisms had acquired a novel intact prophage inserted in the chromosome between CDSs SSPA3930 and SSPA3931. This prophage was 33.8 kb in size and mostly closely related to phage P88 (NC_026014) (Figure 4.4). This novel prophage carried approximately 50 proteins, including a variant of the *sopE* gene, which exhibited 220/240, 222/240 and 221/240 amino acid identity to the *sopE* gene present in SPA-2 phage of the *S*. Paratyphi A reference strain AUK_12601 (FM200053); prophage of *S*. Typhimurium SL1344 (NC_016810.1) and SPI-7 in *S*. Typhi CT18 (AL513382), respectively. SopE is an effector protein secreted via a type III secretion system (SPI-1) and plays an important role in bacterial invasion into non-phagocytic cells by activating the actin cytoskeleton rearrangements and stimulating membrane ruffling. SopE-expressing *S*. Typhimurium has been shown to increase virulence and organisms carrying this additional gene have been associates with several epidemics [303]. It is unclear whether this novel sopE prophage modulates the virulence of these *S*. Paratyphi A. In addition to gaining resistance to nalidixic acid and a novel sopE prophage, there were also a number of non-synonymous mutations and gene disruptions associated with sub-lineage A1 (Table 4.1).

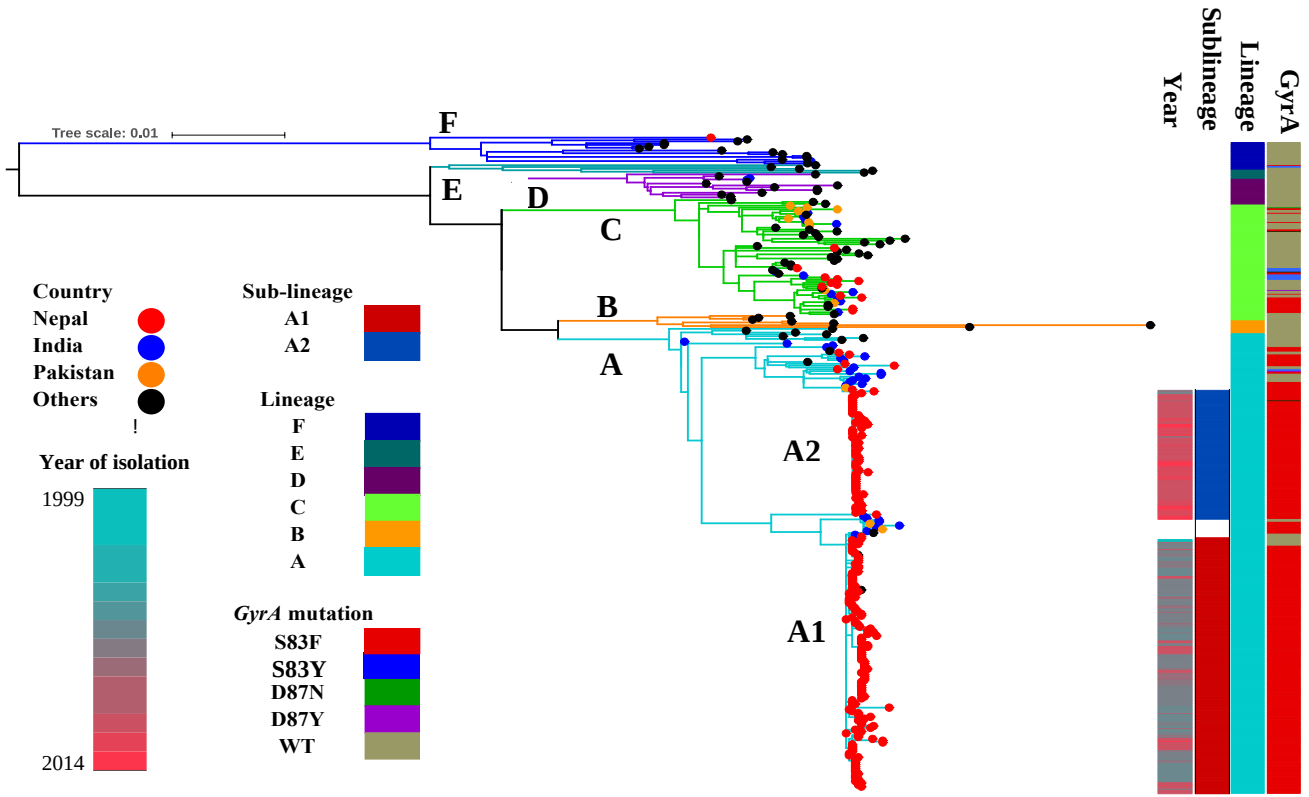

Figure 4.3 Phylogenetics of Nepalese *S.* Paratyphi A isolates in a global context

Maximum likelihood phylogenetic tree of 223 Nepalese *S.* Paratyphi A isolates together with 111 global *S.* Paratyphi A isolates and their corresponding metadata. The tree consisted of six major lineages (A-F) corresponding to different colored branches. The terminal nodes exhibit the country of isolates with isolates originating from Nepal and neighboring countries highlighted. The bars show the year of isolation, two major sublineages A1 (red) and A2 (blue) within lineage A, the division of six major lineages and the presence of gyrA mutations. The scale bar indicates the number of substitutions per variable site.

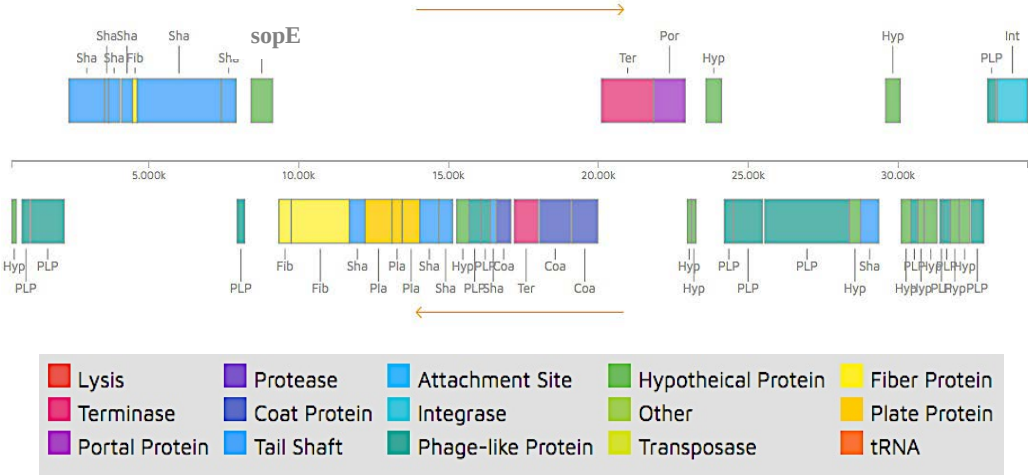

Figure 4.4 Novel sopE prophage of sub-lineage A1

Table 4.1 Nonsynonymous mutations and indels associated with sub-lineage A1

Position in FM200053	CDS	Product	Mutation	Ref	Alt
253654	**SSPA0212**	2,3,4,5-tetrahydropyridine-2-carboxylate N-succinyltransferase	Nonsynonymous	C	T
860533	**SSPA0723**	dTDP-glucose 4,6-dehydratase	Nonsynonymous	C	T
2753956	**SSPA2470**	4-aminobutyrate aminotransferase	Nonsynonymous	G	A
1346343	**SSPA1187**	putative NADP-dependent oxidoreductase	Nonsynonymous	C	T
1690227	**SSPA1517**	ABC transporter ATP-binding subunit	Nonsynonymous	G	A
1881695	**SSPA1689**	30S ribosomal protein S1	Nonsynonymous	T	A
2899096	**SSPA2613**	sulfite reductase (NADPH) flavoprotein beta subunit	Nonsynonymous	A	C
3135049	**SSPA2833**	conserved hypothetical protein	Nonsynonymous	G	A
4441670	**SSPA3972**	putative innner membrane protein	Nonsynonymous	G	A
73777	**SSPA0061**	CitG protein	Nonsynonymous	C	T
482924	**SSPA0385**	ethanolamine ammonia-lyase light chain	Nonsynonymous	C	T
2311942	**SSPA2069**	conserved hypothetical protein	Nonsynonymous	G	A
2600568	**SSPA2336**	DNA polymerase III epsilon subunit	Nonsynonymous	C	T
2794835	**SSPA2511**	putative sigma-54-dependent transriptional regulator	Nonsynonymous	G	A
4023577	**SSPA3609**	putative lipoprotein	Nonsynonymous	C	T
4295443	**SSPA3826**	anaerobic C4-dicarboxylate transporter	Nonsynonymous	A	C
4332842	**SSPA3866**	putative amino acid permease	Nonsynonymous	C	A
679384	**SSPA0561**	outer membrane protein C	Nonsynonymous	A	C
4569233	**SSPA4083**	putative two-component response regulator (creB)	Nonsense	C	T
3631814	**SSPA3280**	xylose operon regulatory protein	Insertion	GT	GGT
2809585	**SSPA2524**	formate hydrogenlyase subunit 2	Deletion	CGGGCGGCGC	C

Nonsynonymous mutations and indels in sub-lineage A1 were identified relative to the inferred most recent common ancestor of sub-lineage A1 and sub-lineage A2.

The genomics data suggested that, after clonal expansion, sub-lineage A1 gradually began to disappear from the population after 2011 and was replaced by the distantly related sub-lineage, A2 (Figure 4.3). The main branch of sub-lineage A2 differed by 96 SNPs from sub-lineage A1 and contained 67/210 (32%) of the Nepalese isolates belonging to lineage A. Isolates belonging to sub-lineage A2 was first detected in 2008 and initiated a clonal expansion, which replaced all other lineages from 2011. Sub-lineage A2 became the most dominant genotype accounting for 83-96% of isolates annually between 2012 and 2014. The median pairwise SNP distance of the Nepalese isolates within sub-lineage A2 was only 2 SNPs (IQR: 1-4 SNPs), which was lower than the median pairwise SNP distance of isolates belonging to sub-lineage A1 (5 SNPs, IQR: 4-7 SNPs), indicating a very recent clonal expansion of sub-lineage A2. All Nepalese isolates in sub-lineage A2 also had a Ser83Phe mutation in *gyrA* and were resistant to nalidixic acid. In comparison to the reference strain AKU_12601, isolates within sub-lineage A2 carried a comparable complement of prophage regions and there was no apparent difference in gene content between the first A2 isolate in 2008 and the remaining isolates in later years of this sub-lineage. As shown in Table 4.2, there were various non-synonymous mutations and gene disruptions associated with sub-lineage A2. Gene disruptions included SSPA0470, which encodes for a conserved hypothetical across *Salmonella enterica* and SSPA2905 (*tdcD*), encoding for propionate kinase. The Tdc operon (*tdcABCDEFG*) consists of seven genes which encode enzymes essential for the degradation of short-chain fatty acids, which serve as an alternative carbon and energy source in the absence of a preferred nutrient in human gut. Additionally, a number of non-synonymous mutations were located in genes involved in bacterial virulence, LPS and

gene regulation, such as secreted protein SifA, an effector protein of the SPI-2 Type 3 secretion system, which plays an important role in *Salmonella* virulence [304], WaaP (SSPA3336) involved in lipopolysaccharide core biosynthesis, signal peptidase I (LepAB), sensor protein KdpD in the two-component KdpD/KdpE regulatory system [305]. Details regarding all nonsynonymous mutations and indels associated with sub-lineage A2 are shown in Table 4.2.

Table 4.2 Nonsynonymous mutations and indels associated with sub-lineage A2

Position in FM200053	CDS	Product	Mutation	Ref	Alt
2176873	SSPA1948	rare lipoprotein A precursor	Nonsynonymous	A	G
2412607	SSPA2166	proline-specific permease ProY	Nonsynonymous	C	T
475620	SSPA0379	putative aldehyde dehydrogenase	Nonsynonymous	T	C
1671505	SSPA1500	putative MutT-family protein	Nonsynonymous	G	A
3035334	SSPA2730	conserved hypothetical protein	Nonsynonymous	G	A
1978639	SSPA1777	hypothetical ABC transporter ATP-binding protein	Nonsynonymous	A	G
3156617	SSPA2852	possible lipoprotein	Nonsynonymous	T	A
4160748	SSPA3728	phosphoribosylaminoimidazolecarboxamide formyltransferase and IMP cyclohydrolase (bifunctional enzyme)	Nonsynonymous	C	A
320372	SSPA0268	signal peptidase I	Nonsynonymous	G	A
2118532	SSPA1899	membrane-associated protein kinase KdpD	Nonsynonymous	G	C
2118529	SSPA1899	membrane-associated protein kinase KdpD	Nonsynonymous	C	T
1035960	SSPA0902	aspartyl-tRNA synthetase	Nonsynonymous	C	T
1798103	SSPA1620	putative 4-hydroxyphenylacetate permease	Nonsynonymous	T	C
2936045	SSPA2640	L-serine dehydratase 2 (L-serine deaminase 2)	Nonsynonymous	G	A
825360	SSPA0698	conserved hypothetical protein	Nonsynonymous	C	T
3321069	SSPA3005	oxaloacetate decarboxylase beta chain	Nonsynonymous	C	G
3204223	SSPA2895	putative membrane protein	Nonsynonymous	T	C
999519	SSPA0865	putative glucose-6-phosphate dehydrogenase	Nonsynonymous	C	T
1684077	SSPA1511	putative virulence determinant	Nonsynonymous	G	A
3109130	SSPA2806	possible transferase	Nonsynonymous	G	A
1133549	SSPA1000	alanine racemase	Nonsynonymous	C	T
1106787	SSPA0973	PTS system, mannose-specific IIAB component	Nonsynonymous	T	C
3289612	SSPA2977	nitrogen regulatory IIA protein	Nonsynonymous	A	G
279714	SSPA0232	putative secreted chitinase	Nonsynonymous	A	T
1216127	SSPA1073	anthranilate synthase component I	Nonsynonymous	G	A
3210225	SSPA2900a	conserved hypothetical transport protein (pseudogene)	Nonsynonymous	G	A
3700432	SSPA3336	lipopolysaccharide core biosynthesis protein	Nonsynonymous	C	T
578579	SSPA0470	conserved hypothetical protein	Nonsense	G	T
3215599	SSPA2905	propionate kinase tdcD	Deletion	AT	T

Nonsynonymous mutations and indels in sub-lineage A2 were identified relative to the inferred most recent common ancestor of sub-lineage A1 and sub-lineage A2.

4.2.4 Genetic relatedness between acute and carriage *S*. Paratyphi A isolates in Nepal

The majority of *S*. Paratyphi A carriage isolates belonged to sub-lineage A1 (53%, 9/17) and clustered within the acute isolates, of which 8/9 were resistant to nalidixic acid and had the same mutation Ser83Phe in *gyrA*. These carriage isolates probably resulted from recent asymptomatic infections for the following reasons. First, as mentioned above, the resistance mutation Ser83Phe in *gyrA* likely occurred and subsequently become fixed in the bacterial population after 2005. Second, these carriage isolates were collected between 2007 and 2009. There was one nalidixic acid susceptible isolate from a carrier (GB624) that was collected in 2009 and clustered within the global nalidixic acid susceptible isolates. The duration of asymptomatic carriage by this isolate is at least 5 years considering the establishment of nalidixic resistant sub-lineage A1 from 2005. The median pairwise SNP distance between acute isolates within this sub-lineage A1 was 5 SNPs (IQR: 3-7 SNPs), which was significantly lower than the median pairwise SNP distance between carrier isolates (7 SNPs, IQR: 5-7 SNPs) (Wilcoxon rank sum test, p= 0.003). My data demonstrated that the genetic diversity within the carriage isolates was higher compared to acute isolates in sub-lineage A1 considering the number of acute isolates was much higher.

Additionally, there was one carrier isolate collected in 2009 (GB672) that clustered within sub-lineage A2. This isolate was resistant to nalidixic acid (mutation Ser83Phe) and was only 1 SNP different from the first acute isolate collected in 2008, therefore this organism likely originated from recent asymptomatic infection. Three other carriage isolates (GB193, GB640, and GB726) belonged to global lineage A,

but they were distantly related to Nepalese isolates and clustered tightly with Indian isolates. The remaining 4 carrier isolates belonged to global lineage C (3) and lineage F (1).

Overall, my data demonstrated that the carriage isolates belonged to a wide range of genotypes; the most common genotype was sub-lineage A1. Additionally, the genotypic distribution of carrier isolates reflected the general genetic structure in the bacteria population. I found sufficient evidence to suggest that half of carrier isolates likely originated from recently asymptomatic carriage. Further, the fact that carriage isolates within sub-lineage A1 displayed higher diversity than acute isolates suggests that the carriage isolates may have been exposed to distinctive evolutionary process to induce adaptation for survival and replication in the gall bladder. Generally, the bacterial population of *S.* Paratyphi A in Kathmandu was highly diverse and dynamic with evidence of clonal expansion followed by clonal replacement. Sub-lineage A2 replaced all other pre-existing lineages since 2011 suggesting that carrier isolates seemed to have limited role in maintaining the bacterial diversity or facilitating the transmission in this population.

Most carriage and acute Nepalese isolates within lineage C clustered in proximity to other South Asia global isolates. These isolates were mostly detected in 2005 and gradually disappeared from this population in Kathmandu after 2011, this may be associated with the expansion and replacement of other successful genotypes. The only carriage strain in lineage F was distantly related and highly distinctive from the rest of the Nepalese isolates.

4.2.5 Spatial and spatiotemporal distribution of S. Paratyphi A genotypes

4.2.5.1 Genotypic subgrouping

To investigate the short-term evolution and dynamics of bacterial transmission in this setting, I firstly reconstructed the phylogeny for Nepalese isolates using a maximum-likelihood method and further subdivided sub-lineage A1 and A2 into different monophyletic clusters (designated as A1.1-11 and A2.1) based on moderate-to-high supported bootstrap values (60-99%). Subsequently, a SNP alignment of organisms belonging to sub-lineage A1 and A2 was extracted and used to build a minimal spanning tree using the goeBURST algorithm. Subgroups identified using this method were identical to the monophyletic clusters within the maximum-likelihood phylogeny; the minimal spanning tree was displayed for better visualization (Figure 4.5). I then combined this genotype information with individual GPS location data to investigate the spatiotemporal distribution and specific demographic and clinical characteristics associated with the common genotypes.

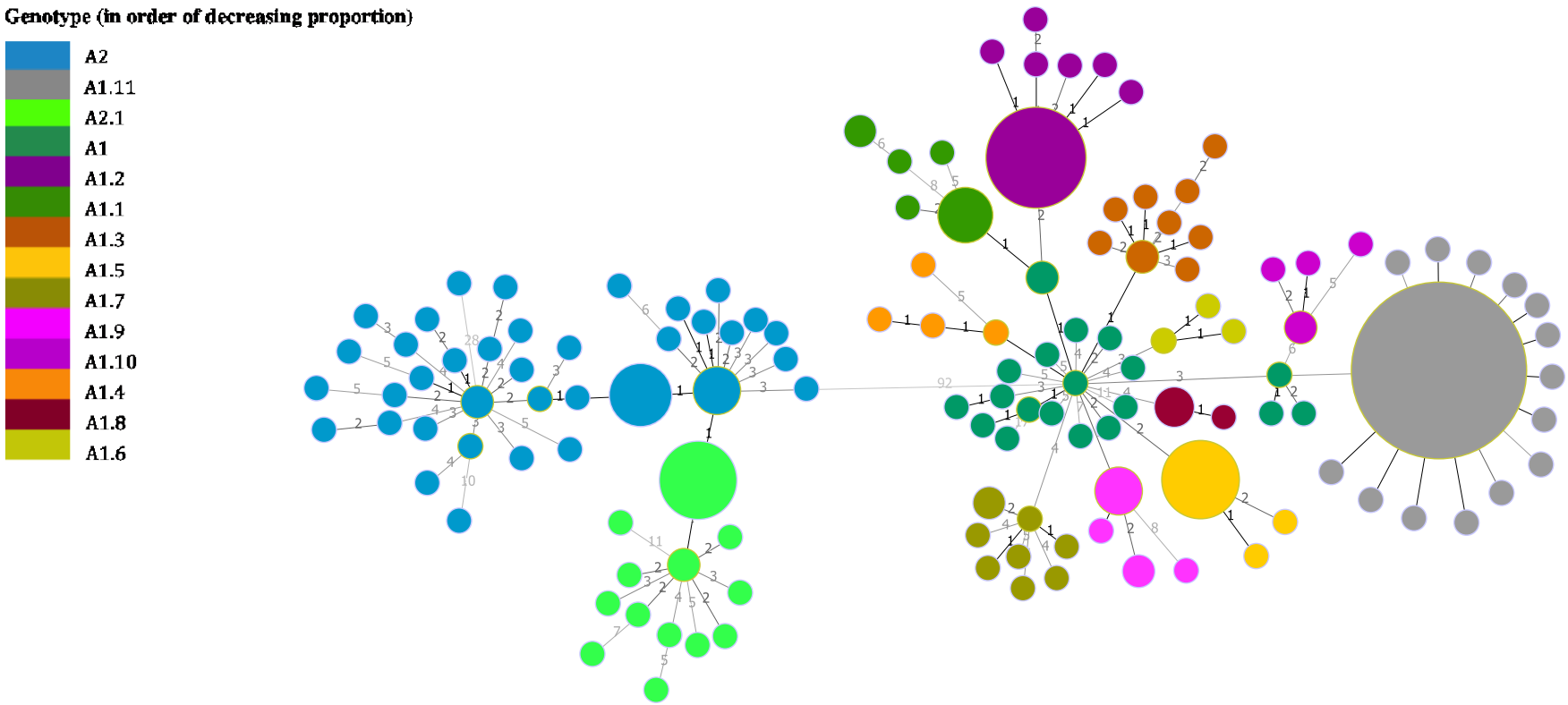

Figure 4.5 Minimal spanning tree showing different clonal clusters within sub-lineages A1 and A2

Minimum spanning tree showing various clonal clusters within sub-lineage A1 (A1.1-A1.11) and sub-lineage A2 (A2.1). The different clonal clusters are color-coded for reference. The number on each of the branches indicates the number of SNPs between each cluster.

I identified a total of 16 different genotypes of *S.* Paratyphi A. Between 2005 and 2009, sub-clade A1 was dominant and accounted for 91-96% (2005-2008) and 67% (2008-2009) of all *S.* Paratyphi A, respectively. During this period, this sub-lineage diversified and formed 11 distinct clonal clusters (A1.1-A1.11). There were substantial changes in the dominant cluster over time as well as the formation and extinction of minor clusters (Figure 4.6). For example, cluster A1.2 was the most common genotype in 2005, accounting for 27.9% of all isolates collected this year; however, this genotype subsequently disappeared and was replaced by another common cluster (A1.11) in the following years. Overall, the diversity of sub-lineage A1 decreased with time, declining from 9 different clusters in 2005 to only 3 in 2009. Between 2011-2014 there were dramatic changes in the bacterial population structure with the sudden increase of sub-lineage A2 in 2011 (62.7%, 32/51), which was concurrent with the disappearance of clonal clusters within sub-lineage A1. During this period (2011-2014), sub-lineage A2 predominated and a clonal cluster (A2.1) was represented at an almost constant rate annually (22-33%). Lineage C was constantly detected over the whole study period but represented only a small proportion of organisms.

After comparing some characteristics between sub-lineage A1 and A2, I found that the median age of patients within sub-lineage A1 was significantly older (21 years (IQR: 15-25 years) versus 18 years (IQR: 11-23 years) (p=0.0498, Kruskal-Wallis test) (Table 4.3). Clinical symptoms at presentation between these two sub-lineages were comparable except that organisms within sub-lineage A1 were more likely to be associated with abdominal pain (p=0.044, Chi-squared test). Water use also differed

dramatically, with people infected with sub-lineage A2 more likely to report drinking bottled water (21%, 14/66) compared to those infected with sub-lineage A1 (5.7%, 5/88) (p=0.0037, Chi-squared test). Additionally, a higher proportion of patients infected with sub-lineage A2 (60%, 40/66) organisms reported using municipal supply as the main water source compared to 47.7% (42/88) to those infected with sub-lineage A1 organisms (p=0.113, Chi-squared test) and were also more likely to filter water (42.4%, 28/66) compared to sub-lineage A1 (28.7%, 25/87) (p=0.078, Chi-squared test). Lastly, those infected with sub-lineage A2 organisms were significantly less likely to use well water as the main water source (3% compared to 15.9%, p=0.015, Fisher's exact test). There was no difference in the antimicrobial resistance profiles between sub-lineage A1 and A2.

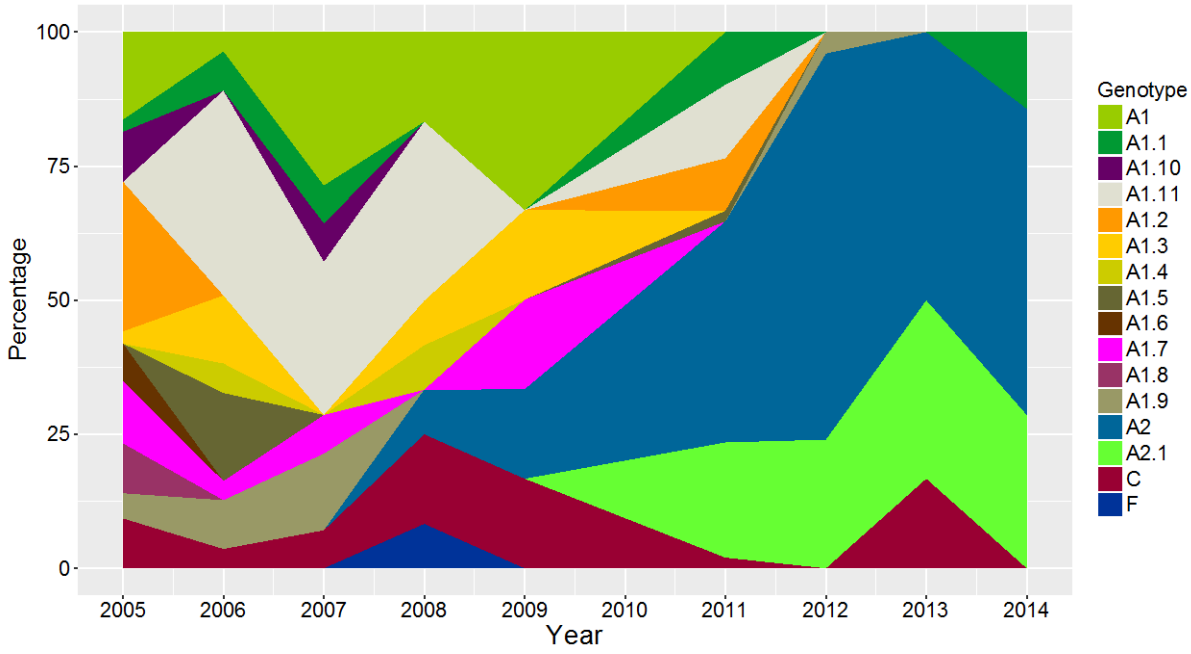

Figure 4.6 Annual distribution of *S.* Paratyphi A genotypes in Nepal

The graph shows the annual percentage of each *S.* Paratyphi A genotype in Nepal between 2005 and 2014. Different genotypes are color-coded for reference.

Table 4.3 Characteristics of two main sub-lineages A1 and A2 of *S.* Paratyphi A in Nepal

	Sub-lineage		
Characteristic	**A1**	**A2**	**p_value**
Age (median(IQR)	21 (15-25)	18 (11-23)	0.0498
Male sex	67% (89/135)	78% (52/67)	0.089
Symptom			
Anorexia	67.7 (86/127)	66.7 (44/66)	
Nausea	37.8 (48/127)	36.4 (24/66)	
Abdominal pain	37 (47/127)	22.7 (15/66)	0.044
Diarrhoea	12.6 (16/127)	12.1 (8/66)	
Constipation	13.4 (17/127)	10.6 (7/66)	
Headache	95.3 (121/127)	87.9 (58/66)	
Water source			
Municipal supply	47.7 (42/88)	60.6 (40/66)	0.113
Bottled	5.7 (5/88)	21.2 (14/66)	0.0037
Stone spout	14.8 (13/88)	10.6 (7/66)	0.446
Tanker	1.1 (1/88)	1.5 (1/66)	
Well	15.9 (14/88)	3.0 (2/66)	0.015
Mix	14.8 (13/88)	3.0 (2/66)	
Water treatment			
Boil	13.8 (12/87)	12.1 (8/66)	
Chlorine	6.9 (6/87)	0 (0/66)	
Filter	28.7 (25/87)	42.4 (28/66)	0.078
Mix	10.3 (9/87)	7.6 (5/66)	
Other	4.6 (4/87)	0 (0/66)	
Untreated	35.6 (31/87)	37.9 (25/66)	

4.2.5.2 Spatial and spatiotemporal mapping of *S*. Paratyphi A genotypes

Overall, there were a number of genotypes that exhibited spatial clustering. For the organisms within sub-lineage A1, genotype A1.2 (11/17 cases) clustered within a radius of 747m (RR = 8.6, p=0.004), genotype A1.3 (9/10 cases) also exhibited significant clustering within a radius of 696m (RR = 37.7, p<0.0001). Genotypes A1.6 (3/3 cases) and A1.10 (3/5 cases) were associated with smaller clusters containing almost all identified patients with radii <150m (RR = ∞, p<0.001 and RR = 110.0, p=0.002, respectively). For sub-lineage A2, I detected a cluster of genotype A2.1 (10/21 cases) in the south of Lalitpur with a radius of 769m (RR=12.6, p<0.001). I also investigated the remaining isolates within sub-lineage A2 and identified one cluster in the North of Lalitpur (14/46 cases) with a radius of 1.2 km (RR=3.76, p=0.003) and another smaller cluster (6/46 cases) with a radius of 157m (RR=5.2, p=0.01) nested within the cluster of genotype A1.3. The location and the magnitude of significant clusters of different genotypes over the entire period of study is shown in Figure 4.7.

I also found a number of genotypes that clustered spatiotemporally over the study period. Spatiotemporal clusters were far more common in the beginning of study period compared to later years. For example, I identified 4 distinct and significant spatiotemporal clusters of genotypes A1.2, A1.6, A1.8 and A1.10 in 2005; noticeably all clusters were concentrated between June and September. Similarly in 2006 there were two significant clusters of genotypes A1.3 and A1.11, which were also found during the same period between May and August. Sub-lineage A2 started to replace other genotypes and expand in the bacterial population from 2011 with one significant spatiotemporal cluster detected (genotype A2.1). This cluster incorporated the

majority of patients presenting between May and September in 2011. In later years, sub-lineage A2 disseminate throughout the geographical study area, but I identified a significant spatiotemporal cluster of sub-lineage A2 spanning 2011 to 2013 in the north of Lalitpur, suggesting this area was a hot spot of S. Paratyphi A infection in these years.

The combination of evidence from both spatial and spatiotemporal clustering analyses suggests that localized clusters of a variety of genotypes within sub-lineage A1 were common early on in the study period (2005-2006) and declined in favor of setting-wide dominance of sub-lineage A2 by 2011. Genotypes with overlapping spatial and spatiotemporal clusters (namely, A1.2, A1.3, A1.6, A1.0, A2.1) showed the strongest evidence of concentrated, localized outbreaks synced in time and space. Genotypes with only spatiotemporal clustering (A1.4 and A1.11) were diffuse over space, but more concentrated temporally. Sub-lineage A2 clustered spatiotemporally most obviously in 2011 when it started to replace the majority of other genotypes and subsequently dispersed over the study area.

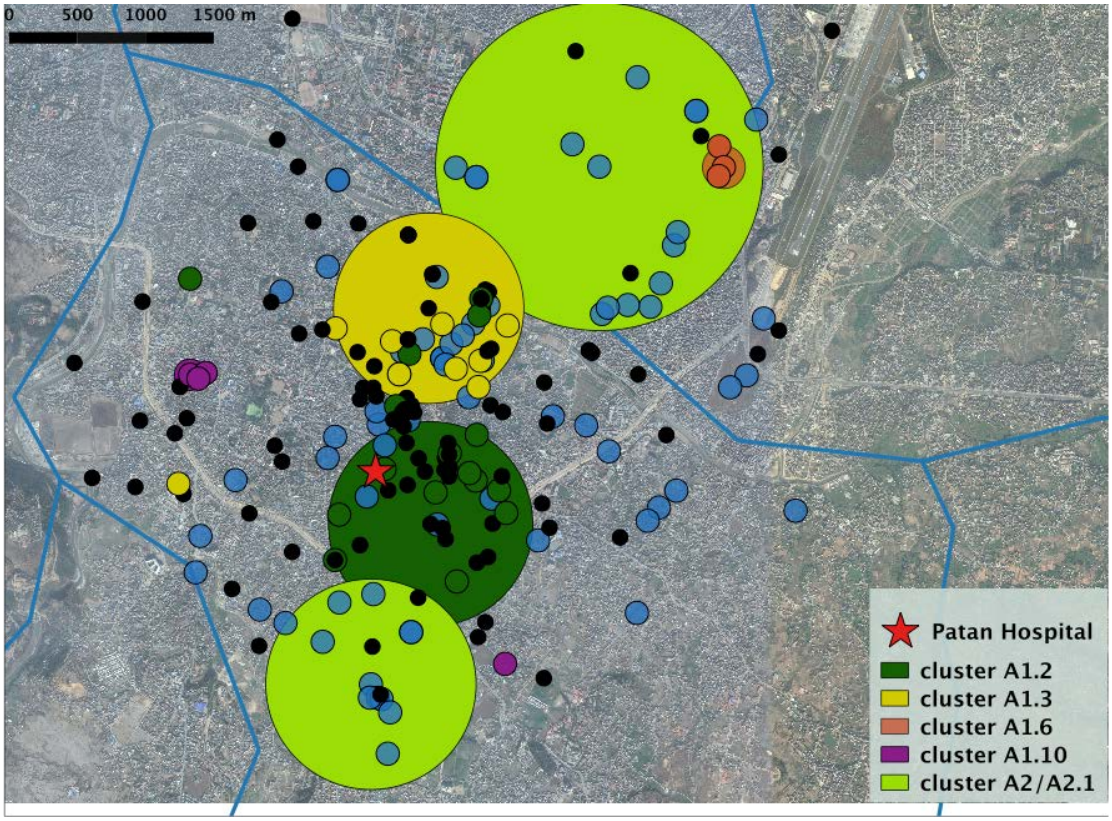

Figure 4.7 Clustering of *S.* Paratyphi A genotypes over space between 2005 and 2014

Clusters are shown in different colours, as indicated by the legend in the lower right. Significant clusters are highlighted by circles of various radii. The red star shows the location of Patan Hospital.

4.3 Discussion

S. Paratyphi A has become an increasingly important agent of typhoid fever, particularly with respect to disease prevalence and antimicrobial resistance. However, research on this pathogen is lagging behind that of *S.* Typhi and even other *Salmonella* serovars as a result of lack of public attention and a challenge of robust methodology. Previous epidemiological investigations of *S.* Paratyphi A lack a detailed characterization of the pathogen population and therefore do not provide any information about the disease dynamics. Here, I utilized a combination of bacterial genomics and more conventional epidemiology to provide new insights into the structure and dynamics of *S.* Paratyphi A in Nepal, which is essential for understanding the pattern of disease transmission in this endemic setting. I found that the *S.* Paratyphi A population in Nepal was diverse and highly dynamic and I observed evidence of a clonal expansion in early of study period followed by the replacement and expansion of a distinct clone in later years. *S.* Paratyphi A spread very rapidly throughout the geographical area during the clonal expansions.

From the global phylogeny, there were at least nine occasions where isolates from Nepal clustered with other South Asian isolates, mostly from India and Pakistan, suggesting that organism transfer between these neighbouring countries was common. Sub-lineages A1 and A2, which expanded in this location, shared common ancestors with Indian isolates. Therefore, it is unclear whether these expansions were due to successful clones locally or as a result of organism importation from neighbouring countries.

None of these Nepalese *S.* Paratyphi A in this study was resistant to any first-line drugs. Therefore, I propose that a mutation in *gyrA* and the acquisition of a novel sopE prophage were factors associated with the expansion of sub-lineage A1, which may indicate a strong selective pressure for fluoroquinolones in this setting. Previous surveillance between 1993-2003 in Nepal found that multi-drug resistant *S.* Paratyphi A were diminishing in this location population, this was occurring concurrently with decreased susceptibility against to fluoroquinolones [298]. Such phenomenon was also observed in Nepal for *S.* Typhi and is concerning given that fluoroquinolones are widely used to treat typhoid fever and can be purchased without prescription [306]. The replacement and rapid expansion of sub-lineage A2 throughout the study area from 2011 warrants further investigations as such events are rarely observed for *S.* Paratyphi A. These data imply that organisms within sub-lineage A2 have entered into a population that was either naïve or had lower immunity. *S.* Paratyphi A has been endemic in Nepal for many years and sub-lineage A1was historically the dominant lineage, therefore this genotype shift may be associated with a subtle change in antigenicity of the organisms.

Defining the population structure of *S.* Paratyphi A in Nepal allowed me to further classify these organisms into different genotypes for investigating their short-term spatiotemporal distribution, which could not be performed by conventional approaches. The spatiotemporal clustering analysis demonstrated that a substantial number of genotypes clustered by space and time, suggesting a high level of acute, person-to-person transmission of *S.* Paratyphi A in Kathmandu. Previous case-control studies in Indonesia and Nepal also found that person-to-person transmission plays an important role in *S.* Paratyphi A infection with risk factors also including flooding and

contaminated street food, whereas contaminated drinking water were more associated with *S*. Typhi infection [129,142]. Noticeably, all localized *S*. Paratyphi A outbreaks occurred during monsoon months between May and September, which corresponds with the seasonal distribution of typhoid fever in this setting [302]. Previous study also demonstrated a high level of faecal contamination of the water supply, which correlated with increased rainfall [307]. Given that only half of patients in this study access to the municipal water, the contamination of drinking water and water used in food preparation during that rainy season was likely the most important source of infection.

Outbreaks of typhoid fever caused by *S*. Paratyphi A are not uncommon. Very recently, an outbreak of *S*. Paratyphi A infection caused by an identical organisms was reported in Cambodia, which spread to several provinces and sickened dozens of European, American, Japanese travellers returning home [308]. Even though the source of this outbreak remains unknown, contaminated food is the most reasonable cause of this large-scale infection. Such outbreaks may go unobserved in Nepal, therefore improvements in food safety should be considered as an important primary control measure for *S*. Paratyphi A infection. While improvements in safe water, food safety and proper sanitation seems to be the most effective way to manage typhoid fever, it is unlikely that it would be an imminent solution for Nepal. As a result, there is an urgent need of *S*. Paratyphi A vaccine, or more ideally, a bivalent vaccine which can prevent both *S*. Typhi and *S*. Paratyphi A infections.

The relative role of typhoid carriage in the disease transmission, as well as maintaining genetic diversity and generating novel genotypes, has been extensively

questioned. By whole-genome sequencing 17 carriage isolates and relating these genomic data with acute isolates, I have, for the first time, provided new insights into the genetic relatedness between acute and carriage *S.* Paratyphi A isolates and then assessed the potential role of carriage in this endemic setting. I found a high genetic diversity within the carriage isolates from this location; the genetic diversity suggested that the majority of these isolates (59%, 10/17) were likely associated with recent asymptomatic/acute infections. However, the exact duration of carriage of these individuals is unknown as the majority did not have recent history of typhoid. The remainders of the carriage isolates were generally distantly related to the acute isolates. However, the genetic diversity within the Nepalese *S.* Paratyphi A reduced substantially after 2011 with almost all isolates falling within sub-lineage A2. Furthermore, the carriage isolates appear to have accumulated a set of differing mutations during gall bladder carriage and have become genetically distinct from contemporary acute isolates. My limited data suggests that *S.* Paratyphi A carriers did not play an important role in maintaining the genetic diversity nor in the disease transmission in this endemic setting.

In conclusion, this chapter represented a detailed molecular epidemiological investigation of *S.* Paratyphi A in Nepal. The structure and dynamics of the pathogen population was described and distinguishing by both clonal expansion and clonal replacement. I found a high rate of reduced fluoroquinolone susceptibility in *S.* Paratyphi A in this location, which likely contributed to the successful expansion of these organisms. The rapid dissemination within local population and diversification of dominant clones resulted in short-term spatiotemporal clusters of various genotypes, indicating the importance of human-to-human transmission for *S.*

Paratyphi A in this setting. My data also demonstrated that in a highly endemic setting, *S.* Paratyphi A carriage likely plays a limited contribution to the disease transmission and maintenance of the bacterial diversity.

Chapter 5

The molecular and spatial epidemiology of typhoid fever in rural Cambodia

5.1 Introduction

Typhoid fever is mainly contracted through the consumption of contaminated water or food. Therefore, the disease is largely preventable through the provision of safe water, food safety and adequate sanitation. However, such interventions are still challenging in many endemic locations given the huge economic costs and long timelines [309]. Licensed typhoid vaccines have been proved to be an effective short-term strategy in reducing typhoid burden and are currently recommended by the World Health Organization (WHO) in areas with high disease burden and increasing antimicrobial resistance [33]. Despite this recommendation, the programmatic use of typhoid vaccines is very limited, largely due to insufficient data on the epidemiology of the disease and prevalence of antimicrobial resistant organisms for evidence generation and policy making [146]. Additionally, identification of high-risk areas and vulnerable populations for targeted vaccination appears as a big challenge in resource-limited settings where usually lack of systemic disease surveillance. In such circumstances, case detection and appropriate antimicrobial therapy become crucial for typhoid management in endemic locations. However, the effectiveness of these strategies is diminished due to a lack of accurate and inexpensive rapid diagnostic tests for typhoid as well as the emergence of antimicrobial resistance organisms [109,310].

Most studies on typhoid have focused on urban slum populations, where high incidence rates have been reported [133,311,312]. However, the epidemiological features of typhoid fever in different environmental settings remain unclear, especially in rural areas with poor access to healthcare. Investigation on the transmission patterns and risk factors of typhoid fever in both urban and rural areas is critical for effective disease control and prevention [309]. Over the last decade, advances of molecular genotyping methods such as Single Nucleotide Polymorphisms (SNPs) typing and whole genome sequencing (WGS) have successfully provided unprecedented insights into the bacterial population structure and evolutionary relationships between isolates [128,244]. SNP-based typing and WGS have played a significant role in studying the molecular epidemiology of typhoid fever, discovering a diverse range of *S.* Typhi haplotypes that are commonly found co-circulating in different geographical settings [131–134,246]. These studies have outlined the importance of environment transmission within the localized human populations. Such molecular investigations are also critical for understanding the bacterial population dynamics and tracking the emergence of antimicrobial resistant organisms. The current population structure of *S.* Typhi has been primarily driven by series of clonal expansions and scattering of a specific haplotype (H58) in Asia and Africa. This globally dominant H58 haplotype mostly exhibits reduced susceptibility to fluoroquinolones and is commonly multidrug resistant against first-line agents (ampicillin, chloramphenicol, and trimethoprim-sulphamethoxazole) [248].

Typhoid fever is endemic in Cambodia but epidemiological data regarding to the mortality, morbidity and risk factors for the disease are limited. Between 2006 and 2008, a community-based study was conducted near Phnom Penh (the capital city)

and multidrug resistant *S*. Typhi with reduced susceptibility to ciprofloxacin was routinely detected from patients presenting with acute fever [313]. Additionally, a hospital-based surveillance of pediatric bloodstream infections in Siem Reap (Northwest of Cambodia) from 2007 and 2011 reported that *S*. Typhi was the most common pathogen causing pediatric bloodstream infection and H58 *S*. Typhi exhibiting reduced susceptibility to ciprofloxacin was dominant (accounting for 96% of *S*. Typhi isolates) [247]. In this investigation, I utilized whole genome sequencing to further characterize the H58 *S*. Typhi population in Siem Reap, Cambodia and combined the resulting phylogenetic information with additional epidemiological approaches to investigate the spatiotemporal distribution of *S*. Typhi and identify population-level risk factors associated with typhoid fever in this location. The data presented in this study was published in PLoS Negleted Tropical Diseases in June 2016 (Appendix G).

5.2 Results

5.2.1 Baseline characteristics

Between 2007 and 2014, there were 284 microbiologically confirmed cases of typhoid fever caused by *S.* Typhi at Angkor Hospital for Children (AHC) in Siem Reap (method section 2.4.3). *S.* Paratyphi A was uncommon, with only three cases in 2008 followed by an isolated outbreak in 2013-2014 (38 cases). A total of 262/284 (93%) of the confirmed *S*. Typhi cases lived within a 100 km radius of AHC and spanned 78 communes; these 78 communes were selected for a spatial investigation and to determine population level risk factors for typhoid fever. During this same period there were 19,877 admissions with an ICD-10 discharge diagnosis other than typhoid fever originating from the same geographic area. The baseline characteristics

of all communes and those with at least one case of typhoid fever are shown in Table 5.1.

Of the 262 cases of typhoid fever living within a 100 km radius of AHC, the median age was 8.2 years (interquartile range (IQR): 5.1-11.5 years). Additionally, 62/262 (24%) of the cases were less than five years of age and 142/262 (54%) were female. As shown in Figure 5.1a, the absolute number of confirmed cases of typhoid fever increased dramatically (from 12 cases per year to 71 cases per year) between 2009 and 2012, but then declined in 2013 and 2014 (28 and 45 cases in 2013 and 2014; respectively); data from my non-confirmed typhoid cases also reflected this trend. Over this same time period (2009 to 2014) the number of patients attending AHC for other conditions (control population) mirrored the distribution of the cases (Figure 5.1b). There was seasonal variation in the number of typhoid cases, with the majority of the cases (178/262; 68%) occurring during the early monsoon months (April, May, June and July) (Figures 5.1c & 5.1d). In late monsoon months (August to October), the number of cases declined to less than two cases per month and generally remained below this threshold in the dry season (November to March) (Figures 5.1c & 5.1d).

Table 5.1 Baseline characteristics of all communes and those with at least one case of typhoid fever

Characteristic	All communes		Typhoid communes	
	median	IQR	median	IQR
	n=243		n=78	
Population density/km^2	105.7	53-210	119.4	60-214
Elevation, m	17	12-28	18	11-35
Distance to lake, km	45.3	24-63	30.0	14-49
Average household size	4.8	4.6-5.0	5.0	4.8-5.1
Percent of population <15 yr	36.4%	34-39%	36.9%	35-39%
Median age of population, yr	19.5	18-21	19.4	18-20
Adult literacy	72.8%	59-82%	69.3%	57-77%
Female adult literacy	65.3%	50-75%	62.5%	49-71%
Total attending school	28.6%	26-31%	28.0%	24-31%
Female attending school	26.8%	24-29%	26.0%	23-28%
Female education >25 years /1,000 population				
Primary not completed	85.6	63-101	78.5	59-101
Primary/Lower secondary	27.8	16-55	22.0	16-39
Secondary or above	0.51	0.1-1.6	0.4	0.1-1.4
Toilet, % of households				
None	83.1%	63-92%	85.2%	68-92%
Sewage	5.3%	2-14%	4.6%	2-15%
Septic tank	3.9%	1-16%	2.8%	1-11%
Pit latrine	2.0%	1-5%	1.2%	1-5%
Drinking water, % of households				
Piped	1.5%	1-4%	2.1%	1-4%
Tube/pipe well	10.2%	3-28%	23.6%	8-63%
Dug well	26.9%	11-56%	23.6%	9-63%
Spring/river	23.9%	4-54%	4.7%	1-27%
Drinking water location, % of households				
Within premises	19.3%	10-35%	27.8%	18-56%
Near premises	31.1%	22-40%	28.0%	21-34%
Away premises	41.8%	23-56%	35.0%	11-52%

IQR: interquartile range.

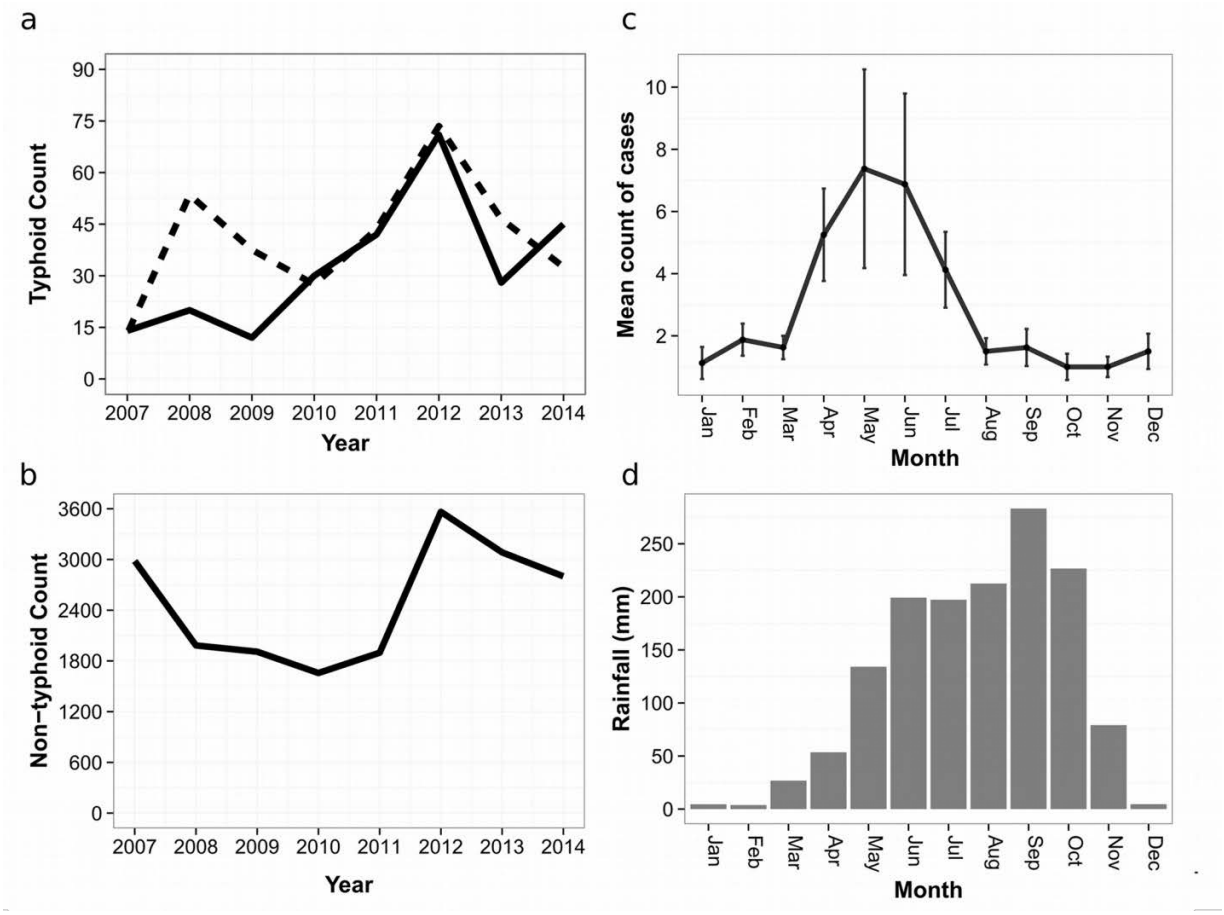

Figure 5.1 The annual and seasonal distribution of typhoid fever cases at Angkor Hospital for Children in Cambodia

a) The annual number of culture confirmed (solid line) and non-confirmed (broken line) typhoid cases at AHC from 2007 to 2014. b) The annual number of total admissions at AHC from 2007 to 2014. c) The mean monthly count of typhoid cases aggregated from 2007 to 2014. d) The average monthly rainfall (mm) per month over the study period.

5.2.2 Spatiotemporal clustering of typhoid fever cases

The majority of *S.* Typhi cases (241/284; 85%) originated from communes located within Siem Reap province (Figure 5.2). The median population density in communes with at least once case of typhoid fever was 119 people/km^2 (IQR: 60-212), and 70/78 (90%) of communes with a typhoid fever case were classified as rural. Compared to typhoid cases, the non-typhoid fever population controls came from a larger area (243 communes), the median population density of which was lower at 106 people/km^2 (IQR: 53-210); however, a similar proportion of these communes (220/243; 91%) was also classified as rural (Figure 5.2).

The estimated median commune level minimum incidence of reported cases of typhoid fever over the study period was 0.62/1,000 children aged <15 years (IQR: 0.37-1.02; range: 0.5-11.36). The reported incidence varied significantly across the 78 communes. Kampong Kleang commune (Soutr Nikom district, Siem Reap) showed the highest incidence of typhoid fever over the study period with 11.36 cases of typhoid fever /1,000 population of children aged <15 years (Figure 5.2c). This area is renowned for its floating villages and is situated on the edge of Tonle Sap Lake, approximately 35 km southeast of Siem Reap City. The second highest incidence was identified in Kaoh Chiveang commune (Aek Phnum district, Battambang, 33 km southwest of Siem Reap City) with 4.1 cases/1000 people aged <15 years over the study period (Figure 5.2c). Both of these areas experience heavy flooding when the Tonle Sap Lake expands during the rainy season.

Overall, there was some evidence of positive spatial autocorrelation (case clustering) across the 78 communes that had at least one case of typhoid fever between 2007 and

2014 (Moran's I=0.11, p<0.056). The magnitude of this autocorrelation varied over time and was the most significant in 2013 (Moran's I=0.19, p<0.019) but was non-significant in other years. I was able to identify three significant spatiotemporal clusters associated with high rates of typhoid fever. The first occurred in 2008 toward the west of the study area and had a radius of 23.8 km; this cluster had 1.27 predicted cases and 10 observed cases (relative risk [RR]=8.17, p=0.002). The second cluster occurred in 2012 in the central northern area and had a radius of 10.8 km, with 1.67 predicted cases and 12 observed cases (RR=7.47, p<0.001). The final cluster occurred in 2013 in the southeastern area and had a radius of 15.5 km, with 0.88 predicted cases and 14 observed cases (RR=16.8, p<0.0001) (Figure 5.2d).

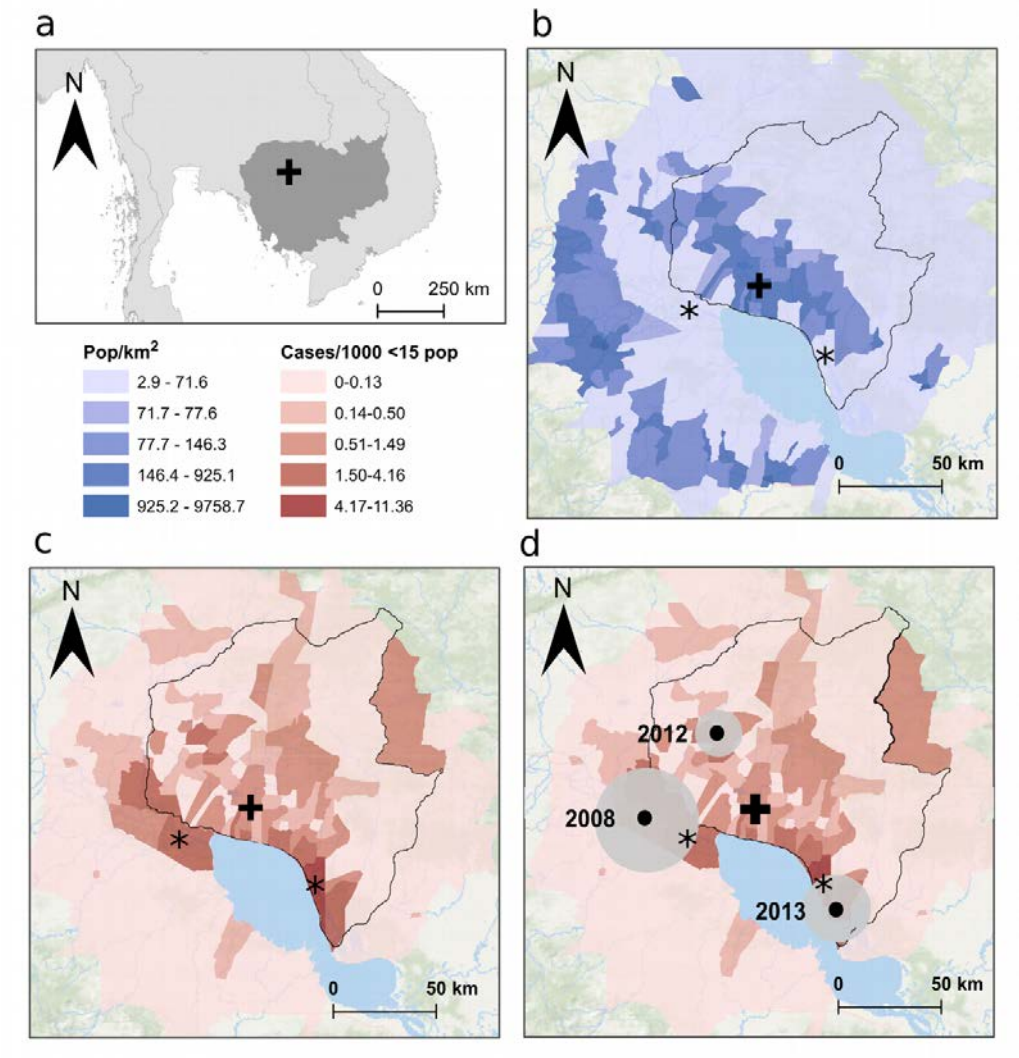

Figure 5.2 The spatial distribution of typhoid fever cases in Siem Reap province, Cambodia

a) North oriented map of Cambodia, the black cross shows the location of AHC. b) Map showing the population density (people/km^2, color-coding in key) of the 78 communes within the typhoid study area. AHC is shown by the black cross, the black border denotes Siem Reap province and the left and right asterisks are mark the locations of the communes with highest incidence of typhoid fever; Kaoh Chiveang and Kampong Kleang, respectively. c) Map of the study area showing the rate of reported typhoid cases per 1,000 population under the age of 15 years (color-coding in key). d) Map of the study area showing significant spatiotemporal clusters of typhoid during the study period, the size of the grey circles corresponds to the radius of the cluster and the years of the clusters are denoted.

5.2.3 The population structure of *Salmonella* Typhi in Siem Reap province, Cambodia

The resulting WGS data demonstrated that 97% (203/209) of the sequenced Cambodian isolates could be attributed to haplotype H58. The majority (199/203, 98%) of the H58 isolates exhibited intermediate susceptibility against fluoroquinolones (0.12-0.5 μg/mL) via the common amino acid substitution of serine to phenylalanine at codon 83 (S83F) in the DNA gyrase protein encoded by *gyrA*. There was a strong association between haplotype H58 and an IncHI1 plasmid, which confers an MDR phenotype, with 89% (180/203) of the H58 isolates harboring the common IncHI1 plasmid and the corresponding antimicrobial resistance phenotype. For the six non-H58 isolates, no mutations were observed in the *gyrA* gene, while two (33%) carried the same IncHI1 plasmid as found in the H58 isolates. I identified 188 SNPs across the H58 population and, from a SNP-based phylogeny**,** identified the circulation of at least four lineages of H58 circulating in the selected area of Cambodia between 2007 and 2012 (Fig 5.3a). These lineages, designated here as I-IV, differed from each other by as little as three to five SNPs and were phylogenetically well-supported (bootstrap values ≥ 87%). The majority of the H58 isolates fell into lineage IV (152/203, 75%) and lineage III (41/203, 20%).

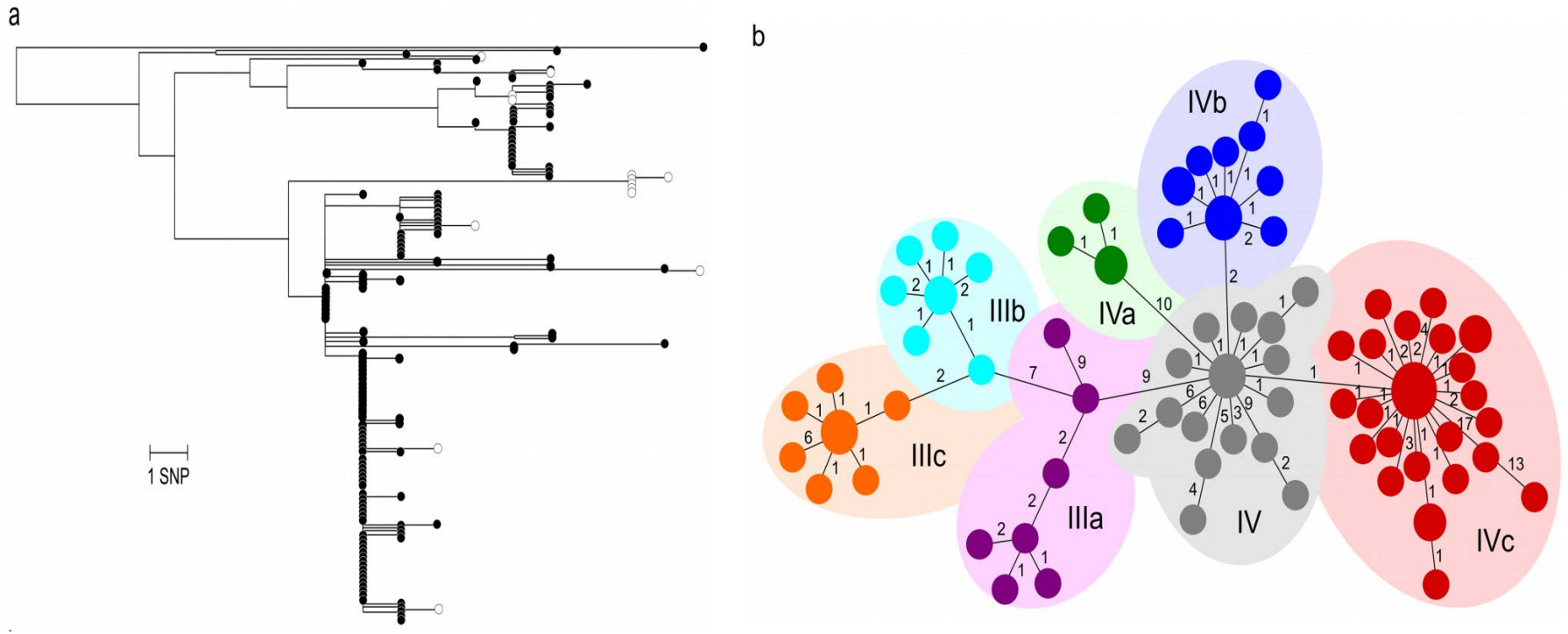

Figure 5.3 The phylogenetic structure of the H58 lineage of Cambodian *Salmonella* Typhi

a) Maximum likelihood phylogenetic tree of the 203 H58 isolates identified during this project (scale bar denotes SNP differences). The sub-lineages are shown on the major branches. Isolates exhibiting a multi-drug resistance (MDR) phenotype are indicated by black nodes. The tree is midpoint-rooted for the purpose of clarity. Bootstrap values >85% are indicated by an asterisk. b) Minimum spanning tree subdividing H58 lineage III and IV into the various sublineages (IIIa, IIIb, IIIc, IV, IVa, IVb, IVc). The various sublineages are color-coded for reference and the number of each variant is indicated by the cluster size. The number on each of the branches signifies the number of SNPs between each cluster.

5.2.4 The spatiotemporal distribution of *Salmonella* Typhi genotypes

To investigate short-term evolutionary traits within the identified lineages, I constructed a SNP-based minimum spanning tree (Figure 5.3b). Using these data, I was able to investigate the local population dynamics and detected several clonal clusters emerging from lineage III (IIIa-IIIc) and lineage IV (IVa-IVc). My data show a complex temporal distribution of *S.* Typhi H58 sublineages circulating in this location between 2007 and 2012 (Figure 5.4a). The distribution of these various strains was highly dynamic, with strain replacements, potential extinctions and the specific microevolution and expansion of H58-IVc (Figure 5.4a). In 2011 and 2012, H58-IVc became the dominant genotype, accounting for 44% (18/42) and 85% (61/72) of all *S.* Typhi isolates in these years, respectively.

I next aimed to identify spatiotemporal clustering of the various *S.* Typhi H58 sublineages, and found that IIIc, IV, IVb and IVc all displayed significant evidence of clustering over space and time. Notably, the locations of these clusters were generally different between sublineages, signifying some degree of geographical variation of the circulating *S.* Typhi strains. For example, I identified significant clustering of H58-IIIc in the western part of the study area in 2011 ($p<0.001$, RR: 26.7, radius: 36km) (Figure 5.4b) and clustering of the emergent H58-IVc strain in both 2011 (Kampong Khleang commune, $p<0.001$, RR: 39.4, radius: <1km) and in two locations in 2012 (smaller cluster, $p=0.017$, RR: 5.17, radius: 6.2km; larger cluster, $p<0.001$, RR: 5.87, radius: 33.9km).

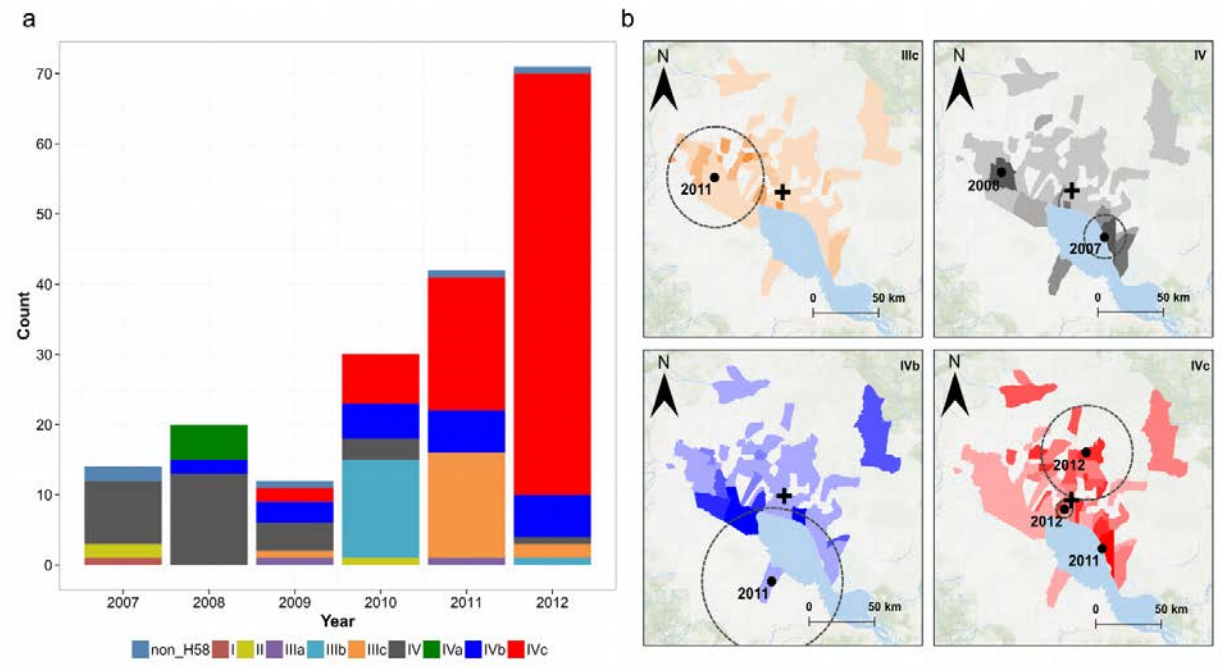

Figure 5.4 The spatiotemporal distribution of the various *Salmonella* Typhi lineages/sublineages in Siem Reap province, Cambodia

a) Bar chart shows the annual distribution of the various *S.* Typhi lineages/sublineages from 2007 to 2012; sublineages are color-coded as in Figure 3b. b) Maps showing significant spatiotemporal clusters identified for sublineages IIIc, IV, IVb and IVc. The timing of each cluster is shown by the year in black text and the dotted circle represents the radius of the detected cluster. Background colors represent the rate of each sublineage per 1,000 population aged under 15 years. The incidence rates vary between sublineages, ranging from 0 to a maximum of 0.8 (IIIc), 3.12 (IV), 2.56 (IVb) and 5.84 (IVc) 5.84 cases/1,000 population aged under 15 years.

5.2.5 Population risk factors for typhoid fever

I additionally investigated associations between rates of typhoid in children and demographic and sanitation variables at the commune level. I found a number of significant risk factors (e.g. low female education level and collection of drinking water near the household premises) and protective factors (e.g. higher population density, elevation, distance from lake and attendance at school) associated with the rate of typhoid hospitalizations in the univariate analysis (Table 5.2). However, after controlling for confounders, I found that the distance of the centroid of the commune to the perimeter of the lake was strongly and significantly associated with rate of typhoid cases (10km increase in distance from the lake, incidence rate ratio (IRR): 0.38, 95%CI 0.26-0.55, $p<0.001$) (Table 5.2). Furthermore, the relative numbers of households within the commune connected to public sewage services and households using a sunken well were also strongly protective, however these associations were reversed through interaction with increasing number of households using wells and distance from the lake, respectively (Table 5.2). Finally, a high number of households reporting drinking water retrieval from 'within the household premises' were also associated with a significant protective effect (log households/1,000 households, IRR: 0.65, 95%CI: 0.49-0.86, $p=0.003$).

Table 5.2 Regression results of highlighting factors associated with typhoid cases

Commune characteristic	Univariable		Multivariable	
	IRR (95%CI)	p	IRR (95%CI)	P
Population density^	0.81 (0.70-0.95)	0.008		
Elevation, 10m	0.89 (0.81-0.99)	0.026		
Distance to lake, 10km	0.81 (0.74-0.89)	<0.001	0.38 (0.26-0.55)	<0.001
Average household size	1.54 (0.66-3.57)	0.317		
Total attending school/1,000^	0.11 (0.04-0.33)	<0.001		
Female education >25 years /1000 population^				
Primary not completed	2.59 (1.53-4.38)	<0.001		
Primary/Lower secondary	0.94 (0.70-1.25)	0.654		
Secondary or above	0.96 (0.80-1.17)	0.714		
Toilets per 1000 people^				
None	1.24 (0.84-1.79)	0.246		
Sewage	0.96 (0.81-1.15)	0.676	0.44 (0.25-0.80)	0.007
x households with wells			1.19 (1.07-1.32)	0.001
Septic tank	0.89 (0.78-1.02)	0.090		
Pit latrine	0.95 (0.79-1.15)	0.621		
Drinking water, hh/1000 hh^				
Piped	0.87 (0.74-1.03)	0.104		
Tube/pipe well	0.82 (0.73-0.92)	0.001		
Dug well	0.83 (0.64-0.83)	<0.001	0.31 (0.19-0.50)	<0.001
x distance to lake			1.16 (1.09-1.25)	<0.001
Spring/river	1.15 (1.05-1.25)	0.003		
Drinking water location, hh/1000 hh^				
Within premises	0.71 (0.55-0.93)	0.013	0.65 (0.49-0.86)	0.003
Near premises	3.38 (2.24-5.10)	<0.001		
Away premises	0.88 (0.73-1.06)	0.177		

^log of the variable was included ; hh: household

5.3 Discussion

In this study I combined conventional epidemiological methods, current genome sequencing tools and geospatial mapping to add insight into the epidemiology of typhoid fever in atric patients attending a single healthcare facility in central Cambodia. The majority of recent typhoid fever studies originate from urban locations in low-income countries. This study provides a new perspective into this important community-acquired infection from a predominantly rural setting. The primary finding of this study is that there is a considerable and widespread burden of pediatric typhoid fever in rural Cambodia, thus questioning the dogma that typhoid fever is predominantly geographically restricted to urban populations with poor sanitation systems [133,314]. My data are consistent with findings from a recent study conducted across sub-Saharan Africa [315]. The Typhoid Surveillance in Africa Programme (TSAP) found a large burden of typhoid fever in younger children and almost equivalent population incidences between urban and rural settings. This distribution was most apparent in West Africa (Burkina Faso and Ghana) and was similarly restricted to children aged less than 15 years [315]. Therefore, I infer that the epidemiology of typhoid fever in Cambodia may be more similar to contemporary observations from sub-Saharan Africa, as opposed to the urban distribution that has commonly been observed across much of Asia [133,316].

The impending availability of Vi-conjugate vaccine raises the question of who should be given this vaccine and when it should be given to obtain maximum benefit in the control of typhoid fever [317,318]. This issue is complicated by a lack of population-based incidence data and a poor understanding of the burden of disease in school and preschool aged

children, for whom the conjugated form of the Vi polysaccharide vaccine would be particularly beneficial [319]. My data indicate a substantial burden of typhoid fever in school and preschool aged children in this area, with a hospital-based incidence (i.e. a minimum population incidence) of 11.36 cases of typhoid fever /1,000 population in children aged <15 years over the study period. The overall burden of typhoid fever in this population is likely to be greater than I have estimated due to poor sensitivity of blood culture and restriction of the study to a single healthcare centre. Siem Reap province could be a suitable location in which to trial, or even introduce, the next generation typhoid vaccines in Cambodia that have been tested elsewhere [317]. Further, I suggest that immunizing school-aged children in the period prior to the wet season may provide the most economic and prudent approach for vaccine introduction.

Between 2007 and 2012, I observed a sharp increase in the number of typhoid cases concurrent with an increasing geographic expansion. I also observed that typhoid fever in this population followed a seasonal pattern, suggesting an association with rainfall and potentially with localized flooding and the contamination of water sources. The population-based risk factors support these hypotheses, as living further away from Tonle Sap Lake and access to water within the household were highly protective. Additionally, I found that two communes located next to the lake (Kaoh Chiveang and Kampong Kleang) had the highest incidence of typhoid fever and had large clusters of cases in 2008 and 2013. This case clustering in specific locations warrants further investigation at the household level to understand specific sanitation-associated risk factors and likely exposures to *S.* Typhi in this setting [307]. It appears that access to lake water in some of

these communes, such as Kaoh Chiveang, is vital for the household water supply and I hypothesize that the lake water is more prone to localised fecal contamination at specific times throughout the year.

Using targeted SNP-specific PCR, I have previously shown that MDR H58 *S.* Typhi strains dominate in this population [247]. My WGS investigation confirmed these findings and identified additional diversification in this population. I was able to separate these H58 strains into seven (IIIa, IIIb, IIIc, IV, IVa, IVb, IVc) major sublineages. These discrete groups varied in size and were segregated by only limited numbers of SNPs. I did observe some evidence of expansion of sublineage IVc between 2009 and 2012; this correlated with several spatiotemporal clusters suggesting small disease outbreaks. I currently cannot explain the expansion of this group and my strain selection for sequencing was limited by the availability of strains isolated only up to 2012. Despite some clustering of closely related strains, the overall temporal and spatial distribution of strains was random, with a range of *S.* Typhi H58 sublineages circulating throughout the study period, which is similar to patterns described in urban settings in Asia [132,289].

This study has some limitations. First the data originated from patients attending a single healthcare facility, without the added support of healthcare utilization data. This approach, while cost-effective, induces bias in the spatial and risk factor analyses. Furthermore, while the associations identified in the regression analysis are plausible and provide direction for future investigations, they should be viewed with caution. The population level census data does not allow examination of exposures at an individual or

household-level and provides only broad epidemiological evidence. However, the association with distance to the lake and water and sanitation variables suggests these factors should be examined more rigorously in the future with respect to the dynamics of typhoid fever outbreaks. Similarly, the identification and location of the spatiotemporal clusters should be interpreted with some degree of caution. Communes without cases were not included in the cluster analyses due to a lack of data as to whether these regions truly lacked typhoid cases. A dataset with more complete spatial information on presence and absence of typhoid would permit a more reliable analysis.

In summary, I found a large burden of typhoid fever in children in rural Cambodia. My conventional population-based risk factor analysis identified access to water in the household and increasing distance from Tonle Sap Lake as protective against typhoid fever in communes. Spatial mapping and WGS provided additional resolution to investigate these findings and confirmed that proximity to that lake was associated with discrete disease clusters. I confirmed the dominance of MDR H58 *S.* Typhi in this location and found a substantial amount of diversification within this lineage. My data provide a platform for additional studies in the Cambodian population and suggest that this is a suitable location in which to introduce Vi conjugate vaccines for school children.

Chapter 6

Genetic traits of *Salmonella* Typhi gallbladder carriage isolates and their role in disease transmission in Kathmandu, Nepal

6.1 Introduction

Typhoid carriage has been recognized as an important public health problem for over a century. Mary Mallon (better known as Typhoid Mary), a cook in New York, and Mr N, a milker in England, were associated with the first true epidemiological investigations of infections and typhoid outbreaks caused by asymptomatic carriers in the early 1900s [119,320]. Typhoid carriage is typically thought to occur in patients who fail to fully clear the organisms after recovering from their acute phase of illness. These people become asymptomatic carriers and can shed the bacteria for up to three months. A subset of these individuals (~2- 5 percent) may develop into a chronic carrier state, whereby they intermittently shed the bacteria via stool and urine for more than a year [34]. Chronic carriage may provide an ecological niche that facilitates the transmission and persistence of the bacteria in human populations, potentially posing a major public health threat [321]. As a result, to prevent disease transmission and eventually eradicate the causative pathogens, it is essential to detect and provide treatment for the chronic carriers. However, this remains a huge challenge due to several impeding factors exist such as a lack of epidemiological data on the long-term follow-up of typhoid patients, intermittent fecal shedding, sup-optimal diagnostic methods with poor sensitivity to detect *Salmonella*

carriage, and a lack of scientific data on effectiveness of antimicrobial treatment for chronic carriage [322,323]. The situation is further exacerbated given the fact that infections can be symptomless and up to 25 percent of carriers have no history of acute typhoid [34,119].

Some understanding of typhoid carriage has only been gleaned in recent decades. Evidence from mouse models of *Salmonella* infection and clinical investigations have demonstrated that the gallbladder is the primary permissive niche for long-term bacterial persistence [122–126,130,324]. Various epidemiological investigations have shown an association between typhoid carriers and individuals with gallstones in their gallbladder [120,122,130,325–327]. *Salmonella* can form dense biofilms on the gallstone surface and specifically bind to cholesterol gallstones, which is considered as a hallmark of typhoid carriage [124,130,324,328,329]. Moreover, recent studies have suggested that biofilms form on the gallbladder epithelium and it has been suggested that intracellular invasion of epithelial cells lining the gallbladder may be another possible mechanism for persistence [125,330]. Despite these advances, there are still unresolved questions regarding the development of a chronic carrier state such as host responses to long-term colonization of *S.* Typhi, bacterial adaptive mechanisms for surviving within the gallbladder environment, the role of chronic carriers in disease transmission in endemic areas, regulatory mechanisms of biofilm formation, and the interaction between *S.* Typhi and the human gallbladder. Genomic characterization of individual carriage isolates as well as genomic comparison between acute and carrier isolates have been performed to investigate the genetic signatures associated with typhoid carriage [331–334]. However, these

studies provided limited information because of the limited number of isolates, making the population structure and genetic background of the acute and chronic isolates difficult to resolve. In this chapter, I performed the genome sequencing and analyses of 24 isolates from typhoid carriers and 96 isolates from acute patients collected during the time same period in Kathmandu, Nepal to describe the population structure and characterize the genetic relatedness between acute and carrier *S.* Typhi isolates. This investigation also aimed to decipher potential genetic traits associated with carrier isolates and estimate the role of carriage in the disease transmission in an area that is highly endemic for typhoid fever.

6.2 Results

6.2.1 The phylogenetic structure of Nepalese acute and carrier *Salmonella* Typhi isolates between 2007 and 2010

My genomic and phylogenetic analyses of 120 *S.* Typhi isolates (24 from typhoid carriers and 96 from acute patients) demonstrated that genotype 4.3.1 (H58) was the most dominant genotype isolated during this study period, constituting 65.6% (63/96) of all acute isolates and 62.5% (15/24) of all carriage isolates. The second most common genotype was 3.3.0 (H1), accounting for 14.6% (14/96) and 12.5% (3/24) of all acute and carrier isolates, respectively. Additionally, there was high genetic diversity within this bacterial population with various minor genotypes co-circulating, including the subclades 3.2.2; 3.0.1; 2.2.2; 2.2.1; clades 4.1; 3.1; 2.2 and primary cluster 2 (Figure 6.1). Of these minor genotypes, the carriage isolates fell within subclade 3.2.2 (8.3%; 2/24), 2.2.2 (4.2%; 1/24), clade 2.2 (8.3%; 2/24) and primary cluster 2 (4.2%; 1/24). Overall, the

carriage isolates were not significantly more likely to be H58 in comparison to non-H58 organisms (p=0.774, Chi-squared test). These data suggest that *S*. Typhi carriage isolates were not restricted to any particular bacterial genotype; instead, the genotype distribution among carrier isolates mirrored the general genetic structure of the sampled bacterial population.

I hypothesized that carriage isolates may be more resistant to antimicrobials, therefore I firstly selected to investigate the presence of fluoroquinolone-resistance mutations. The genome sequence data showed that mutations conferring non-susceptiblity to fluoroquinolones were predominately found within the H58 organisms (64/78, 82.1%), rather than the non-H58 organisms (1/42, 2.4%). The majority of the mutations occurred at codon 83 of the DNA gyrase gene *gyrA*, changing serine to phenylalanine (S83F) in the protein. Comparing *gyrA* mutations between the acute and carriage isolates, respectively, within the H58 organisms, I found that 76.2% (48/63) and 60% (9/15) contained the S83F mutation; 7.9% (5/63) and 13.3% (2/15) had a S83Y mutation and 15.9% (10/63) and 26.7% (4/15) had no mutation. There was no significant difference (p=0.327, Chi-squared test) in the presence of fluoroquinolone resistance-associated mutations between acute and carrier isolates in this group.

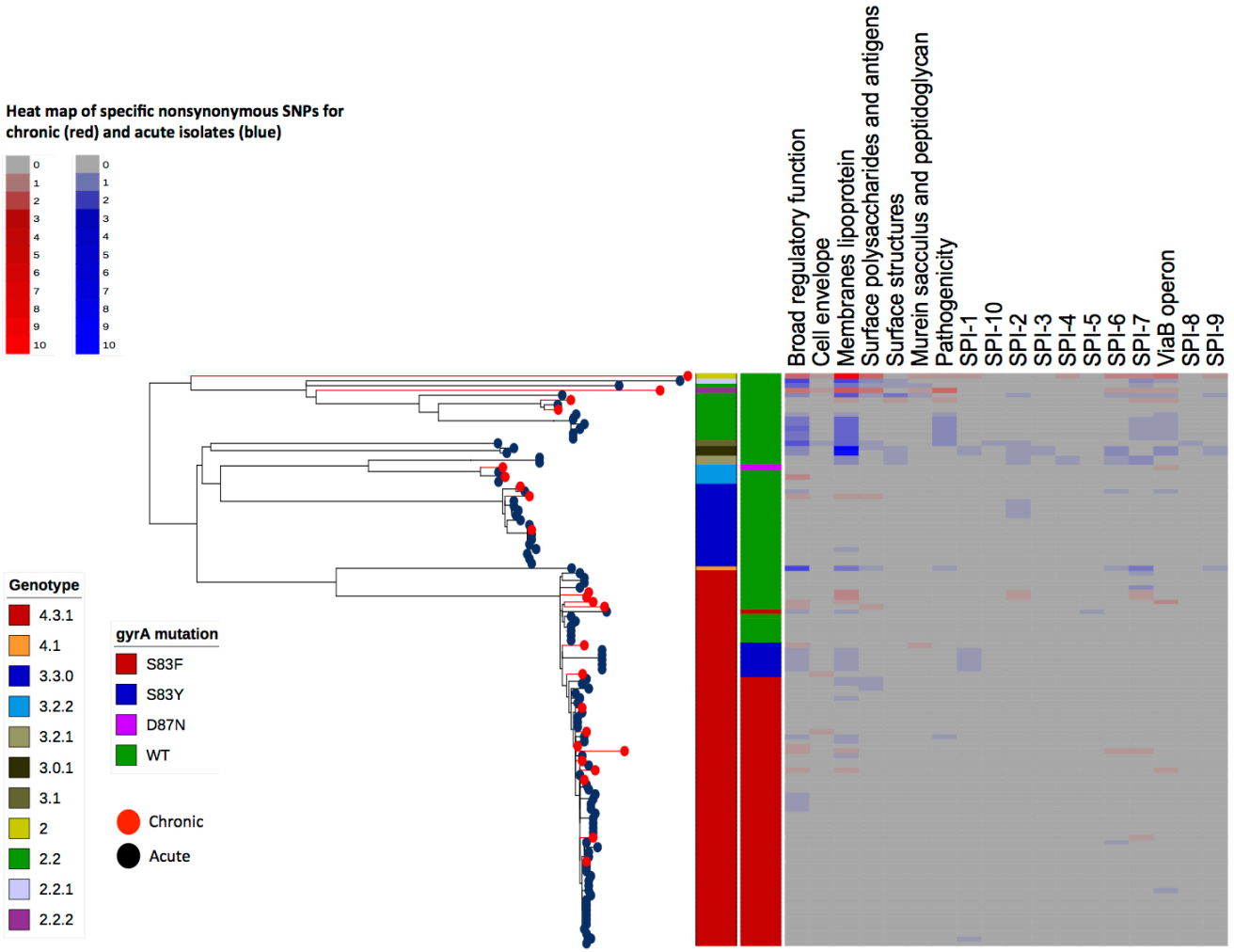

Figure 6.1 The phylogenetic structure of carriage and acute *S.* Typhi isolates collected between 2007 and 2010

Mid-point rooted maximum likelihood tree based on core-genome SNPs of 120 *Salmonella* Typhi isolates including 24 from typhoid carriers and 96 from acute patients with the corresponding metadata: genotype, gyrA mutations and a heat map of functional distribution of nonsynonymous SNPs that were specific to chronic and acute isolates. The heat map is based on the number of nonsynonymous SNPs specific to chronic (represented as red bars) and acute isolates (represented as blue bars) in each functional class, with an emphasis on functions related to bacterial regulation, surface structures, pathogenicity and virulence. Chronic and acute isolates are shown as red and dark circles at the terminal nodes, respectively. Terminal branches leading to chronic isolates are highlighted in red.

6.2.2 Dissecting the genetic traits of *Salmonella* Typhi carriage isolates

There was a total of 2186 SNPs (in comparison with reference genome CT18) identified across this collection of 120 genomes from acute and carriage *S.* Typhi carriage isolates. 1038/2186 (47.5%) of these SNPs were nonsynonymous (44.7% (977/2186) missense and 2.8% (61/2186) were nonsense); 32.4% (708/2186) were synonymous and 20% (440/2186) were in the intergenic regions and pseudogenes. The mean dN/dS for acute isolates was slightly less than the mean dN/dS for carrier isolates (0.478 vs 0.495, respectively), although both of these figures suggest that most of the mutations were neutral. In order to better understand the adaptive point mutations associated with typhoid carriage, all nonsynonymous SNPs (NSs) occurring exclusively within carrier *S.* Typhi genomes were identified. Genes containing these SNPs were further grouped by their predicted functions according to the *S.* Typhi functional classification scheme generated by the Sanger Institute. A comparable analysis was performed for all NSs in the acute *S.* Typhi isolates. There were a total of 228 carriage-specific NSs (212 missense and 16 nonsense mutations, appendix E) and 469 acute-specific NSs (437 missense and 32 nonsense mutations, appendix F). There was no significant difference (p=0.924, Chi-square test) in the proportion of nonsense mutations out of total specific NSs in the acute versus the carriage isolates across the whole phylogenetic structure. However, for genotype 4.3.1 (H58), the proportion of nonsense mutations out of total specific NSs for carrier isolates was significantly higher than that of acute isolates (10/60 compared to 2/67, Fisher exact test, p=0.009). These data suggested that gene degradation by nonsense mutations was more common in carriage isolates in comparison to the acute isolates within genotype 4.3.1. The inactivated genes among the carriage isolates included genes

involved in the synthesis of peptidoglycan (*pbpC*), vitamin B12 receptor (*btuB*), general stress response regulator (*rpoS*), laterally acquired protein in SPI-7 (STY4562), membrane transport protein (STY3932), central metabolism (STY0230, *ggt*), hypothetical proteins (STY0929, STY4178) and osmotically inducible lipoprotein E precursor (*osmE*).

Overall, the NSs associated with acute versus carrier *S.* Typhi isolates could be grouped into 78 functions. The highest prevalence of NSs was found in gene functions related to hypothetical proteins, membranes lipoproteins, unknown function, transport/binding proteins, SPI-7, general regulatory function, surface polysaccharides and antigens, carbohydrate degradation, and DNA replication/modification (Figure 6.2). There was no statistically significant difference found in the prevalence of these NSs across all gene functions between acute and carrier isolates. However, the data showed that the proportion of NSs in the *viaB* operon was significantly higher in carriage isolates in comparison to the acute isolates (9/228 compared to 7/469, Chi squared test, p=0.04). This finding was also reflected for *S.* Typhi isolates belonging to genotype 4.3.1, with carriage isolates having more specific NSs in the *viaB* operon than the acute isolates (5/60 compared to 1/67, Fisher's exact test, p=0.08). Additionally, there were two carriage isolates (GB428 and GB003) that had lost the Vi capsular polysaccharide due to the deletion of the entire SPI-7 region (134kb). Their loss of Vi expression was confirmed by agglutination test using Vi antisera under different osmolarity conditions (method section 2.3). The *viaB* operon is a ~14 kb region located within SPI-7 in the *S.* Typhi chromosome containing five genes responsible for Vi polysaccharide synthesis

(*tviABCDE*) and a further five genes (*vexABCDE*) involved in transportation of the Vi capsule to the cell surface. The Vi capsular polysaccharide is an important virulence factor enhancing bacterial resistance to complement and phagocytic killing [335]. Anti-Vi antibody titers have been found to be abnormally high in the blood of typhoid carriers and have been used as a serological marker for the detection of typhoid carriage [336,337]. Here, I show that multiple *S.* Typhi carriage isolates had lost the ability to express Vi.

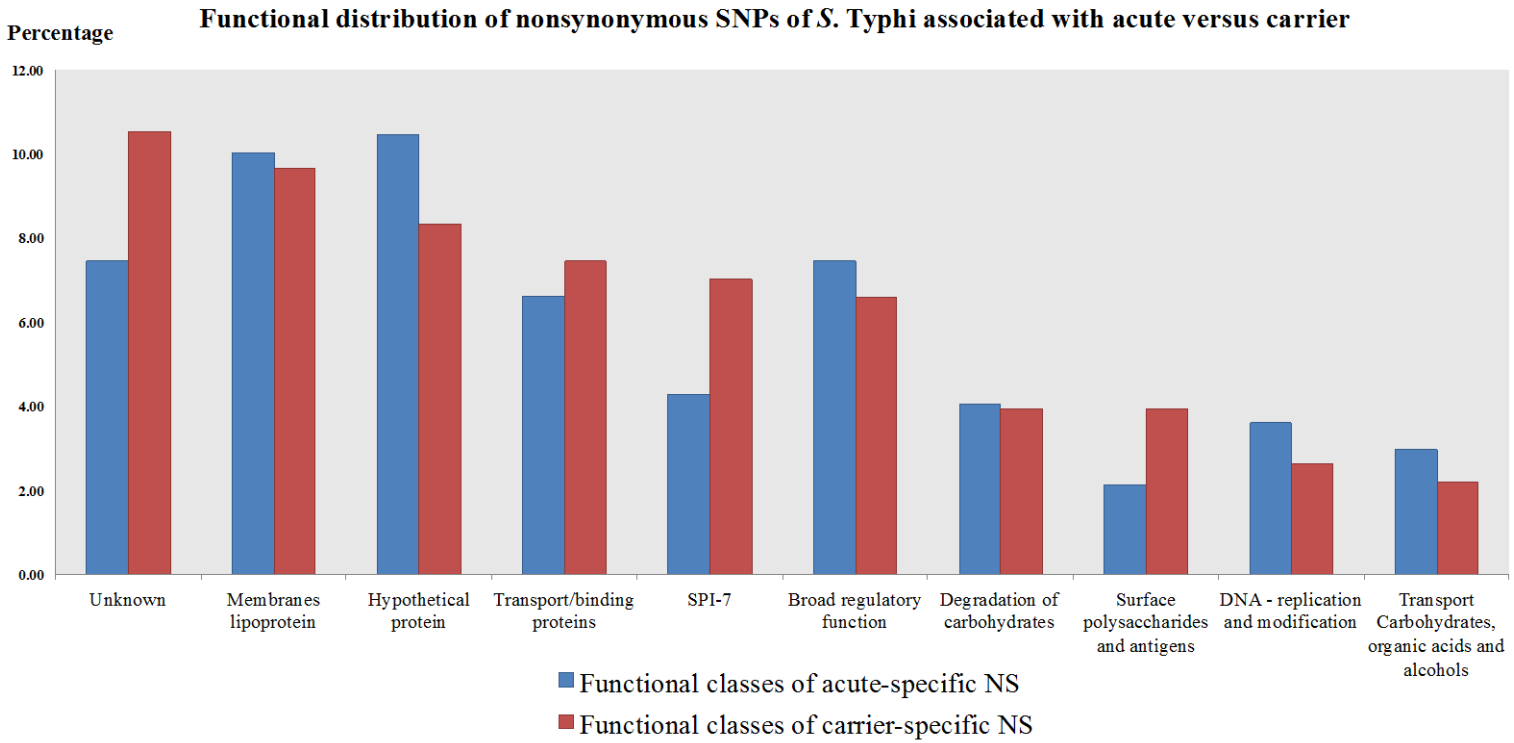

Figure 6.2 Top ten functional classes with highest prevalence of acute-specific nonsynonymous SNPs versus carrier-specific nonsynonymous SNPs

6.2.3 Positive selection associated with typhoid carriage

Next, I investigated signatures of positive selection by trying to identify similar genetic changes within different carrier isolates. Among the carrier-specific NSs identified, there were a number of different mutations occurring in the same gene or the same biological pathways in at least two phylogenetically unlinked carrier isolates. For example, within the *viaB* operon as mentioned above, there were two NSs at codon 137 and 462 in the *tviE* gene (isolates GB580 and GB026) and six NSs in codons 166, 504, 506, 508, 665, 752 in the *tviD* gene (isolates GB005, GB026, GB076, GB125 and GB281). Both genes are known to be involved in the polymerization and translocation of the Vi capsule [338]. Convergent NSs also occurred the *rpoS* gene of isolates GB125 (nonsense mutation at codon 247) and GB705 (NSs at codon 94 and 250). The *rpoS* gene encodes the sigma factor sigma-38, a central regulatory protein of the general stress responses (temperature, pH, osmolarity, redox state, antimicrobial peptide) and nutrient starvation. RpoS has been showed to down-regulate Vi expression and RpoS-negative strains of *S.* Typhi overexpressed Vi polysaccharide at low and medium osmolarities, which can affect their invasiveness and macrophage resistance [339,340]. A further example was NSs at codon 59 and 230 in the *degS* gene (serine protease) (isolates GB005 and GB169). *DegS* is a component of DegS-DegU two-component system, which is involved in expression of several degradative enzymes for salt stress responses and growth-limiting conditions in Gram-positive bacteria, suggesting it might have an important role for bacterial survival in the high salt concentration within the gall bladder. Additionally, three isolates (GB005, GB026, and GB705) had different NSs in codons 335, 406, 946, respectively, in STY1242 (*ptsG* - glucose-specific PTS system IIBC component). The PtsG enzyme is a

component of the glucose-specific phosphotransferase system, plays a role in phosphorylation and translocation of glucose across the bacterial membrane, and is induced in carbon-limited conditions [341]. None of the acute isolates had mutations in this gene. There were also several other genes containing NSs in more than two carriage isolates, such as STY0429 (*SbcC* - exonuclease), STY0661 (*dmsC* - molybdopterin containing oxidoreductase membrane anchor subunit), STY1447 (putative ribulose-5-phosphate 3-epimerase) and STY2760 (*ratA* - putative exported protein).

With respect to convergent mutations within the same biological pathways, there were a number of carriage-specific NSs involved in LPS O-antigen synthesis and modification. For example, a NS in the *rfc* gene (regulator of O-antigen polymerization) in isolate GB441; a NS in the STY2629 gene (LPS modification acyltransferase) of isolate GB335; two NSs in the *rfbE* (CDP-tyvelose-2-epimerase) and *rfaG* genes (lipopolysaccharide core biosynthesis protein) in isolate GB281 and three NSs in the *rfbK* (phosphomannomutase), *manB* (phosphomannomutase) and *rfaD* genes (ADP-L-Glycero-D-mannoheptose-6-epimease) in isolate GB026. *RfbK* and *manB* are both related to GDP-mannose synthesis for the LPS and *rfaD* is an enzyme that catalyzes the conversion of ADP-D-glycerol-D-mannoheptose to ADP-L-glycerol-D-mannoheptose, a precursor for the synthesis of inner-core LPS. The enrichment of NSs related to LPS structure suggests that LPS has an important role for the long-term colonization of these carrier isolates.

6.2.4 Estimating the role of typhoid carriage in disease transmission in Kathmandu, Nepal

Although the majority of acute and carriage isolates generally clustered together within the same genotypes across the phylogenetic tree of Nepalese *S.* Typhi, many carriage isolates were associated with atypically long terminal branches, suggesting they exhibited a greater degree of genetic variation from the acute isolates (Figure 6.1). One rational explanation for this phenomenon was that many of the *S.* Typhi isolates from typhoid carriers have been colonizing the gallbladder for a prolonged period and have undergone distinct mutational accumulation for adaptation, whereby they have gradually become distantly related to contemporary acute isolates. Additionally, there were very few cases where contemporary acute isolates were directly linked to carriage isolates within the phylogenetic structure. Taken together, these results potentially indicate that these typhoid carriers had limited contribution to the transmission of disease in Kathmandu.

Alternatively, there may be other explanations for the presence of long terminal branches in the phylogenetic tree of carriage isolates. For example, carriers may have been colonized by imported variants with a different genetic background, or more simply, the *S.* Typhi isolates from acute patients may have been under-sampled or not representative of the diversity circulating in the environment. I therefore aimed to identify the most probable molecular mechanisms contributing to the variation in branch lengths. The number of *S.* Typhi isolates was not evenly distributed among different genotypes, therefore, for this sub-analysis I selected the acute and carrier isolates belonging to only genotype 4.3.1 (H58), which is also the most successful and globally dominant genotype.

First, I reconstructed a global phylogeny of 4.3.1 *S.* Typhi isolates including 78 isolates from this study (63 from acute patients and 15 from typhoid carriers) and a global collection of 798 published previously isolates [248]. The phylogenetic reconstruction showed that the majority of Nepalese 4.3.1 isolates belonged to a dominant group comprising of 9/15 carrier isolates and 45/63 acute isolates (Figure 6.3). For the remaining isolates, two carrier isolates (GB003 and GB044) clustered together with Indian isolates; the other four carrier isolates (GB076, GB441, GB266, and GB387) were distantly related with respect to the rest of the 4.3.1 isolates and consistently associated with long terminal branch lengths; the acute isolates were found to cluster with Indian (n=6) and Southeast Asian isolates (n=6) and other minor groups within the Nepalese 4.3.1 population. This finding suggested that there has been a clonal expansion of 4.3.1 *S.* Typhi in Kathmandu, with some evidence of inter-country typhoid transmission. However, it was difficult to determine the direction of transmission without accompanying epidemiological information. Furthermore, there were little data to support that long terminal branches associated with carrier isolates were driven by colonization of imported strains.

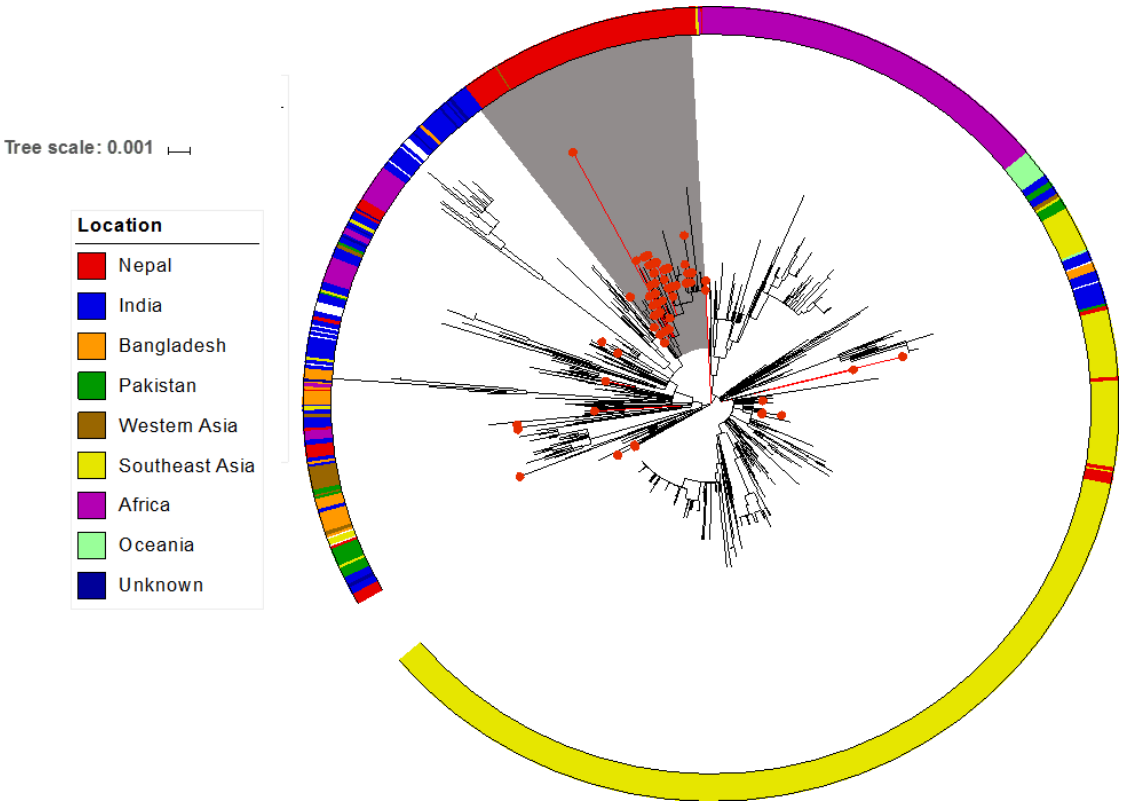

Figure 6.3 Phylogenetic structure of acute and carrier 4.3.1 *Salmonella* Typhi isolates from Nepal in the global context

Maximum likelihood phylogenetic tree of Nepalese H58 *S.* Typhi isolates (63 from acute patients and 15 from typhoid carriers) in the global context. H58 *S.* Typhi isolates from this study are highlighted in red circles at the terminal nodes. Terminal branches associated with carriage isolates are shown in red color. The outer ring exhibits the location of the isolates from Nepal and its neighboring countries as well as other regions in the world.

I next extracted and compared the branch lengths from the most recent common ancestor (tMRCA) for acute versus carriage isolates belonging to genotype 4.3.1. My data showed that the mean branch length from the tMRCA for the acute isolates was 0.00447 substitutions/site (range: 0.00196-0.0083) equivalent to 12 SNPs (range: 5-21 SNPs). The mean branch length from the tMRCA for the carriage isolates was 0.0052 substitutions/site (range: 0.00276-0.01108) equivalent to 14 SNPs (range: 7-28 SNPs). Therefore, although some carriage isolates within genotype 4.3.1 were associated with longer branch lengths in comparison to acute isolates, this difference was not significant. One explanation for this may be a variable duration of *S.* Typhi carriage.

Lastly, I estimated and compared the pairwise genetic distances within the acute isolates versus the pairwise genetic distances within carriage isolates. The rationale for this was that the number of acute isolates (n=63) was higher than chronic isolates (n=15), therefore the genetic diversity within acute isolates should be higher if there was no selection and the mutation rates were comparable. However, the median pairwise SNP distance within the acute isolates was 13 SNPs (IQR: 8-19 SNPs), which was significantly lower than the median pairwise SNP distance within the carriage isolates: 21 SNPs (IQR: 12-24) (Wilcoxon rank sum test, $p=2.8\text{x}10^{-9}$) (Figure 6.4). Additionally, the median pairwise SNP difference between the acute and the carriage isolates was 17 SNPs, suggesting some degree of overlap in SNP content between these two groups. These findings demonstrate that there was more genetic diversity among carriage isolates than acute isolates, which may result from an accumulation of adaptive point mutations in

response to alternative selective pressures within the gallbladder.

These data suggest that carriage isolates likely follow different evolutionary pathways from acute isolates. During carriage *S.* Typhi consistently accumulates potentially adaptive point mutations, which may assist long-term survival within human gallbladder. My data proposes that the longer carrier isolates reside within the gallbladder, the more genetically distinct they are from acute isolates. Furthermore, it suggests that typhoid carriage may not be an important source of disease transmission in Kathmandu.

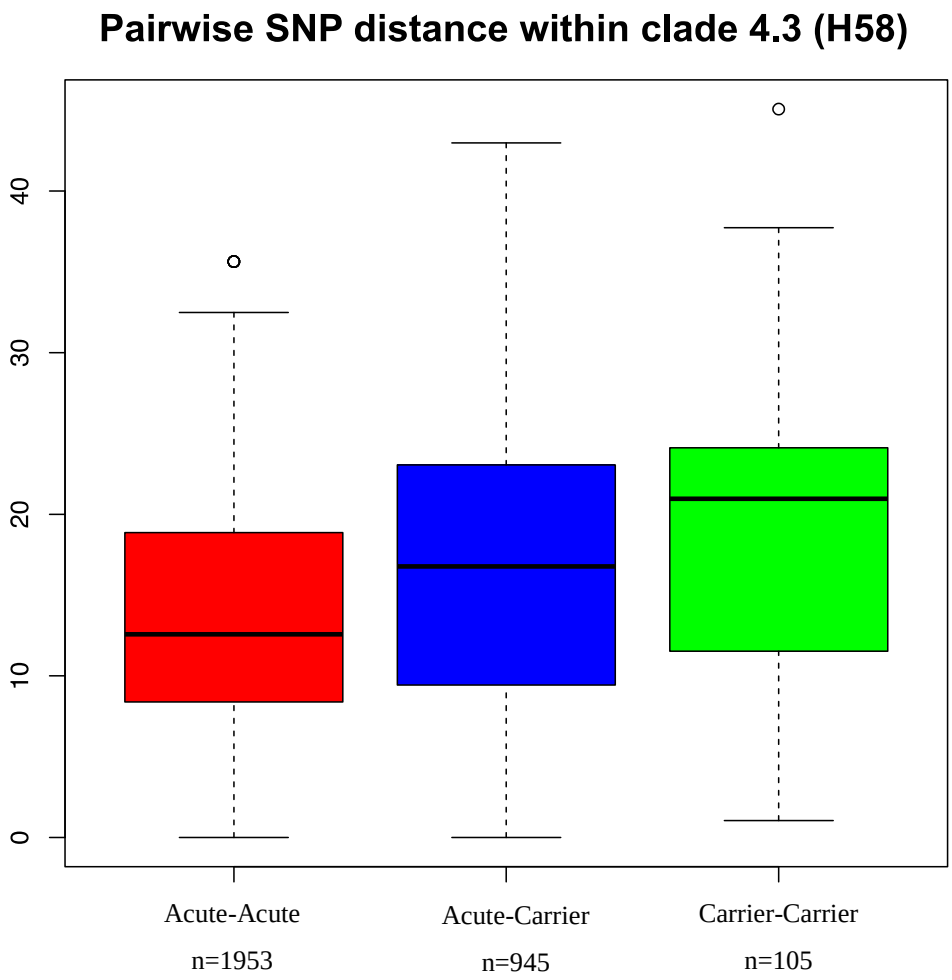

Figure 6.4 Distribution of pairwise SNP distances within and between acute and carrier isolates

6.3 Discussion

Understanding the molecular mechanisms associated with chronic typhoid carriage represents a challenging research area given that this population is difficult to identify prospectively. Additionally, there has been as a lack of robust genetic methodology to investigate these questions due to the genetic conservation of *S.* Typhi. The environmental factors driving the evolution of *S.* Typhi within the gallbladder are poorly understood and little is known about the adaptive mechanisms that promote long-term survival. Previous genomic comparisons between *S.* Typhi isolates recovered from acute patients and typhoid carriers have failed to take into consideration the heterogeneity of the bacterial population, which may have a large effect on determining genetic variation between acute and carrier isolates. An ideal approach for genomic comparison should include a characterization of the bacterial population structure in the context of typhoid endemicity. In this study, my data demonstrated that typhoid carriage was induced by a diverse range of bacterial genotypes. Further, the development of a carrier state was neither restricted to any particular genotype nor to a fluoroquinolone-resistant phenotype. My study thus rejects the notions that typhoid carriage may be associated with one particular genotype or that a shift towards fluoroquinolone therapy has reduced carriage. In fact, the genetic distribution of carriage isolates largely reflected the overall genetic structure in the bacterial population, with the most dominant genotype being 4.3.1. Genotype 4.3.1 (H58) started to emerge and disseminate across South Asia in the early 1990s [176]. Until now, 4.3.1 has been replacing other older preexisting genotypes to become the dominant genotype circulating in South Asia, Southeast Asia and Africa. According to my data, it is likely that the continued spread of 4.3.1 *S.* Typhi will

dominate over other minor genotypes in Kathmandu. This may also increase the detection of 4.3.1 *S*. Typhi in typhoid carriers over time, compromising the potential role that typhoid carriers play in maintaining bacterial diversity.

Assessing the role of chronic carriage in disease transmission represents one of the most important public health issues in typhoid control. Typhoid carriers have been widely considered as an important source of infection; however, their exact contribution to newly infected cases in endemic areas like Kathmandu remains questionable. Previous molecular epidemiological studies in endemic regions in Nepal, Vietnam, and Indonesia have shown the abundance of environmental transmission in these areas, with a wide diversity of co-circulating bacterial genotypes identified among acute typhoid patients [131–134]. These studies have also implied a minimal contribution of person-to-person transmission to new typhoid cases in endemic settings. Here, a phylogenetic reconstruction for non-recombinant non-repetitive core genomes demonstrated a high level of genetic diversity of *S*. Typhi genotypes circulating in Kathmandu. More importantly, the phylogenetic structure of the acute and chronic isolates collected during the same period in a single location provided a unique opportunity to investigate the phylogenetic relationship between these isolates and assess the role of carriage in disease transmission. There were very few examples in the phylogenetic tree where the carriage isolates clustered in close proximity or directly gave rise to acute isolates. Further, many carriage isolates were associated with atypically long terminal branches. While there might be several factors that may lead to this phenomenon, my investigations suggested that the accumulation of adaptive point mutations in the carriage isolates was the most

likely molecular mechanism. Considering acute and carrier isolates are circulating in different ecological niches, long-term exposure to different selective pressures exerted by these environments may result in a difference in their accumulated adaptive mutations over time. This also means that the longer the duration of carriage in the gallbladder, the more genetically distantly related they are in comparison to contemporary acute isolates. This speculation is supported by the fact that genotype 4.3.1 carrier isolates exhibited a significantly higher level of genetic diversity in comparison to the acute isolates. Taken together, my analysis strongly advocates that *S.* Typhi have undergone a distinctive evolutionary pathway during the carrier state and play a limited role in disease transmission in Kathmandu.

I additionally aimed to investigate the genetic signatures associated with typhoid carriage with a focus on genetic function, particularly at the single nucleotide level. As typhoid carriage is not restricted to any bacterial genotype, gene acquisition is unlikely to contribute to the development of a carrier state. By identifying NS mutations occurring specifically in carrier isolates and classifying them into predicted functional classes for comparisons with those of acute isolates, I found that gene degradation by nonsense mutations was significantly higher in carriage than acute isolates within genotype 4.3.1. The phenotypic effect of gene inactivation on the bacterial phenotype and carriage is unknown. However, it is an intriguing phenomenon and worthy of further investigation, as gene inactivation has been shown to be an important evolutionary mechanism in the adaptation of *S.* Typhi [231,240]. There was also evidence regarding the enrichment of NS mutations related to the Vi polysaccharide capsule in the carrier isolates. Vi antigen is

immunogenic and anti-Vi antibody gradually wanes in acute typhoid patients after recovery, but is persistent in the blood of chronic carriers [337,342]. Despite the fact that anti-Vi antibody is not a reliable serological marker for the detection of typhoid carriers in endemic areas, data from sero-surveillance studies for chronic carriage in these areas have commonly reported a high prevalence of elevated levels of anti-Vi antibodies in healthy individuals, which may be associated with both carriers and repeatedly infected persons [343,344]. Immunofluorescent staining of biofilms produced by *S.* Typhi on the surface of human gallstones has shown the abundance of Vi capsule in the biofilm extracellular matrix, suggesting that *S.* Typhi consistently express Vi antigen during the carrier state [329]. The increased frequency of nonsynonymous mutations in the viaB operon (*tviB, tviD* and *tviE* gene) of carrier isolates in this study together with their high level of anti-Vi antibodies in the blood (unpublished data) suggest that *S.* Typhi residing in the gallbladder are under consistent selection pressure imposed by the human immune response.

Identifying genes that may be under positive selection among carriage isolates is crucial for understanding the evolutionary forces and bacterial adaptation to the gallbladder environment during the carrier state. Signatures of positive selection were detected in a number of genes containing differing carriage-specific NS mutations in at least two phylogenetically unlinked carriage isolates. Many of these genes were associated with gene regulation under stress conditions and the expression of virulence genes. For example, the global regulatory gene *rpoS* is not only responsible for general stress responses and nutrient starvation but also regulates genes involved in biofilm formation,

the colonization of Peyer's patches, persistence of *S.* Typhimurium in the spleen, and virulence and Vi polysaccharide synthesis of *S.* Typhi [345–347]. Furthermore, the *degS* gene is involved in salt stress responses and growth-limiting conditions in Gram-positive bacteria; STY1242 (*ptsG* - glucose-specific PTS system IIBC component) is activated under the stress of carbon starvation. These findings suggest that *S.* Typhi is exposed to a range of differing stresses within the human gallbladder. Furthermore, the genes responsible for LPS biosynthesis and modification also displayed a marked accumulation of NS mutations in the carriage isolates. LPS is a major component of the outer membrane of all Gram-negative bacteria and represents one of the main factors contributing to the resistance for high concentrations of bile salt in the gallbladder [348,349]. LPS is also a key structural component of the biofilm extracellular matrix which forms on human gallstones [329]. The disruption of genes involved in LPS biosynthesis of *S.* Typhimurium may have a negative effect on the production of biofilms and the attachment of bacteria on contact surfaces [350]. The enrichment of NS mutations in genes involved in LPS biosynthesis and modification can lead to structural changes of the LPS and thereby might enhance bile resistance or affect biofilm formation. There were also several other genes potentially under positive selection which could have unknown effects on carriage, including STY0429 (*SbcC* - exonuclease), STY0661 (*dmsC* - molybdopterin containing oxidoreductase membrane anchor subunit), STY1447 (putative ribulose-5-phosphate 3-epimerase), and STY2760 (*ratA* - putative exported protein).

This study has some limitations. First, the number of carrier and acute isolates was relatively small and thus might not truly reflect the actual genetic structure in the

bacterial population. This sample size may also affect the interpretation of the phylogenetic distances between acute and carrier isolates and the functional analyses of the specific NS mutations in these isolates. Second, the duration of carriage is impossible to assess as most typhoid carriers do not recall a history of typhoid. My data suggested that the duration of carriage was likely to be highly variable, which consequently led to variable terminal branch lengths. It was not possible to measure the evolutionary rate of *S.* Typhi during the carriage state without knowing the duration of carriage in the gallbladder. Therefore, it is necessary for all future epidemiological investigations of typhoid fever to include a follow-up period of at least one year. This is essential not only to provide better estimate of the evolutionary rate of *S.* Typhi in different ecological niches but also to understand bacterial adaptation during carriage. Despite these obvious limitations, this was the first ever genomic investigation of the genetic characteristics of *S.* Typhi carriage isolates and the phylogenetic relatedness between carrier and acute isolates circulating over the same time period in a typhoid-endemic area. This study also provided valuable evidence for assessing the role of typhoid carriers in disease transmission in Kathmandu, Nepal.

In conclusion, typhoid carriage is not associated with any particular genotype nor driven by fluoroquinolone resistance. Additionally, I found strong evidence that typhoid carriers are likely not an important source of new infections in endemic settings such as Kathmandu. As a result, public health control measures should focus on providing people with safe water, food safety and vaccination for disease prevention. *S.* Typhi is exposed

to a variety of stressful conditions within the gallbladder and undergoes distinctive evolutionary processes for better adaptation.

Chapter 7

General discussion

The emergence and spread of a novel subclade of ciprofloxacin-resistant H58 *S.* Typhi in Nepal and a neighboring country demonstrates that fluoroquinolone resistance in this pathogen has become a serious problem across South Asia. More importantly, such organisms were found to be associated with fluoroquinolone treatment failure. My data strongly advocate that fluoroquinolones should no longer be used as empirical therapy for typhoid in this region; therefore, more clinical studies should be conducted to evaluate the effectiveness of alternative treatment options. Additionally, with the first typhoid conjugate vaccine Typbar-TCV that can be used for children aged less than two years and has been prequalified by WHO, it is necessary to conduct vaccine trials to measure the efficacy and immunogenicity of this new vaccine in Nepal and to develop a sustainable vaccine procurement and financing mechanisms. Programmatic use of the new typhoid vaccine targeting high-risk populations may not only reduce the disease burden but also minimize the antimicrobial use and selective pressure in the bacterial population. Meanwhile, molecular typing and phenotypic characterization of *S.* Typhi needs to be performed routinely in South Asian countries and other reference laboratories outside this region to keep track of the spread of this novel H58 subclade to provide early warning for public health authorities and healthcare providers. These potentially dangerous organisms are likely to be widespread across the Indian subcontinent and may become a global health threat if they reach other vulnerable populations in Asia and Africa, as we have seen previously with the other H58 *S.* Typhi.

Despite the progress in *S.* Typhi vaccine development, there is still no vaccine against *S.* Paratyphi A. Further, the epidemiological characteristics and burden of disease caused by *S.* Paratyphi A are not well described, and the dynamics of bacterial populations in endemic areas are often overlooked. Using a genomic approach combined with conventional epidemiological tools, my study provided a detailed characterization of the phylogenetic structure and spatiotemporal distribution of *S.* Paratyphi A isolates from Kathmandu. This Nepalese *S.* Paratyphi A population was highly dynamic with evidence of regular bacterial transmission between Nepal and neighboring countries, resulting in clonal expansions of distinct genotypes at different time periods. Lineage A was the most common genotype found in this collection of *S.* Paratyphi A, which consisted of two dominant sub-lineages (A1 and A2), whereby a single sub-lineage (A2) has rapidly replaced all other preexisting genotypes from 2011. The emergence and rapid lineage replacement of sublineage A2 in Nepal is an intriguing phenomenon and requires further research into its virulence and antigenicity. For example, in my laboratory we are working on the characterization of LPS structure, epithelial cellular invasion and macrophage killing of *S.* Paratyphi A isolates belonging to different genotypes. Further, human protective immunity against differing *S.* Paratyphi A genotypes as well as the clinical and epidemiological features associated with infections caused by sub-lineage A2 need to be further investigated. The introduction of the new typhoid conjugate vaccine against *S.* Typhi in areas where *S.* Typhi and *S.* Paratyphi A are co-circulating may additionally lead to an increase in *S.* Paratyphi A infections. While *S.* Paratyphi A infections have become a growing problem in many parts of Asia, vaccines against this

pathogen are lagging far behind. Therefore, routine surveillance and improved public health practices in endemic areas can be an immediate solution for disease control and management.

Most previous typhoid studies have primarily focused on urban slums in low and middle-income countries where high rates of typhoid fever have often been reported. There is a general lack of epidemiological data about the disease burden and antimicrobial resistance of *S.* Typhi in rural settings with limited access to healthcare. My study provided unprecedented insights into the *S.* Typhi population structure and epidemiological features of typhoid fever in rural areas in Siem Reap province, Cambodia. Typhoid fever is widespread in rural areas of Cambodia, causing a significant disease burden in children aged less than 15 years. Several spatiotemporal outbreaks were identified in communes located near Tonle Sap Lake and proximity to the lake was associated with increased risk of infection. There was also a wide geographic distribution and high prevalence of MDR H58 *S.* Typhi with reduced susceptibility to fluoroquinolone in this setting. Despite these novel findings, the study was based on hospital surveillance data, which was limited to a single healthcare setting and lacked complete geospatial data. Therefore, community surveillance at individual or household-level in communes with high typhoid fever incidence (such as Kaoh Chiveang and Kampong Kleang) is necessary to identify specific risk behaviors and spatiotemporal case clusters. Furthermore, the level of fecal contamination and the presence of *S.* Typhi in water resources in these areas should also be examined. Such valuable information would be essential for guiding public health interventions as well as identifying high-risk

populations for typhoid vaccine introduction, given the fact that the new typhoid conjugate vaccine has been prequalified and registered in Cambodia.

The scope of my study also aimed to address one of the most important typhoid research questions regarding to role of typhoid carriage in disease transmission in endemic settings. My genomic and phylogenetic analyses of *S.* Typhi and *S.* Paratyphi A isolates recovered from asymptomatic carriers and acute patients suggested that typhoid carriers are not likely to contribute significantly to new infections in endemic area or the maintenance of bacterial genetic diversity. *S.* Typhi is likely to be under stress within the gallbladder and forge a distinct evolutionary pathway for better adaptation within the gallbladder. One of the important limitations of my study was a lack of information on the duration of carriage, which is crucial to estimate the mutation rate of *S.* Typhi during the carrier state and to better understand the phylogenetic relationships between carriage and acute isolates. However, it is very difficult to identify chronic typhoid carriers prospectively considering the very low rate of chronic carriage, intermittent fecal shedding, and a lack of a robust detection method. To follow up on the findings described in my thesis, studies are currently being performed in my laboratory to further characterize and compare the capacity to invade epithelial cells, kill macrophages and form biofilms between the acute and carrier isolates belonging to the same lineages. Such experiments are necessary to better understand how *S.* Typhi behave within the gallbladder environment during the carrier state. Additional research is also being conducted to identify specific biological markers associated with chronic carriage that can be used to develop a diagnostic test.

In conclusion, the advance of next-generation sequencing has revolutionized epidemiological research and bacterial genomics has become an essential tool for understanding the circulation of pathogens and the impact of antimicrobial resistant organisms on disease outcomes. Using genomic approaches combined with other clinical and epidemiological data, my research contributed significantly to a better understanding of the emergence of antimicrobial resistance and epidemiological features of typhoid fever in both rural and urban settings. For the first time, I measured an effect of AMR and bacterial genotype on the treatment outcome of typhoid fever, which is likely to have a substantial impact on clinical and public health practices.

In conclusion, my thesis strongly advocates the discontinued use of fluoroquinolones as empirical treatment for typhoid fever in South Asia and highlights the necessity for alternative antimicrobial therapies, as well as the introduction of new conjugate typhoid vaccines in this region. Further molecular investigations on disease transmission and AMR surveillance are required to routinely monitor the emergence and spread of these clinically important pathogens. The research methods described in my thesis have proven to be powerful, adaptable, robust and should be applied for similar studies on alternative pathogens in the future. Only by understanding the interaction of antimicrobial treatment and the infecting organism can we begin to curtail the current antimicrobial resistance crisis.

Chapter 8

References

1 AWWA (American Water Works Association). Waterborne Pathogens - Manual of Water Supply Practices, 2nd edn. American Water Works Association, 2006.

2 Ochman H, Wilson AC. Evolution in bacteria: Evidence for a universal substitution rate in cellular genomes. *J Mol Evol* 1987; **26**: 74–86.

3 Sofos J. Advances in Microbial Food Safety. Elsevier Science, 2013.

4 Greenwood D, Slack RCB, Peutherer JF. Medical microbiology: a guide to microbial infections : pathogenesis, immunity, laboratory diagnosis, and control. Churchill Livingstone, 2002.

5 Popoff MY, Le Minor L. Antigenic formulas of the Salmonella serovars 8th ed. World Health Organization Collaborating Center for Reference and Research on Salmonella. *Res Microbiol* 2001; **152**: 907–9.

6 Braden CR. Annual Summary 2002. *Reading* 2002.

7 Andrewes FW. Studies in group-agglutination I. The salmonella group and its antigenic structure. *J Pathol Bacteriol* 1922; **25**: 505–21.

8 Popoff MY, Bockemühl J, Gheesling LL. Supplement 2002 (no. 46) to the Kauffmann-White scheme. *Res Microbiol* 2004; **155**: 568–70.

9 Grimont P, Weill F-X. Antigenic formulae of the Salmonella servovars. *WHO Collab Cent Ref Res Salmonella* 2008; : 1–167.

10 Centre for Disease Control and Prevention. National Enteric Disease Surveillance : STEC Surveillance Overview. *Cent Dis Control* 2011; : 0–3.

11 Kauffmann F. The Bacteriology of Enterobacteriaceae: Collected Studies of the Author and His Coworkers. Williams & Wilkins, 1966.

12 Brenner FW, Villar RG, Angulo FJ, Tauxe R, Swaminathan B. GUEST COMMENTARY Salmonella Nomenclature. *J Clin Microbiol* 2000; **38**: 2465–7.

13 Crosa ' JH, Brenner DJ, Ewing WH, Falkow S. Molecular Relationships Among the Salmonelleae.

J Bacteriol 1973; **115**: 307–15.

14 Le Minor L, Popoff MY. Designation of Salmonella enterica sp. nov., nom. rev., as the Type and Only Species of the Genus Salmonella: Request for an Opinion. *Int J Syst Bacteriol* 1987; **37**: 465–8.

15 Reeves MW, Evins GM, Heiba a. a., Plikaytis BD, Farmer JJ. Clonal nature of Salmonella typhi and its genetic relatedness to other salmonellae as shown by multilocus enzyme electrophoresis and proposal of Salmonella bongori comb. nov. *J Clin Microbiol* 1989; **27**: 313–20.

16 Grimont PAD, Weill F-X. Antigenic formulae of the Salmonella serovars, 2007, 9th Edition. *WHO Collab Cent Ref Res Salmonella ParisPasteur* https://www.pasteur.fr/sites/default/files/veng_0.pdf (accessed Oct 18, 2017).

17 Farmer JJ, McWhorter AC, Morris GK, Brenner DJ. The Salmonella-Arizona group of Enterobacteriaceae: nomenclature, classification, and reporting. *Clin Microbiol Newsl* 1984; **6**: 63–6.

18 The Community Summary Report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in the European Union in 2008. *EFSA J* 2010; **8**: 1496.

19 Fierer J, Guiney DG. Diverse virulence traits underlying different clinical outcomes of Salmonella infection. *J Clin Invest* 2001; **107**: 775–80.

20 Agbaje M, Begum RH, Oyekunle MA, Ojo OE, Adenubi OT. Evolution of Salmonella nomenclature: a critical note. DOI:10.1007/s12223-011-0075-4.

21 Kingsley RA, Bäumler AJ. Host adaptation and the emergence of infectious disease: The Salmonella paradigm. *Mol Microbiol* 2000; **36**: 1006–14.

22 Umler AJB, Tsolis EM, Ficht TA, Adams ALG. MINIREVIEW Evolution of Host Adaptation in Salmonella enterica. *Infect Immun* 1998; **66**: 4579–87.

23 Dunstan SJ, Hue NT, Han B, *et al.* Variation at HLA-DRB1 is associated with resistance to enteric fever. *Nat Genet* 2014; **46**: 1333–6.

24 Butler T, Islam A, Kabir I, Jones PK. Patterns of morbidity and mortality in typhoid fever dependent on age and gender: Review of 552 hopitalized patients with diarrhea. *Rev Infect Dis* 1990; **13**: 85–90.

25 Parry CM, Thompson C, Vinh H, *et al.* Risk factors for the development of severe typhoid fever in Vietnam. *BMC Infect Dis* 2014; **14**: 73.

26 Waddington CS, Darton TC, Jones C, *et al.* An outpatient, ambulant-design, controlled human infection model using escalating doses of salmonella typhi challenge delivered in sodium bicarbonate solution. *Clin Infect Dis* 2014; **58**: 1230–40.

27 Gotuzzo E, Frisancho O, Sanchez J, *et al.* Association Between the Acquired Immunodeficiency Syndrome and Infection With Salmonella typhi or Salmonella paratyphi in an Endemic Typhoid Area. *Arch Intern Med* 1991; **151**: 381–2.

28 Lin FYC, Ho VA, Bay P V., *et al.* The epidemiology of typhoid fever in the Dong Thap Province, Mekong Delta region of Vietnam. *Am J Trop Med Hyg* 2000; **62**: 644–8.

29 Bhutta ZA. Impact of age and drug resistance on mortality in typhoid fever. *Arch Dis Child* 1996; **75**: 214–7.

30 Stuart BM, Pullen RL. Typhoid: Clinical Analysis of Three Hundred and Sixty Cases. *Arch Intern Med* 1946; **78**: 629–61.

31 Sinha a, Sazawal S, Kumar R, *et al.* Typhoid fever in children aged less than 5 years. *Lancet* 1999; **354**: 734–7.

32 Ferreccio C, Levine MM, Manterola A. Benign bacteremia caused by Salmonella typhi and paratyphi in children younger than 2 years. *J Pediatr* 1984; **104**: 899–901.

33 World Health Organization. Background document: The diagnosis, treatment and prevention of typhoid fever. *World Health* 2003; : 38.

34 Parry CM, Hien TT, Dougan G, *et al.* Typhoid fever. *N Engl J Med* 2002; **347**: 1770–82.

35 Van Den Bergh ETAM, Hussein Gasem M, Keuter M, Dolmans M V. Outcome in three groups of patients with typhoid fever in Indonesia between 1948 and 1990. *Trop Med Int Heal* 1999; **4**: 211–5.

36 STUART BM, PULLEN RL. Typhoid; clinical analysis of 360 cases. *Arch Intern Med (Chic)* 1946; **78**: 629–61.

37 Hoffman SL, Punjabi NH, Kumala S, *et al.* Reduction of mortality in chloramphenicol-treated severe typhoid fever by high-dose dexamethasone. *N Engl J Med* 1984; **310**: 82–8.

38 Rogerson SJ, Spooner VJ, Smith TA, Richens J. Hydrocortisone in chloramphenicol-treated severe typhoid fever in papua new guinea. *Trans R Soc Trop Med Hyg* 1991; **85**: 113–6.

39 Hoa NTT, Diep TS, Wain J, *et al.* Community-acquired septicaemia in southern Viet Nam: The importance of multidrug-resistant Salmonella typhi. *Trans R Soc Trop Med Hyg* 1998; **92**: 503–8.

40 Bhutta ZA. Impact of age and drug resistance on mortality in typhoid fever. *Arch Dis Child* 1996; **75**: 214–7.

41 Butler T, Islam A, Kabir I, Jones PK. Patterns of morbidity and mortality in typhoid fever dependent on age and gender: review of 552 hospitalized patients with diarrhea. *Rev Infect Dis*; **13**: 85–90.

42 Butler T, Rumans L, Arnold K. Response of typhoid fever caused by chloramphenicol-susceptible and chloramphenicol-resistant strains of salmonella typhi to treatment with trimethoprim-sulfamethoxazole. *Rev Infect Dis* 1982; **4**: 551–61.

43 Gotuzzo E, Morris JG, Benavente L, *et al.* Association between Specific Plasmids and Relapse in Typhoid Fever. *J Clin Microbiol* 1987; : 1779–81.

44 Yew FS, Chew SK, Goh KT, Monteiro EH, Lim YS. Typhoid fever in Singapore: a review of 370 cases. *J Trop Med Hyg* 1991; **94**: 352–7.

45 Acharya G, Butler T, Ho M, *et al.* Treatment of typhoid fever: Randomized trial of a three-day course of ceftriaxone versus a fourteen-day course of chloramphenicol. *Am J Trop Med Hyg* 1995; **52**: 162–5.

46 Hien TT, Bethell DB, Hoa NTT, *et al.* Short course of ofloxacin for treatment of multidrug-resistant typhoid. *Clin Infect Dis* 1995; **20**: 917–23.

47 Wain J, Hien TT, Connerton P, *et al.* Molecular typing of multiple-antibiotic-resistant Salmonella enterica serovar Typhi from Vietnam: Application to acute and relapse cases of typhoid fever. *J Clin Microbiol* 1999; **37**: 2466–72.

48 Parry CM, Ho VA, Phuong LT, *et al.* Randomized controlled comparison of ofloxacin, azithromycin, and an ofloxacin-azithromycin combination for treatment of multidrug-resistant and nalidixic acid-resistant typhoid fever. *Antimicrob Agents Chemother* 2007; **51**: 819–25.

49 Levine MM, Tacket CO, Sztein MB. Host-Salmonella interaction: Human trials. *Microbes Infect*

2001; **3**: 1271–9.

50 Hornick RB, Greisman SE, Woodward TE, DuPont HL, Hawkins AT, Snyder MJ. Typhoid Fever: Pathogenesis and Immunologic Control. *N Engl J Med* 1970; **283**: 739–46.

51 Woodward WE. Volunteer studies of typhoid fever and vaccines. *Trans R Soc Trop Med Hyg* 1980; **74**: 553–6.

52 Glynn JR, Hornick RB, Levine MM, Bradley DJ. Infecting dose and severity of typhoid: Analysis of volunteer data and examination of the influence of the definition of illness used. *Epidemiol Infect* 1995; **115**: 23–30.

53 Jones B, Pascopella L, Falkow S. Entry of microbes into the host: using M cells to break the mucosal barrier. *Curr Opin Immunol* 1995; **7**: 474–8.

54 House D, Bishop A, Parry C, Dougan G, Wain J. Typhoid fever: pathogenesis and disease. *Curr Opin Infect Dis* 2001; **14**: 573–8.

55 Dougan G, Baker S. Salmonella enterica serovar Typhi and the pathogenesis of typhoid fever. *Annu Rev Microbiol* 2014; : 317–36.

56 Schwan WR, Huang XZ, Hu L, Kopecko DJ. Differential bacterial survival, replication, and apoptosis-inducing ability of Salmonella serovars within human and murine macrophages. *Infect Immun* 2000; **68**: 1005–13.

57 Nga TVT, Karkey A, Dongol S, *et al.* The sensitivity of real-time PCR amplification targeting invasive Salmonella serovars in biological specimens. *BMC Infect Dis* 2010; **10**: 125.

58 Wain J, Bay PVB, Vinh H, *et al.* Quantitation of bacteria in bone marrow from patients with typhoid fever: Relationship between counts and clinical features. *J Clin Microbiol* 2001; **39**: 1571–6.

59 Wain J, Diep TS, Ho VA, *et al.* Quantitation of bacteria in blood of typhoid fever patients and relationship between counts and clinical features, transmissibility, and antibiotic resistance. *J Clin Microbiol* 1998; **36**: 1683–7.

60 Hoffman SL, Edman DC, Punjabi NH, *et al.* Bone marrow aspirate culture superior to streptokinase clot culture and 8 ML 1:10 blood-to-broth ratio blood culture for diagnosis of typhoid fever. *Am J Trop Med Hyg* 1986; **35**: 836–9.

61 Vallenas C, Hernandez H, Kay B, Black R, Gotuzzo E. Efficacy of bone marrow, blood, stool and duodenal contents cultures for bacteriologic confirmation of typhoid fever in children. *Pediatr Infect Dis J* 1985; **4**: 496–8.

62 Gilman R, Terminel M, Levine M, Hernandez-Mendoza P, Hornick R. RELATIVE EFFICACY OF BLOOD, URINE, RECTAL SWAB, BONE-MARROW, AND ROSE-SPOT CULTURES FOR RECOVERY OF SALMONELLA TYPHI IN TYPHOID FEVER. *Lancet* 1975; **305**: 1211–3.

63 Lanata C, Ristori C, Jimenez L, *et al.* Vi SEROLOGY IN DETECTION OF CHRONIC SALMONELLA TYPHI CARRIERS IN AN ENDEMIC AREA. *Lancet* 1983; **322**: 441–3.

64 Ferreccio C, Levine M, Astroza L, *et al.* The detection of chronic Salmonella typhi carriers: a practical method applied to food handlers. *Rev Med Chil* 1990; **118**: 33–7.

65 Kiran Y, Yadav SK, Geeta P. A comparative study of typhidot and widal test for rapid diagnosis of typhoid fever. *Int J Curr Microbiol Appl Sci* 2015; **4**: 34–8.

66 Narayanappa D, Sripathi R, Jagdishkumar K, Rajani HS. Comparative study of dot enzyme immunoassay (Typhidot-M) and Widal test in the diagnosis of typhoid fever. *Indian Pediatr* 2010; **47**: 331–3.

67 Khoharo HK. A comparative study of the typhidot (Dot-EIA) and Widal tests in blood culture positive cases of typhoid fever. *Trop Doct* 2011; **41**: 136–8.

68 Sherwal B, Dhamija R, Randhawa V, Jais M, Kaintura A, Kumar M. A Comparative Study of Typhidot and Widal Test in Patients of Typhoid Fever. *JIACM* 2004; **5**: 244–6.

69 Naheed A, Ram PK, Brooks WA, *et al.* Clinical value of Tubex™ and Typhidot® rapid diagnostic tests for typhoid fever in an urban community clinic in Bangladesh. *Diagn Microbiol Infect Dis* 2008; **61**: 381–6.

70 Dutta S, Sur D, Manna B, *et al.* Evaluation of new-generation serologic tests for the diagnosis of typhoid fever: data from a community-based surveillance in Calcutta, India. *Diagn Microbiol Infect Dis* 2006; **56**: 359–65.

71 Siba V, Horwood PF, Vanuga K, *et al.* Evaluation of serological diagnostic tests for typhoid fever in Papua new guinea using a composite reference standard. *Clin Vaccine Immunol* 2012; **19**: 1833–7.

72 Parry CM, Nguyen Thi Tuyet H, Diep TS, *et al.* Value of a single-tube Widal test in diagnosis of typhoid fever in Vietnam. *J Clin Microbiol* 1999; **37**: 2882–6.

73 Clegg A, Passey M, Omena M, Karigifa K, Suve N. Re-evaluation of the widal agglutination test in response to the changing pattern of typhoid fever in the highlands of Papua New Guinea. *Acta Trop* 1994; **57**: 255–63.

74 Hoffman SL, Flanigan TP, Klaucke D, *et al.* The Widal slide agglutination test, a valuable rapid diagnostic test in typhoid fever patients at the Infectious Diseases Hospital of Jakarta. *Am J Epidemiol* 1986; **123**: 869–75.

75 Massi MN, Shirakawa T, Gotoh A, Bishnu A, Hatta M, Kawabata M. Quantitative detection of Salmonella enterica serovar Typhi from blood of suspected typhoid fever patients by real-time PCR. *Int J Med Microbiol* 2005; **295**: 117–20.

76 Liang L, Juarez S, Nga TVT, *et al.* Immune profiling with a Salmonella Typhi antigen microarray identifies new diagnostic biomarkers of human typhoid. *Sci Rep* 2013; **3**: 1043.

77 Tran Vu Thieu N, Trinh Van T, Tran Tuan A, *et al.* An evaluation of purified Salmonella Typhi protein antigens for the serological diagnosis of acute typhoid fever. J. Infect. 2017. DOI:10.1016/j.jinf.2017.05.007.

78 Näsström E, Vu Thieu NT, Dongol S, *et al.* Salmonella typhi and Salmonella paratyphi a elaborate distinct systemic metabolite signatures during enteric fever. *Elife* 2014; **3**. DOI:10.7554/eLife.03100.

79 Lin FYC, Ho V a., Bay P V., *et al.* The epidemiology of typhoid fever in the Dong Thap Province, Mekong Delta region of Vietnam. *Am J Trop Med Hyg* 2000; **62**: 644–8.

80 Butler T. Treatment of typhoid fever in the 21st century: Promises and shortcomings. *Clin Microbiol Infect* 2011; **17**: 959–63.

81 Smith MD, Duong NM, Hoa NTT, *et al.* Comparison of ofloxacin and ceftriaxone for short-course treatment of enteric fever. *Antimicrob Agents Chemother* 1994; **38**: 1716–20.

82 Phongmany S, Phetsouvanh R, Sisouphone S, *et al.* A randomized comparison of oral chloramphenicol versus ofloxacin in the treatment of uncomplicated typhoid fever in Laos. *Trans R Soc Trop Med Hyg* 2005; **99**: 451–8.

83 Chinh NT, Solomon T, Thong MX, *et al.* Short courses of ofloxacin for the treatment of enteric fever. *Trans R Soc Trop Med Hyg* 1997; **91**: 347–9.

84 Arnold K, Hong CS, Nelwan R, *et al.* Randomized comparative study of fleroxacin and chloramphenicol in typhoid fever. *Am J Med* 1993; **94**: 195S–200S.

85 Cristiano P, Iovene MR, Imperatore C, *et al.* Pefloxacin versus chloramphenicol in the therapy of typhoid fever. *Infection* 1995; **23**: 103–6.

86 Phoung CXT, Kneen R, Anh NT, Luat TD, White NJ, Parry CM DN, Group PCTS. A comparative study of ofloxacin and cefixime for treatment of typhoid fever in children. *Pediatr Infect Dis J* 1999; **18**: 245–8.

87 Chinh NT, Parry CM, Ly NT, *et al.* A randomized controlled comparison of azithromycin and ofloxacin for treatment of multidrug-resistant or nalidixic acid-resistant enteric fever. *Antimicrob Agents Chemother* 2000; **44**: 1855–9.

88 Parry CM, Ho VA, Phuong LT, *et al.* Randomized Controlled Comparison of Ofloxacin, Azithromycin, and an Ofloxacin-Azithromycin Combination for Treatment of Multidrug-Resistant and Nalidixic Acid-Resistant Typhoid Fever. *Antimicrob Agents Chemother* 2007; **51**: 819–25.

89 Threlfall EJ, Ward LR, Skinner JA, Smith HR, Lacey S. Ciprofloxacin-resistant salmonella typhi and treatment failure. *Lancet* 1999; **353**: 1590–1.

90 Crump JA, Kretsinger K, Gay K, *et al.* Clinical response and outcome of infection with Salmonella enterica serotype typhi with decreased susceptibility to fluoroquinolones: A United States FoodNet multicenter retrospective cohort study. *Antimicrob Agents Chemother* 2008; **52**: 1278–84.

91 Frenck RW, Mansour A, Nakhla I, *et al.* Short-course azithromycin for the treatment of uncomplicated typhoid fever in children and adolescents. *Clin Infect Dis* 2004; **38**: 951–7.

92 Dutta P, Mitra U, Dutta S, De A, Chatterjee MK, Bhattacharya SK. Ceftriaxone therapy in ciprofloxacin treatment failure typhoid fever in children. *Indian J Med Res* 2001; **113**: 210–3.

93 Effa EE, Bukirwa H. Azithromycin for treating uncomplicated typhoid and paratyphoid fever (enteric fever). Cochrane Database Syst. Rev. 2008. DOI:10.1002/14651858.CD006083.pub2.

94 Frenck RW, Nakhla I, Sultan Y, *et al.* Azithromycin versus Ceftriaxone for the Treatment of Uncomplicated Typhoid Fever in Children. *Clin Infect Dis* 2000; **31**: 1134–8.

95 Lasserre R, Sangalang RP, Santiago L. Three-day treatment of typhoid fever with two different doses of ceftriaxone, compared to 14-day therapy with chloramphenicol: A randomized trial. *J Antimicrob Chemother* 1991; **28**: 765–72.

96 Islam A, Butler T, Kabir I, Alam NH. Treatment of typhoid fever with ceftriaxone for 5 days or chloramphenicol for 14 days: A randomized clinical trial. *Antimicrob Agents Chemother* 1993; **37**: 1572–5.

97 Soe GB, Overturf GD. Treatment of typhoid fever and other systemic salmonelloses with cefotaxime, ceftriaxone, cefoperazone, and other newer cephalosporins. *Rev Infect Dis* 1987; **9**: 719–736.

98 Pape JW, Gerdes H, Oriol L, Johnson WD. Typhoid fever: Successful therapy with cefoperazone. *J Infect Dis* 1986; **153**: 272–6.

99 Girgis NI, Tribble DR, Sultan Y, Farid Z. Short course chemotherapy with cefixime in children with multidrug-resistant Salmonella typhi septicaemia. *J Trop Pediatr* 1995; **41**: 364–5.

100 BHUTTA ZA, KHAN IA, MOLLA AM. Therapy of multidrug-resistant typhoid fever with oral cefixime vs. intravenous ceftriaxone. *Pediatr Infect Dis J* 1994; **13**: 990–3.

101 Gotuzzo E, Guerra JG, Benavente L, *et al.* Use of norfloxacin to treat chronic typhoid carriers. *J Infect Dis* 1988; **157**: 1221–5.

102 Gotuzzo E, Carrillo C. Quinolones in typhoid fever. *Infect Dis Clin Pract* 1994; **3**: 345–51.

103 Ferreccio C, Morris JG, Valdivieso C, *et al.* Efficacy of ciprofloxacin in the treatment of chronic typhoid carriers. *J Infect Dis* 1988; **157**: 1235–9.

104 Hu L, Kopecko DJ. Salmonella Typhi and Paratyphi. *Evaluation* 2001. DOI:10.1006/bknlm.2001.0066.

105 John Rice Mine. THE INCUBATION PERIOD OF TYPHOID FEVER. *J Infect Dis* 1922; **31**: 296–301.

106 Naylor GRE. INCUBATION PERIOD AND OTHER FEATURES OF FOOD-BORNE AND WATER-BORNE OUTBREAKS OF TYPHOID FEVER IN RELATION TO PATHOGENESIS AND GENETICS OF RESISTANCE. *Lancet* 1983; **321**: 864–6.

107 Wait DA, Sobsey MD. Comparative survival of enteric viruses and bacteria in Atlantic Ocean

seawater. In: Water Science and Technology. 2001: 139–42.

108 Cho JC, Kim SJ. Viable, but non-culturable, state of a green fluorescence protein-tagged environmental isolate of Salmonella typhi in groundwater and pond water. *FEMS Microbiol Lett* 1999; **170**: 257–64.

109 Bhan MK, Bahl R, Bhatnagar S. Typhoid and paratyphoid fever. In: Lancet. 2005: 749–62.

110 Kabwama SN, Bulage L, Nsubuga F, *et al.* A large and persistent outbreak of typhoid fever caused by consuming contaminated water and street-vended beverages: Kampala, Uganda, January - June 2015. *BMC Public Health* 2017; **17**. DOI:10.1186/s12889-016-4002-0.

111 Mermin JH, Villar R, Carpenter J, *et al.* A massive epidemic of multidrug-resistant typhoid fever in Tajikistan associated with consumption of municipal water. *J Infect Dis* 1999; **179**: 1416–22.

112 Caumes E, Ehya N, Nguyen J, Bricaire F. Typhoid and paratyphoid fever: a 10-year retrospective study of 41 cases in a Parisian hospital. *J Travel Med* 2001; **8**: 293–7.

113 Gupta SK, Medalla F, Omondi MW, *et al.* Laboratory-Based Surveillance of Paratyphoid Fever in the United States: Travel and Antimicrobial Resistance. *Clin Infect Dis* 2008; **46**: 1656–63.

114 Lynch MF, Blanton EM, Bulens S, *et al.* Typhoid fever in the United States, 1999-2006. *JAMA - J Am Med Assoc* 2009; **302**: 859–65.

115 Mermin JH, Townes JM, Gerber M, Dolan N, Mintz ED, Tauxe R V. Typhoid fever in the United States, 1985-1994: Changing risks of international travel and increasing antimicrobial resistance. *Arch Intern Med* 1998; **158**: 633–8.

116 Basnyat B, Maskey AP, Zimmerman MD, Murdoch DR. Enteric (typhoid) fever in travelers. *Clin Infect Dis* 2005; **41**: 1467–72.

117 Steinberg EB, Bishop R, Haber P, *et al.* Typhoid Fever in Travelers: Who Should Be Targeted for Prevention? *Clin Infect Dis* 2004; **39**: 186–91.

118 Crump J a, Crump J a, Luby SP, Luby SP, Mintz ED, Mintz ED. The global burden of typhoid fever. *Bull World Health Organ* 2004; **82**: 346–53.

119 Ledingham JCG. Mr N the milker , and Dr Koch ' s concept of the healthy carrier. *Lancet* 1999; **353**: 1354–6.

120 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.

121 Vaishnavi C, Kochhar R, Singh G, Kumar S, Singh S, Singh K. Epidemiology of typhoid carriers among blood donors and patients with biliary, gastrointestinal and other related diseases. *Microbiol Immunol* 2005; **49**: 107–12.

122 Schiøler H, Christiansen ED, Høybye G, Rasmussen SN, Greibe J. Biliary calculi in chronic *Salmonella* carriers and healthy controls: a controlled study. *Scand J Infect Dis* 1983; **15**: 17–9.

123 Marshall JM, Flechtner AD, La Perle KM, Gunn JS. Visualization of extracellular matrix components within sectioned Salmonella biofilms on the surface of human gallstones. *PLoS One* 2014; **9**. DOI:10.1371/journal.pone.0089243.

124 Crawford RW, Rosales-Reyes R, Ramirez-Aguilar M d. l. L, Chapa-Azuela O, Alpuche-Aranda C, Gunn JS. Gallstones play a significant role in Salmonella spp. gallbladder colonization and carriage. *Proc Natl Acad Sci* 2010; **107**: 4353–8.

125 Gonzalez-Escobedo G, Gunn JS. Gallbladder epithelium as a niche for chronic salmonella carriage. *Infect Immun* 2013; **81**: 2920–30.

126 Basnyat B, Baker S. Typhoid carriage in the gallbladder. Lancet. 2015; **386**: 1074.

127 Gonzalez-Escobedo G, Marshall JM, Gunn JS. Chronic and acute infection of the gall bladder by Salmonella Typhi: understanding the carrier state. *Nat Rev Microbiol* 2011; **9**: 9–14.

128 Roumagnac P, Weill F-X, Dolecek C, *et al.* Evolutionary history of Salmonella typhi. *Science* 2006; **314**: 1301–4.

129 Karkey A, Thompson CN, Tran Vu Thieu N, *et al.* Differential epidemiology of Salmonella Typhi and Paratyphi A in Kathmandu, Nepal: a matched case control investigation in a highly endemic enteric fever setting. *PLoS Negl Trop Dis* 2013; **7**: e2391.

130 Dongol S, Thompson CN, Clare S, *et al.* The Microbiological and Clinical Characteristics of Invasive Salmonella in Gallbladders from Cholecystectomy Patients in Kathmandu, Nepal. *PLoS One* 2012; **7**. DOI:10.1371/journal.pone.0047342.

131 Holt KE, Baker S, Dongol S, *et al.* High-throughput bacterial SNP typing identifies distinct clusters of Salmonella Typhi causing typhoid in Nepalese children. *BMC Infect Dis* 2010; **10**: 144.

132 Holt KE, Dolecek C, Chau TT, *et al.* Temporal Fluctuation of Multidrug Resistant Salmonella

Typhi Haplotypes in the Mekong River Delta Region of Vietnam. *PLoS Negl Trop Dis* 2011; **5**: e929.

133 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.

134 Baker S, Holt K, Van De Vosse E, *et al.* High-throughput genotyping of Salmonella enterica serovar Typhi allowing geographical assignment of haplotypes and pathotypes within an urban district of Jakarta, Indonesia. *J Clin Microbiol* 2008; **46**: 1741–6.

135 Olsen SJ, Bleasdale SC, Magnano a R, *et al.* Outbreaks of typhoid fever in the United States, 1960-99. *Epidemiol Infect* 2003; **130**: 13–21.

136 Buckle GC, Walker CLF, Black RE. Typhoid fever and paratyphoid fever: Systematic review to estimate global morbidity and mortality for 2010. *J Glob Health* 2012; **2**: 10401.

137 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 Heal* 2014; **2**: e570–80.

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

139 Shuan C, Teh J, Chua KH, Thong KL. Paratyphoid Fever : Splicing the Global Analyses. 2014; **11**. DOI:10.7150/ijms.7768.

140 Ochiai RL, Wang X, Von Seidlein L, *et al.* Salmonella paratyphi A rates, Asia. *Emerg Infect Dis* 2005; **11**: 1764–6.

141 Fangtham M, Wilde H. Emergence of Salmonella paratyphi A as a major cause of enteric fever: Need for early detection, preventive measures, and effective vaccines. *J Travel Med* 2008; **15**: 344–50.

142 Vollaard AM, Ali S, Van Asten HAGH, *et al.* Risk Factors for Typhoid and Paratyphoid Fever in Jakarta, Indonesia.

http://pmmp.cnki.net/Resources/CDDPdf/evd%5C200801%5CJAMA%5C□ □ □ □ □ %5Cjama 2004291062607.pdf (accessed May 29, 2017).

143 Dong B-Q, Yang J, Wang X-Y, *et al.* Trends and disease burden of enteric fever in Guangxi

province, China, 1994-2004. *Bull World Health Organ* 2010; **88**: 689–96.

144 JAMA. Typhoid in the large cities of the united states in 1918: Seventh annual report. *J Am Med Assoc* 1919; **72**: 997–9.

145 Nga TVT, Parry CM, Le T, *et al.* The decline of typhoid and the rise of non-typhoid salmonellae and fungal infections in a changing HIV landscape: Bloodstream infection trends over 15 years in southern Vietnam. *Trans R Soc Trop Med Hyg* 2012; **106**: 26–34.

146 Record E. Typhoid vaccines: WHO position paper. *Wkly Epidemiol Rec* 2008; **83**: 49–59.

147 Ochiai RL, Acosta CJ, Agtini M, *et al.* The use of typhoid vaccines in Asia: the DOMI experience. *Clin Infect Dis* 2007; **45 Suppl 1**: S34–8.

148 Scobie HM, Nilles E, Kama M, *et al.* Impact of a targeted typhoid vaccination campaign following cyclone Tomas, Republic of Fiji, 2010. *Am J Trop Med Hyg* 2014; **90**: 1031–8.

149 Duncan Steele A, Hay Burgess DC, Diaz Z, Carey ME, M Zaidi AK. Challenges and Opportunities for Typhoid Fever Control: A Call for Coordinated Action. DOI:10.1093/cid/civ976.

150 Mohan VK, Varanasi V, Singh A, *et al.* Safety and Immunogenicity of a Vi Polysaccharide–Tetanus Toxoid Conjugate Vaccine (Typbar-TCV) in Healthy Infants, Children, and Adults in Typhoid Endemic Areas: A Multicenter, 2-Cohort, Open-Label, Double-Blind, Randomized Controlled Phase 3 Study. *Clin Infect Dis* 2015; **61**: 393–402.

151 Bhutta ZA, Capeding MR, Bavdekar A, *et al.* Immunogenicity and safety of the Vi-CRM197 conjugate vaccine against typhoid fever in adults, children, and infants in south and southeast Asia: results from two randomised, observer-blind, age de-escalation, phase 2 trials. *Lancet Infect Dis* 2014; **14**: 119–29.

152 Eng F, Ing Y, In CL, *et al.* THE EFFICACY OF A SALMONELLA TYPHI Vi CONJUGATE VACCINE IN TWO-TO-FIVE-YEAR-OLD CHILDREN Background Typhoid fever is common in devel. *N Engl J Med* 2001; **344**. http://www.nejm.org/doi/pdf/10.1056/NEJM200104263441701 (accessed May 12, 2017).

153 William O, Thomas M. The Principles and practice of medicine, 9th edn. New York, London, D. Appleton and company, 1920.

154 WOODWARD TE, SMADEL JE, LEY HL, GREEN R, MANKIKAR DS. PRELIMINARY

REPORT ON THE BENEFICIAL EFFECT OF CHLOROMYCETIN IN THE TREATMENT OF TYPHOID FEVER. *Ann Intern Med* 1948; **29**: 131.

155 WOODWARD TE, SMADEL JE. MANAGEMENT OF TYPHOID FEVER AND ITS COMPLICATIONS. *Ann Intern Med* 1964; **60**: 144–57.

156 Colquhoun J, Weetch RS. Resistance to Chloramphenicol developing during Treatment of Typhoid Fever. *Lancet* 1950; **25**: 621–3.

157 Murti BR, Rajyalakshmi K, Bhaskaran CS. Resistance of Salmonella typhi to chloramphenicol. I. A preliminary report. *J Clin Pathol* 1962; **15**: 544–51.

158 Paniker CKJ, Vimala KN. Transferable chloramphenicol resistance in Salmonella typhi. *Nature* 1972; **239**: 109–10.

159 Chun D, Seol SY, Cho DT, Tak R. Drug resistance and R plasmids in Salmonella typhi isolated in Korea. *Antimicrob Agents Chemother* 1977; **11**: 209–13.

160 Butler T, Arnold K, Linh N, Pollack M. CHLORAMPHENICOL-RESISTANT TYPHOID FEVER IN VIETNAM ASSOCIATED WITH R FACTOR. *Lancet* 1973; **302**: 983–5.

161 Olarte J, Galindo E. Salmonella typhi resistant to chloramphenicol, ampicillin, and other antimicrobial agents: strains isolated during an extensive typhoid fever epidemic in Mexico. *Antimicrob Agents Chemother* 1973; **4**: 597–601.

162 Anderson ES. The problem and implications of chloramphenicol resistance in the typhoid bacillus. *J Hyg, Camb* 1975; **74**. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2130376/pdf/jhyg00071-0137.pdf (accessed Sept 26, 2017).

163 Multi-drug resistant typhoid: a global problem. *J Med Microbiol*; **44**: 17–3.

164 Rahman M, Siddique a K, Shoma S, *et al.* Emergence of multidrug-resistant Salmonella enterica serotype Typhi with decreased ciprofloxacin susceptibility in Bangladesh. *Epidemiol Infect* 2006; **134**: 433–8.

165 Connerton P, Wain J, Hien TT, *et al.* Epidemic typhoid in Vietnam: Molecular typing of multiple-antibiotic- resistant Salmonella enterica serotype typhi from four outbreaks. *J Clin Microbiol* 2000; **38**: 895–7.

166 Kariuki S, Gilks C, Revathi G, Hart CA. Genotypic analysis of multidrug-resistant Salmonella

enterica Serovar typhi, Kenya. *Emerg Infect Dis* 2000; **6**: 649–51.

167 Shanahan PM, Karamat K a, Thomson CJ, Amyes SG. Characterization of multi-drug resistant Salmonella typhi isolated from Pakistan. *Epidemiol Infect* 2000; **124**: 9–16.

168 Rathish KC, Chandrashekar MR, Nagesha CN. Multidrug resistant Salmonella typhi in Bangalore, south India. *Indian J Med Sci* 1994; **48**: 85–8.

169 Coovadia YM, Gathiram V, Bhamjee A, *et al.* An outbreak of multiresistant salmonella typhi in south africa. *QJM* 1992; **82**: 91–100.

170 Albert MJ, Haider K, Nahar S, Kibriya AKMG, Hossain MA. Multiresistant salmonella typhi in bangladesh. J. Antimicrob. Chemother. 1991; **27**: 554–5.

171 Threlfall EJ, Ward LR, Rowe B, *et al.* Widespread occurrence of multiple drug-resistant Salmonella typhi in India. *Eur J Clin Microbiol Infect Dis* 1992; **11**: 990–3.

172 Shanahan PM, Jesudason M V, Thomson CJ, Amyes SG. Molecular analysis of and identification of antibiotic resistance genes in clinical isolates of Salmonella typhi from India. *J Clin Microbiol* 1998; **36**: 1595–600.

173 Mirza S, Kariuki S, Mamun KZ, Beeching NJ, Hart CA. Analysis of plasmid and chromosomal DNA of multidrug-resistant Salmonella enterica serovar typhi from Asia. *J Clin Microbiol* 2000; **38**: 1449–52.

174 Hermans PWM, Saha SK, Van Leeuwen WJ, Verbrugh HA, Van Belkum A, Goessens WHF. Molecular typing of Salmonella typhi strains from Dhaka (Bangladesh) and development of DNA probes identifying plasmid-encoded multidrug-resistant isolates. *J Clin Microbiol* 1996; **34**: 1373–9.

175 Chiou C-S, Alam M, Kuo J-C, Liu Y-Y, Wang P-J. Chromosome-mediated multidrug resistance in Salmonella enterica serovar Typhi. *Antimicrob Agents Chemother* 2015; **59**: 721–3.

176 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.

177 Pokharel BM, Koirala J, Dahal RK, Mishra SK, Khadga PK, Tuladhar NR. Multidrug-resistant and extended-spectrum beta-lactamase (ESBL)-producing Salmonella enterica (serotypes Typhi and

Paratyphi A) from blood isolates in Nepal: surveillance of resistance and a search for newer alternatives. *Int J Infect Dis* 2006; **10**: 434–8.

178 Butt T, Ahmad RN, Salman M, Kazmi SY. Changing trends in drug resistance among typhoid salmonellae in Rawalpindi, Pakistan. *East Mediterr Heal J* 2005; **11**: 1038–44.

179 Chandel DS, Chaudhry R, Dhawan B, Pandey A, Dey AB. Drug-resistant Salmonella enterica serotype paratyphi A in India. *Emerg Infect Dis* 2000; **6**: 420–1.

180 Holt KE, Thomson NR, Wain J, *et al.* Multidrug-Resistant Salmonella enterica Serovar Paratyphi A Harbors IncHI1 Plasmids Similar to Those Found in Serovar Typhi. *J Bacteriol* 2007; **189**: 4257–64.

181 Brown JC, Shanahan PM a, Jesudason M V., Thomson CJ, Amyes SGB. Mutations responsible for reduced susceptibility to 4-quinolones in clinical isolates of multi-resistant Salmonella typhi in India. *J Antimicrob Chemother* 1996; **37**: 891–900.

182 Mehta G, Randhawa VS, Mohapatra NP. Intermediate susceptibility to ciprofloxacin in Salmonella typhi strains in India. *Eur J Clin Microbiol Infect Dis* 2001; **20**: 760–1.

183 Wain J, Hoa NT, Chinh NT, *et al.* Quinolone-resistant Salmonella typhi in Viet Nam: molecular basis of resistance and clinical response to treatment. *Clin Infect Dis* 1997; **25**: 1404–10.

184 Renuka K, Sood S, Das BK, Kapil A. High-level ciprofloxacin resistance in Salmonella enterica serotype Typhi in India. J. Med. Microbiol. 2005; **54**: 999–1000.

185 Koirala KD, Thanh DP, Thapa SD, *et al.* Highly resistant Salmonella enterica serovar typhi with a novel gyrA mutation raises questions about the long-term efficacy of older fluoroquinolones for treating typhoid fever. *Antimicrob Agents Chemother* 2012; **56**: 2761–2.

186 Menezes GA, Harish BN, Khan MA, Goessens WHF, Hays JP. Antimicrobial resistance trends in blood culture positive Salmonella Typhi isolates from Pondicherry, India, 2005-2009. *Clin Microbiol Infect* 2012; **18**: 239–45.

187 Lee CJ, Su LH, Huang YC, Chiu CH. First isolation of ciprofloxacin-resistant Salmonella enterica serovar Typhi in Taiwan. *J Microbiol Immunol Infect* 2013; **46**: 469–73.

188 García-Fernández A, Gallina S, Owczarek S, *et al.* Emergence of ciprofloxacin-resistant Salmonella enterica serovar Typhi in Italy. *PLoS One* 2015; **10**.

DOI:10.1371/journal.pone.0132065.

189 Harish BN, Menezes GA, Kengamuthu S, Parija SC. A case report and review of the literature: ciprofloxacin resistant Salmonella enterica serovar Typhi in India. (Special Issue: Enteric fever.). *J Infect Dev Ctries* 2008; **2**: 324–7.

190 Ahmed D, D'Costa LT, Alam K, Nair GB, Hossain MA. Multidrug-resistant Salmonella enterica serovar Typhi isolates with high-level resistance to ciprofloxacin in Dhaka, Bangladesh [2]. Antimicrob. Agents Chemother. 2006; **50**: 3516–7.

191 Dutta S, Sur D, Manna B, *et al.* Emergence of highly fluoroquinolone-resistant Salmonella enterica serovar Typhi in a community-based fever surveillance from Kolkata, India. Int. J. Antimicrob. Agents. 2008; **31**: 387–9.

192 Harish BN, Menezes GA, Sarangapani K, Parija SC. A case report and review of the literature: ciprofloxacin resistant Salmonella enterica serovar Typhi in India. J. Infect. Dev. Ctries. 2008; **2**: 324–7.

193 Gupta V, Kaur J, Chander J. An increase in enteric fever cases due to Salmonella Paratyphi A in & around Chandigarh. *Indian J Med Res* 2009; **129**: 95–8.

194 Hassing RJ, Menezes GA, Van Pelt W, Petit PL, Van Genderen PJ, Goessens WHF. Analysis of mechanisms involved in reduced susceptibility to ciprofloxacin in Salmonella enterica serotypes Typhi and Paratyphi A isolates from travellers to Southeast Asia. *Int J Antimicrob Agents* 2011; **37**: 240–3.

195 Threlfall EJ, Fisher IST, Berghold C, *et al.* Trends in antimicrobial drug resistance in Salmonella enterica serotypes Typhi and Paratyphi A isolated in Europe, 1999-2001. *Int J Antimicrob Agents* 2003; **22**: 487–91.

196 Gaind R, Paglietti B, Murgia M, *et al.* Molecular characterization of ciprofloxacin-resistant Salmonella enterica serovar Typhi and Paratyphi A causing enteric fever in India. *J Antimicrob Chemother* 2006; **58**: 1139–44.

197 Maskey AP, Day JN, Tuan PQ, *et al.* Salmonella enterica Serovar Paratyphi A and S. enterica Serovar Typhi Cause Indistinguishable Clinical Syndromes in Kathmandu, Nepal. *Clin Infect Dis* 2006; **42**: 1247–53.

198 Nair L, Sudarsana J, Pushpa KK. Epidemic of Salmonella enteritica serotype paratyphi A in Calicut-Kerala. Calicut Med. J. 2004; **2**: e2.

199 Shirakawa T, Acharya B, Kinoshita S, Kumagai S, Gotoh A, Kawabata M. Decreased susceptibility to fluoroquinolones and gyrA gene mutation in the Salmonella enterica serovar Typhi and Paratyphi A isolated in Katmandu, Nepal, in 2003. *Diagn Microbiol Infect Dis* 2006; **54**: 299–303.

200 Acharya D, Bhatta D, Malla S, *et al.* Salmonella enterica serovar Paratyphi A: an emerging cause of febrile illness in Nepal. *Orig Artic Nepal Med Coll J* 2011; **13**: 69–73.

201 Adachi T, Sagara H, Hirose K, Watanabe H. Fluoroquinolone-resistant Salmonella Paratyphi A. *Emerg Infect Dis* 2005; **11**: 172–4.

202 Harish BN, Madhulika U, Parija SC. Isolated high-level ciprofloxacin resistance in Salmonella enterica subsp. enterica serotype Paratyphi A. J. Med. Microbiol. 2004; **53**: 819.

203 Bhatia JK, Mathur AD, Arora MM. Reemergence of chloramphenicol sensitivity in enteric fever. *Med J Armed Forces India* 2007; **63**: 212–4.

204 Lakshmi V, Ashok R, Susmita J, Shailaja V V. Changing trends in the antibiograms of Salmonella isolates at a tertiary care hospital in Hyderabad. *Indian J Med Microbiol* 2006; **24**: 45–8.

205 Weill FX, Hoang HT, Roumagnac P, *et al.* Clonal reconquest of antibiotic-susceptible Salmonella enterica serotype Typhi in Son La province, Vietnam. *Am J Trop Med Hyg* 2007; **76**: 1174–81.

206 Joshi S. Antibiogram of S. enterica serovar Typhi and S. enterica serovar Paratyphi A: a multi-centre study from India Indian Network for Surveillance of Antimicrobial Resistance Group a. *WHO South-East Asia J Public Heal* 2012; **1**: 182–8.

207 Maskey AP, Basnyat B, Thwaites GE, Campbell JI, Farrar JJ, Zimmerman MD. Emerging trends in enteric fever in Nepal: 9124 cases confirmed by blood culture 1993-2003. *Trans R Soc Trop Med Hyg* 2008; **102**: 91–5.

208 Chau TT, Campbell JI, Galindo CM, *et al.* Antimicrobial drug resistance of Salmonella enterica serovar Typhi in Asia and molecular mechanism of reduced susceptibility to the fluoroquinolones. *Antimicrob Agents Chemother* 2007; **51**: 4315–23.

209 Saha SK, Darmstadt GL, Baqui AH, *et al.* Molecular basis of resistance displayed by highly ciprofloxacin-resistant Salmonella enterica serovar Typhi in Bangladesh. *J Clin Microbiol* 2006;

44: 3811–3.

210 Strahilevitz J, Jacoby GA, Hooper DC, Robicsek A. Plasmid-mediated quinolone resistance: A multifaceted threat. Clin. Microbiol. Rev. 2009; **22**: 664–89.

211 Nüesch-Inderbinen M, Abgottspon H, Sägesser G, Cernela N, Stephan R. Antimicrobial susceptibility of travel-related Salmonella enterica serovar Typhi isolates detected in Switzerland (2002-2013) and molecular characterization of quinolone resistant isolates. *BMC Infect Dis* 2015; **15**: 212.

212 Geetha VK, Yugendran T, Srinivasan R, Harish BN. Plasmid-mediated quinolone resistance in typhoidal Salmonellae: a preliminary report from South India. *Indian J Med Microbiol* 2014; **32**: 31–4.

213 Islam A, Butler T, Kabir I, Alam NH. Treatment of typhoid fever with ceftriaxone for 5 days or chloramphenicol for 14 days: A randomized clinical trial. *Antimicrob Agents Chemother* 1993; **37**: 1572–5.

214 Gokul BN, Menezes GA, Harish BN. Acc-1 beta -lactamase-producing Salmonella enterica serovar typhi, India. *Emerg Infect Dis* 2010; **16**: 1170–1.

215 Phoba M-F, Barbé B, Lunguya O, *et al.* Salmonella enterica serovar Typhi Producing CTX-M-15 Extended Spectrum β-Lactamase in the Democratic Republic of the Congo. *Clin Infect Dis* 2017; **8**: 99–106.

216 Pfeifer Y, Matten J, Rabsch W. Salmonella enterica serovar Typhi with CTX-M beta-lactamase, Germany. *Emerg Infect Dis* 2009; **15**: 1533–5.

217 Dilruba A, Ashfaqul H, Razib M, *et al.* Salmonella enterica serovar Typhi strain producing extended-spectrum beta -lactamases in Dhaka, Bangladesh. *J Med Microbiol* 2012; **61**: 1032–3.

218 González-López JJ, Piedra-Carrasco N, Salvador F, *et al.* ESBL-producing salmonella enterica serovar typhi in traveler returning from guatemala to spain. *Emerg Infect Dis* 2014; **20**: 1918–20.

219 Al Naiemi N, Zwart B, Rijnsburger MC, *et al.* Extended-spectrum-beta-lactamase production in a Salmonella enterica serotype typhi strain from the Philippines. *J Clin Microbiol* 2008; **46**: 2794–5.

220 Munir T, Lodhi M, Ansari JK, Andleeb S, Ahmed M. Extended Spectrum Beta Lactamase producing Cephalosporin resistant Salmonella Typhi, reported from Rawalpindi, Pakistan. *J Pak*

Med Assoc 2016; **66**: 1035–6.

221 Mawatari M, Kato Y, Hayakawa K, *et al.* Salmonella enterica serotype paratyphi a carrying CTX-M-15 type extended-spectrum beta-lactamase isolated from a Japanese traveller returning from India, Japan, July 2013. *Eurosurveillance* 2013; **18**.

222 Aga Khan University Team Investigates Outbreak Of Antibiotic Resistant Typhoid Fever – Take on Typhoid. http://www.coalitionagainsttyphoid.org/aga-khan-university-team-investigates-outbreak-of-antibiotic-resistant-typhoid-fever-in-hyderabad/ (accessed Oct 2, 2017).

223 Parry CM, Tran N, Thieu V, *et al.* Clinically and Microbiologically Derived Azithromycin Susceptibility Breakpoints for Salmonella enterica Serovars Typhi and Paratyphi A. DOI:10.1128/AAC.04729-14.

224 Pascual A, Conejo MC, Garcia I, Perea EJ. Factors affecting the intracellular accumulation and activity of azithromycin. *J Antimicrob Chemother* 1995; **35**: 85–93.

225 Molloy A, Nair S, Cooke FJ, *et al.* First report of Salmonella enterica serotype paratyphi A azithromycin resistance leading to treatment failure. *J Clin Microbiol* 2010; **48**: 4655–7.

226 Kobayashi T, Hayakawa K, Mawatari M, *et al.* Case report: failure under azithromycin treatment in a case of bacteremia due to Salmonella enterica Paratyphi A. *BMC Infect Dis* 2014; **14**: 404.

227 Manesh A, Balaji V, Ragupathi D, Kumar N, Rupali P. A case of clinical and microbiological failure of azithromycin therapy in Salmonella enterica serotype Typhi despite low azithromycin MIC. 2017. DOI:10.1016/j.ijid.2016.11.409.

228 Coburn B, Grassl GA, Finlay B. Salmonella, the host and disease: a brief review. *Immunol Cell Biol* 2007; **85**: 112–8.

229 Crump JA, Sjölund-Karlsson M, Gordon MA, Parry CM. Epidemiology, clinical presentation, laboratory diagnosis, antimicrobial resistance, and antimicrobial management of invasive Salmonella infections. *Clin Microbiol Rev* 2015; **28**: 901–37.

230 Selander RK, Beltran P, Smith NH, *et al.* Evolutionary genetic relationships of clones of Salmonella serovars that cause human typhoid and other enteric fevers. *Infect Immun* 1990; **58**: 2262–75.

231 McClelland M, Sanderson KE, Clifton SW, *et al.* Comparison of genome degradation in Paratyphi

A and Typhi, human-restricted serovars of Salmonella enterica that cause typhoid. *Nat Genet* 2004; **36**: 1268–74.

232 Didelot X, Achtman M, Parkhill J, Thomson NR, Falush D. A bimodal pattern of relatedness between the Salmonella Paratyphi A and Typhi genomes: Convergence or divergence by homologous recombination? DOI:10.1101/gr.5512906.

233 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.

234 McClelland M, Sanderson KE, Spieth J, *et al.* Complete genome sequence of Salmonella enterica serovar Typhimurium LT2. *Nature* 2001; **413**: 852–6.

235 Deng W, Liou S, Iii GP, *et al.* Comparative Genomics of Salmonella enterica Serovar Typhi Strains Ty2 and Comparative Genomics of Salmonella enterica Serovar Typhi Strains Ty2 and CT18 †. *Society* 2003; **185**: 2330–7.

236 Andersson JO, Andersson SGE. Pseudogenes, Junk DNA, and the Dynamics of Rickettsia Genomes. *Mol Biol Evol* 2001; **18**: 829–39.

237 Cole ST, Eiglmeier K, Parkhill J, *et al.* Massive gene decay in the leprosy bacillus. *Nature* 2001; **409**: 1007–11.

238 Wei J, Goldberg MB, Burland V, *et al.* Complete genome sequence and comparative genomics of Shigella flexneri serotype 2a strain 2457T. *Infect Immun* 2003; **71**: 2775–86.

239 Parkhill J, Wren BW, Thomson NR, *et al.* Genome sequence of Yersinia pestis, the causative agent of plague. *Nature* 2001; **413**: 523–7.

240 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.

241 Boyd EF, Wang FS, Beltran P, Plock SA, Nelson K, Selander RK. Salmonella reference collection B (SARB): strains of 37 serovars of subspecies I. *JGenMicrobiol* 1993; **139**: 1125–32.

242 Thong KL, Bhutta Z a., Pang T. Multidrug-resistant strains of Salmonella enterica serotype typhi are genetically homogenous and coexist with antibiotic-sensitive strains as distinct, independent clones. *Int J Infect Dis* 2000; **4**: 194–7.

243 Kidgell C, Reichard U, Wain J, *et al.* Salmonella typhi, the causative agent of typhoid fever, is

approximately 50,000 years old. *Infect Genet Evol* 2002; **2**: 39–45.

244 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.

245 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.

246 Holt KE, Dutta S, Manna B, *et al.* High-resolution genotyping of the endemic Salmonella Typhi population during a Vi (typhoid) vaccination trial in Kolkata. *PLoS Negl Trop Dis* 2012; **6**. DOI:10.1371/journal.pntd.0001490.

247 Emary K, Moore CE, Chanpheaktra N, *et al.* Enteric fever in Cambodian children is dominated by multidrug-resistant H58 Salmonella enterica serovar Typhi with intermediate susceptibility to ciprofloxacin. *Trans R Soc Trop Med Hyg* 2012; **106**: 718–24.

248 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.

249 Hendriksen RS, Leekitcharoenphon P, Lukjancenko O, *et al.* Genomic signature of multidrug-resistant salmonella enterica serovar Typhi isolates related to a massive outbreak in Zambia between 2010 and 2012. *J Clin Microbiol* 2015; **53**: 262–72.

250 Keddy KH, Sooka A, Smith AM, *et al.* Typhoid fever in South Africa in an endemic HIV setting. *PLoS One* 2016; **11**. DOI:10.1371/journal.pone.0164939.

251 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.

252 Feasey NA, Gaskell K, Wong V, *et al.* Rapid Emergence of Multidrug Resistant, H58-Lineage Salmonella Typhi in Blantyre, Malawi. DOI:10.1371/journal.pntd.0003748.

253 National Institute of Statistics. General Population Census of Cambodia 2008. *Minist Plan* 2008; : 1–37.

254 Kirsch H. Watershed Inventory Siem Reap , Cambodia : 2008; : 9–14.

255 Bauer AW, Kirby WM, Sherris JC, Turck M. Antibiotic susceptibility testing by a standardized single disk method. *Am J Clin Pathol* 1966; **45**: 493–6.

256 CLSI. Performance Standards for Antimicrobial Susceptibility Testing; Twenty-Third Informational Supplement (CLSI Document M100-S23). 2013.

257 Pandit A, Arjyal A, Day JN, *et al.* An Open Randomized Comparison of Gatifloxacin versus Cefixime for the Treatment of Uncomplicated Enteric Fever. DOI:10.1371/journal.pone.0000542.

258 Arjyal A, Basnyat B, Koirala S, *et al.* Gatifloxacin versus chloramphenicol for uncomplicated enteric fever: an open-label, randomised, controlled trial. *Lancet Infect Dis* 2011; **11**: 445–54.

259 Koirala S, Basnyat B, Arjyal A, *et al.* Gatifloxacin versus ofloxacin for the treatment of uncomplicated enteric fever in Nepal: an open-label, randomized, controlled trial. *PLoS Negl Trop Dis* 2013; **7**: e2523.

260 Arjyal A, Basnyat B, Thi Nhan H, *et al.* Articles Gatifl oxacin versus ceftriaxone for uncomplicated enteric fever in Nepal: an open-label, two-centre, randomised controlled trial. *Lancet Infect Dis* 2016; **16**: 535–45.

261 Kummu M. Water management in Angkor: Human impacts on hydrology and sediment transportation. *J Environ Manage* 2009; **90**: 1413–21.

262 Jarvis A, Reuter HII, Nelson A, Guevara E. Hole-filled seamless SRTM data V4. Int. Cent. Trop. Agric. 2008; : available from http://srtm.csi.cgiar.org.

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

264 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.

265 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.

266 Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015; **32**: 268–74.

267 Zhou Z, McCann a., Weill F-X, *et al.* Transient Darwinian selection in Salmonella enterica serovar Paratyphi A during 450 years of global spread of enteric fever. *Proc Natl Acad Sci* 2014; **111**:

12199–204.

268 Francisco AP, Vaz C, Monteiro PT, Melo-Cristino J, Ramirez M, Carrio J a. PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinformatics* 2012; **13**: 87.

269 Hunt M, Mather AE, Sánchez-Busó L, *et al.* ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *bioRxiv* 2017; : 1–21.

270 Zankari E, Hasman H, Cosentino S, *et al.* Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 2012; **67**: 2640–4.

271 Carattoli A, Zankari E, Garciá-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.

272 Daniel GR, Zerbino R, Birney E, *et al.* Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res March Genome Res May Cold Spring Harb Lab Press Octobercshlp.org Downloaded from Cold Spring Harb Lab Press Octobercshlp.org Downloaded from Cold Spring Harb Lab Press Oct* 2012; **22**: 557–67.

273 Seemann T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 2014; **30**: 2068–9.

274 Page AJ, Cummins CA, Hunt M, *et al.* Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 2015; **31**: 3691–3.

275 Kulldorff M. A spatial scan statistic. *Commun Stat - Theory Methods* 1997; **26**: 1481–96.

276 Pfeiffer D, Robinson T, Stevenson M, Stevens K, Rogers D, Clements A. Spatial Analysis in Epidemiology. New York City: Oxford University Press, 2008.

277 Kulldorff M, Heffernan R, Hartman J, Assunção R, Mostashari F. A Space–Time Permutation Scan Statistic for Disease Outbreak Detection. *PLoS Med* 2005; **2**: e59.

278 FIRTH D. Bias reduction of maximum likelihood estimates. *Biometrika* 1993; **80**: 27–38.

279 L. KJD& PR. The Statistical Analysis of Failure Time Data, 2nd edn. Hoboken, New Jersey, USA: John Wiley & Sons, 2002.

280 Team R. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2012. *Open access available http//cran r-project org* 2012.

281 Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer, 2009.

282 Cooke FJ, Wain J. The emergence of antibiotic resistance in typhoid fever. *Travel Med Infect Dis* 2004; **2**: 67–74.

283 Parry CM, Vinh H, Chinh NT, *et al.* The Influence of reduced susceptibility to fluoroquinolones in Salmonella enterica serovar Typhi on the clinical response to ofloxacin therapy. *PLoS Negl Trop Dis* 2011; **5**. DOI:10.1371/journal.pntd.0001163.

284 Baker S, Duy PT, Nga TVT, *et al.* Fitness benefits in fluoroquinolone-resistant Salmonella Typhi in the absence of antimicrobial pressure. *Elife* 2013; **2**: e01229.

285 Kadhiravan T, Wig N, Kapil A, Kabra SK, Renuka K, Misra A. Clinical outcomes in typhoid fever: adverse impact of infection with nalidixic acid-resistant Salmonella typhi. *BMC Infect Dis* 2005; **5**: 37.

286 Arjyal A, Basnyat B, Koirala S, *et al.* Gatifloxacin versus chloramphenicol for uncomplicated enteric fever: an open-label, randomised, controlled trial. *Lancet Infect Dis* 2011; **11**: 445–54.

287 Dolecek C, La TTP, Rang NN, *et al.* A multi-center randomised controlled trial of gatifloxacin versus azithromycin for the treatment of uncomplicated typhoid fever in children and adults in Vietnam. *PLoS One* 2008; **3**. DOI:10.1371/journal.pone.0002188.

288 Arjyal A, Basnyat B, Nhan HT, *et al.* Gatifloxacin versus ceftriaxone for uncomplicated enteric fever in Nepal: an open-label, two-centre, randomised controlled trial. *Lancet Infect Dis* 2016; **16**: 535–45.

289 Thanh DP, Karkey A, Dongol S, *et al.* A novel ciprofloxacin-resistant subclade of h58. Salmonella typhi is associated with fluoroquinolone treatment failure. *Elife* 2016; **5**. DOI:10.7554/eLife.14003.

290 Parry CM, Thuy CT, Dongol S, *et al.* Suitable disk antimicrobial susceptibility breakpoints defining Salmonella enterica serovar Typhi isolates with reduced susceptibility to fluoroquinolones. *Antimicrob Agents Chemother* 2010; **54**: 5201–8.

291 Koirala KD, Thanh DP, Thapa SD, *et al.* Highly resistant Salmonella enterica serovar Typhi with a novel gyrA mutation raises questions about the long-term efficacy of older fluoroquinolones for treating typhoid fever. *Antimicrob Agents Chemother* 2012; **56**: 2761–2.

292 Lu T, Zhao X, Drlica K. Gatifloxacin activity against quinolone-resistant gyrase: allele-specific

enhancement of bacteriostatic and bactericidal activities by the C-8-methoxy group. *Antimicrob Agents Chemother* 1999; **43**: 2969–74.

293 CLSI. Clinical and Laboratory Standards Institute: Performance standards for antimicrobial susceptibility testing; twentieth informational supplement. 2012.

294 Lee C-J, Su L-H, Huang Y-C, Chiu C-H. First isolation of ciprofloxacin-resistant Salmonella enterica serovar Typhi in Taiwan. *J Microbiol Immunol Infect* 2013; **46**: 469–73.

295 García-Fernández A, Gallina S, Owczarek S, *et al.* Emergence of Ciprofloxacin-Resistant Salmonella enterica Serovar Typhi in Italy. *PLoS One* 2015; **10**: e0132065.

296 Schwartz E. Typhoid and Paratyphoid Fever. *Trop Dis Travel* 2010; **366**: 144–53.

297 Maskey AP, Day JN, Phung QT, *et al.* Salmonella enterica serovar Paratyphi A and S. enterica serovar Typhi cause indistinguishable clinical syndromes in Kathmandu, Nepal. *Clin Infect Dis* 2006; **42**: 1247–53.

298 Maskey AP, Basnyat B, Thwaites GE, Campbell JI, Farrar JJ, Zimmerman MD. Emerging trends in enteric fever in Nepal: 9124 cases confirmed by blood culture 1993–2003. *Trans R Soc Trop Med Hyg* 2008; **102**: 91–5.

299 Harish BN, Madhulika U PS. C orrespondence. *J Med Microbiol* 2004; **53**: 819.

300 Baker S, Favorov M, Dougan G. Searching for the elusive typhoid diagnostic. *BMC Infect Dis* 2010; **10**: 45.

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

302 Karkey A, Arjyal A, Anders KL, *et al.* The Burden and Characteristics of Enteric Fever at a Healthcare Facility in a Densely Populated Area of Kathmandu. 2010. DOI:10.1371/journal.pone.0013988.

303 Mirold S, Rabsch W, Rohde M, *et al.* Isolation of a temperate bacteriophage encoding the type III effector protein SopE from an epidemic Salmonella typhimurium strain. *Proc Natl Acad Sci U S A* 1999; **96**: 9845–50.

304 Zhao W, Moest T, Zhao Y, *et al.* The Salmonella effector protein SifA plays a dual role in

virulence. *Sci Rep* 2015; **5**: 12979.

305 Liu Y, Ho KK, Su J, Gong H, Chang AC, Lu S. Potassium transport of Salmonella is important for type III secretion and pathogenesis. *Microbiol (United Kingdom)* 2013; **159**: 1705–19.

306 Shrestha KL, Pant ND, Bhandari R, Khatri S, Shrestha B, Lekhak B. Re-emergence of the susceptibility of the Salmonella spp. isolated from blood samples to conventional first line antibiotics. *Antimicrob Resist Infect Control* 2016. DOI:10.1186/s13756-016-0121-8.

307 Karkey A, Jombart T, Walker AW, *et al.* The Ecological Dynamics of Fecal Contamination and Salmonella Typhi and Salmonella Paratyphi A in Municipal Kathmandu Drinking Water. *PLoS Negl Trop Dis* 2016; **10**: e0004346.

308 Kuijpers LMF, Le Hello S, Fawal N, *et al.* Genomic analysis of Salmonella enterica serotype Paratyphi A during an outbreak in Cambodia, 2013-2015. *Microb genomics* 2016; **2**: e000092.

309 Crump JA, Mintz ED. Global Trends in Typhoid and Paratyphoid Fever. *Clin Infect Dis* 2010; **50**: 241–6.

310 Parry CM, Wijedoru L, Arjyal A, Baker S. The utility of diagnostic tests for enteric fever in endemic locations. Expert Rev. Anti. Infect. Ther. 2011; **9**: 711–25.

311 Brooks WA, Hossain A, Goswami D, *et al.* Bacteremic typhoid fever in children in an urban slum, Bangladesh. *Emerg Infect Dis* 2005; **11**: 326–9.

312 Khan MI, Ochiai RL, Soofi SB, *et al.* Risk factors associated with typhoid fever in children aged 2-16 years in Karachi, Pakistan. *Epidemiol Infect* 2012; **140**: 665–72.

313 Kasper MR, Sokhal B, Blair PJ, Wierzba TF, Putnam SD. Emergence of multidrug-resistant Salmonella enterica serovar Typhi with reduced susceptibility to fluoroquinolones in Cambodia. *Diagn Microbiol Infect Dis* 2010; **66**: 207–9.

314 Khan MI, Ochiai RL, Soofi SB, *et al.* Risk factors associated with typhoid fever in children aged 2-16 years in Karachi, Pakistan. *Epidemiol Infect* 2012; **140**: 665–72.

315 Nichols C, Cruz Espinoza LM, Von Kalckreuth V, *et al.* Bloodstream infections and frequency of pretreatment associated with age and hospitalization status in Sub-Saharan Africa. *Clin Infect Dis* 2015; **61**: S372–9.

316 Cruz Espinoza LM, Nichols C, Adu-Sarkodie Y, *et al.* Variations of Invasive Salmonella Infections

by Population Size in Asante Akim North Municipal, Ghana. *Clin Infect Dis* 2016; **62 Suppl 1**: S17-22.

317 Date KA, Bentsi-Enchill A, Marks F, Fox K. Typhoid fever vaccination strategies. *Vaccine* 2015; **33**: C55–61.

318 Crump JA. Building the case for wider use of typhoid vaccines. *Vaccine* 2015; **33**: C1–2.

319 Mai NL, Phan VB, Vo AH, *et al.* Persistent efficacy of Vi conjugate vaccine against typhoid fever in young children. *N Engl J Med* 2003; **349**: 1390–1.

320 Carrier of typhoid fevver-1912. ; **i**.

321 Gonzalez-Escobedo G, Marshall JM, Gunn JS. Chronic and acute infection of the gall bladder by Salmonella Typhi: understanding the carrier state. DOI:10.1038/nrmicro2490.

322 Bokkenheuser V. Detection of typhoid carriers. *Am J Public Health* 1964; **54**: 477–86.

323 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.

324 Prouty AM, Schwesinger WH, Gunn JS. Biofilm formation and interaction with the surfaces of gallstones by Salmonella spp. *Infect Immun* 2002; **70**: 2640–9.

325 Mian MF, Pek EA, Chenoweth MJ, Coombes BK, Ashkar AA. Humanized mice for Salmonella typhi infection: new tools for an old problem. *Virulence* 2011; **2**: 248–52.

326 Mateen MA, Saleem S, Rao PC, Reddy PS, Reddy DN. Ultrasound in the diagnosis of Typhoid fever. *Indian J Pediatr* 2006; **73**: 681–5.

327 Mathur R, Oh H, Zhang D, *et al.* A mouse model of salmonella typhi infection. *Cell* 2012; **151**: 590–602.

328 Crawford RW, Gibson DL, Kay WW, Gunn JS. Identification of a bile-induced exopolysaccharide required for salmonella biofilm formation on gallstone surfaces. *Infect Immun* 2008; **76**: 5341–9.

329 Marshall JM, Flechtner AD, La Perle KM, Gunn JS, Heuser J. Visualization of Extracellular Matrix Components within Sectioned Salmonella Biofilms on the Surface of Human Gallstones. *PLoS One* 2014; **9**: e89243.

330 Menendez A, Arena ET, Guttman JA, *et al. Salmonella* Infection of Gallbladder Epithelial Cells Drives Local Inflammation and Injury in a Model of Acute Typhoid Fever. *J Infect Dis* 2009; **200**:

1703–13.

331 Yap KP, Gan HM, Teh CSJ, *et al.* Genome sequence and comparative pathogenomics analysis of a salmonella enterica serovar typhi strain associated with a typhoid carrier in Malaysia. J. Bacteriol. 2012; **194**: 5970–1.

332 Baddam R, Kumar N, Shaik S, Lankapalli AK, Ahmed N. Genome dynamics and evolution of Salmonella Typhi strains from the typhoid-endemic zones. *Sci Rep* 2014; **4**. DOI:10.1038/srep07457.

333 Ong SY, Pratap CB, Wan X, *et al.* The Genomic Blueprint of Salmonella enterica subspecies enterica serovar Typhi P-stx-12. *Stand Genomic Sci* 2013; **7**: 483–96.

334 Baddam R, Kumar N, Shaik S, *et al.* Genome sequencing and analysis of Salmonella enterica serovar Typhi strain CR0063 representing a carrier individual during an outbreak of typhoid fever in Kelantan, Malaysia. *Gut Pathog* 2012; **4**: 20.

335 Sharma A, Qadri A. Vi polysaccharide of Salmonella typhi targets the prohibitin family of molecules in intestinal epithelial cells and suppresses early inflammatory responses. *Proc Natl Acad Sci* 2004; **101**: 17492–7.

336 Losonsky GA, Ferreccio C, Kotloff KL, Kaintuck S, Robbins JB, Levine MM. Development and Evaluation of an Enzyme-Linked Immunosorbent Assay for Serum Vi Antibodies for Detection of Chronic Salmonella typhi Carriers. 1987; **25**: 2266–9.

337 Nolan CM, White PC, Feeley JC, Brown SL, Hambie EA, Wong KH. Vi SEROLOGY IN THE DETECTION OF TYPHOID CARRIERS. *Lancet* 1981; **317**: 583–5.

338 Virlogeux I, Waxin H, Ecobichon C, Popoff MY. Role of the viaB locus in synthesis, transport and expression of Salmonella typhi Vi antigen. *Microbiology* 1995; **141**: 3039–47.

339 Santander J, Wanda SY, Nickerson CA, Curtiss R. Role of RpoS in fine-tuning the synthesis of Vi capsular polysaccharide in Salmonella enterica serotype Typhi. *Infect Immun* 2007; **75**: 1382–92.

340 Hirose K, Ezaki T, Miyake M, *et al.* Survival of Vi-capsulated and Vi-deleted Salmonella typhi strains in cultured macrophage expressing different levels of CD14 antigen. *FEMS Microbiol Lett* 1997; **147**: 259–65.

341 Steinsiek S, Bettenbrock K. Glucose transport in Escherichia coli mutant strains with defects in

sugar transport systems. *J Bacteriol* 2012; **194**: 5897–908.

342 Felix A. DETECTION OF CHRONIC TYPHOID CARRIERS BY AGGLUTINATION TESTS. *Lancet* 1938; **232**: 738–41.

343 House D, Ho VA, Diep TS, *et al.* Antibodies to the Vi capsule of Salmonella Typhi in the serum of typhoid patients and healthy control subjects from a typhoid endemic region. *J Infect Dev Ctries* 2008; **2**: 308–12.

344 Gupta A, My Thanh NT, Olsen SJ, *et al.* Evaluation of community-based serologic screening for identification of chronic Salmonella Typhi carriers in Vietnam. *Int J Infect Dis* 2006; **10**: 309–14.

345 Coynault C, Robbe-Saule V, Norel F. Virulence and vaccine potential of Salmonella typhimurium mutants deficient in the expression of the RpoS (σ(S)) regulon. *Mol Microbiol* 1996; **22**: 149–60.

346 Nickerson CA, Curtiss R. Role of sigma factor RpoS in initial stages of Salmonella typhimurium infection. *Infect Immun* 1997; **65**: 1814–23.

347 Santander J, Wanda S-Y, Nickerson CA, Curtiss R, III. Role of RpoS in fine-tuning the synthesis of Vi capsular polysaccharide in Salmonella enterica serotype Typhi. *Infect Immun* 2007; **75**: 1382–92.

348 Prouty AM, Van Velkinburgh JC, Gunn JS. Salmonella enterica serovar typhimurium resistance to bile: Identification and characterization of the tolQRA cluster. *J Bacteriol* 2002; **184**: 1270–6.

349 Gunn JS. Mechanisms of bacterial resistance and response to bile. *Microbes Infect* 2000; **2**: 907–13.

350 Kim SH, Wei CI. Molecular characterization of biofilm formation and attachment of Salmonella enterica serovar typhimurium DT104 on food contact surfaces. *J Food Prot* 2009; **72**: 1841–7.

Chapter 9

Appendices

Appendix A *Salmonella* Typhi isolates and their corresponding sequencing metadata in chapter 3

Name	Tree ID	Accession number	Trial	Year	Country	gyrA	gyrA	gyrA	gyrB	gyrB	parC	parE	parE	parE	H58	MIC	MIC group*
						S83F	D87N	D87V	S464Y	A574V	S80I	A353V	A364V	E460K		Cipro	
02TY_001	STY1_AL513382	ERR1079230	1	2011	Nepal	0	0	0	0	0	0	0	0	0	1	0.016	1
02TY_006	STY6_AL513382	ERR1079231	1	2011	Nepal	1		0	0	0	0	0	0	0	1	0.38	2
02TY_007	STY7_AL513382	ERR1079232	1	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
02TY_009	13STY9_AL513382	ERR1079233	1	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.38	2
02TY_015	STY15_MERGE_AL513382	ERR1079234	1	2012	Nepal	1	0	0	0	0	0	0	0	0	1	0.38	2
02TY_016	STY16_AL513382	ERR1079235	1	2012	Nepal	1	0	0	0	0	0	0	0	0	1	0.38	2
02TY_018	STY18_AL513382	ERR1079236	1	2012	Nepal	1	0	0	0	0	0	0	0	0	1	0.38	2
02TY_027	STY27_MERGE_AL513382	ERR1079237	1	2012	Nepal	1	0	0	0	0	0	0	0	0	1	0.38	2
02TY_028	STY28_MERGE_AL513382	ERR1079238	1	2012	Nepal	1	0	0	0	0	0	0	0	0	1	0.38	2
02TY_030	STY30_AL513382	ERR1079239	1	2012	Nepal	1	0	0	0	0	0	0	0	0	1	0.38	2
02TY_031	STY31_AL513382	ERR1079240	1	2012	Nepal	1	0	0	0	0	0	0	0	0	1	0.38	2
02TY_032	STY32_AL513382	ERR1079241	1	2012	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
02TY_034	STY34_AL513382	ERR1079242	1	2012	Nepal	1	0	0	0	0	0	0	0	0	1	0.5	2
02TY_037	STY37_AL513382	ERR1079243	1	2012	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
02TY_038	STY38_AL513382	ERR1079244	1	2012	Nepal	1	0	0	0	0	0	0	0	0	1	0.38	2
02TY_045	STY45_MERGE_AL513382	ERR1079245	1	2012	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
02TY_047	STY47_AL513382	ERR1079246	1	2012	Nepal	1	0	0	0	0	0	0	0	0	1	0.38	2

02TY_048	STY48_AL513382	ERR1079247	1	2012	Nepal	1	0	0	0	0	0	0	0	0	1	0.125	2
02TY_051	STY51_AL513382	ERR1079248	1	2012	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
02TY_055	STY55_AL513382	ERR1079249	1	2012	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
02TY_057	STY57_AL513382	ERR1079250	1	2012	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
02TY_058	4STY58D1_AL513382	ERR1079251	1	2012	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
02TY_059	STY59_AL513382	ERR1079253	1	2012	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
02TY_060	STY60_AL513382	ERR1079254	1	2012	Nepal	1	0	0	0	0	0	0	0	0	1	0.19	2
02TY_061	STY61_AL513382	ERR1079255	1	2012	Nepal	0	0	0	0	0	0	0	1	0	0	0.012	1
02TY_078	14STY78_AL513382	ERR1079256	1	2012	Nepal	0	0	0	0	0	0	0	0	0	0	0.016	1
02TY_080	STY80_AL513382	ERR1079257	1	2012	Nepal	1	0	0	0	0	0	0	0	0	1	0.19	2
02TY_083	STY83_AL513382	ERR1079258	1	2012	Nepal	1	0	0	0	0	0	0	0	0	1	0.5	2
02TY_119	15STY119D1_AL513382	ERR1079279	1	2013	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
02TY_125	STY125_MERGE_AL513382	ERR1079260	1	2013	Nepal	1	0	0	0	0	0	0	0	0	1	0.38	2
02TY_126	6STY126_AL513382	ERR1079261	1	2013	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
02TY_127	7STY127_AL513382	ERR1079262	1	2013	Nepal	0	1	0	0	0	0	0	0	0	0	0.125	2
02TY_132	15STY132_AL513382	ERR1079263	1	2013	Nepal	1	1	0	0	0	1	0	0	0	1	32	3
02TY_135	16STY135_AL513382	ERR1079264	1	2013	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
02TY_136	STY136_MERGE_AL513382	ERR1079265	1	2013	Nepal	0	0	0	0	0	0	0	0	0	1	0.023	1
02TY_140	16STY140_AL513382	ERR1079266	1	2013	Nepal	0	0	0	0	0	0	0	0	0	1	0.023	1
02TY_143	9STY143D26_AL513382	ERR1079267	1	2013	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
02TY_144	10STY144_AL513382	ERR1079268	1	2013	Nepal	1	0	0	0	0	0	0	0	0	1	0.19	2
02TY_146	11STY146_AL513382	ERR1079269	1	2013	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
02TY_147	17STY147_AL513382	ERR1079270	1	2013	Nepal	1	0	0	0	1	0	0	1	0	0	1	3
02TY_148	17STY148_AL513382	ERR1079271	1	2013	Nepal	1	0	0	0	1	0	0	1	0	0	0.38	2
02TY_149	12STY149_AL513382	ERR1079272	1	2013	Nepal	1	0	0	0	0	0	0	0	0	1	0.38	2
02TY_150	13STY150_AL513382	ERR1079273	1	2013	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
02TY_156	14STY156_AL513382	ERR1079274	1	2013	Nepal	1	0	0	0	0	0	0	0	0	1	0.38	2
02TY_157	18STY157_AL513382	ERR1079275	1	2013	Nepal	0	0	0	0	0	0	1	1	0	0	0.008	1

02TY_158	15STY158_AL513382	ERR1079276	1	2013	Nepal	1	0	0	0	0	0	0	0	0	1	0.38	2
02TY_159	16STY159_AL513382	ERR1079277	1	2013	Nepal	1	0	0	0	0	0	0	0	0	1	1	3
02TY_160	17STY160_AL513382	ERR1079278	1	2013	Nepal	1	1	0	0	0	1	0	0	0	1	32	3
02TY_163	18STY163_AL513382	ERR1079279	1	2013	Nepal	1	1	0	0	0	1	0	0	0	1	32	3
02TY_164	19STY164_AL513382	ERR1079280	1	2013	Nepal	0	0	0	0	0	0	0	0	0	0	0.016	1
02TY_169	STY169_MERGE_AL513382	ERR1079281	1	2013	Nepal	1	1	0	0	0	1	0	0	0	1	24	3
02TY_173	STY173_MERGE_AL513382	ERR1079282	1	2013	Nepal	1	0	0	0	0	0	0	0	0	1	0.38	2
02TY_174	19STY174_AL513382	ERR1079283	1	2013	Nepal	1	0	1	0	0	1	0	1	0	0	32	3
02TY_176	22STY176D1_AL513382	ERR1079284	1	2013	Nepal	1	1	0	0	0	1	0	0	0	1	32	3
02TY_178	24STY178_AL513382	ERR1079286	1	2013	Nepal	1	0	1	0	0	1	0	1	0	0	32	3
02TY_183	13254_5#37	ERR586912	1	2013	Nepal	1	1	0	0	0	1	0	0	0	1	32	3
02TY_185	25STY185_AL513382	ERR1079287	1	2013	Nepal	0	0	0	1	0	0	0	0	0	0	0.15	2
02TY_186	26STY186_AL513382	ERR1079288	1	2013	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
02TY_188	STY188_MERGE_AL513382	ERR1079289	1	2013	Nepal	1	1	0	0	0	1	0	0	0	1	32	3
02TY_190	13254_5#39	ERR586914	1	2013	Nepal	1	1	0	0	0	1	0	0	0	1	24	3
02TY_195	STY195_MERGE_AL513382	ERR1079290	1	2013	Nepal	1	0	0	0	0	0	0	0	0	1	0.19	2
02TY_196	29STY196_AL513382	ERR1079291	1	2013	Nepal	0	0	0	0	0	0	0	0	0	1	0.016	1
02TY_200	30STY200_AL513382	ERR1079292	1	2013	Nepal	0	0	0	0	0	0	0	0	0	0	0.016	1
02TY_202	20STY202_AL513382	ERR1079293	1	2013	Nepal	1	0	0	0	0	0	0	1	0	0	0.25	2
02TY_210	18STY210_AL513382	ERR1079294	1	2013	Nepal	1	1	0	0	0	1	0	0	0	1	32	3
02TY_213	24STY213_AL513382	ERR1079295	1	2014	Nepal	0	0	0	0	0	0	0	1	0	0	0.016	1
02TY_216	31STY216_AL513382	ERR1079296	1	2014	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
02TY_219	21STY219D1_AL513382	ERR1079297	1	2014	Nepal	1	1	0	0	0	1	0	0	0	1	32	3
02TY_222	23STY222_AL513382	ERR1079299	1	2014	Nepal	1	0	0	0	0	0	0	0	1	1	0.38	2
02TY_226	32STY226_AL513382	ERR1079300	1	2014	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
02TY_229	20STY229_AL513382	ERR1079301	1	2014	Nepal	1	1	0	0	0	1	0	0	0	1	32	3
02TY_232	STY232_MERGE_AL513382	ERR1079302	1	2014	Nepal	1	1	0	0	0	1	0	0	0	1	32	3
02TY_235	24STY235D1_AL513382	ERR1079303	1	2014	Nepal	1	0	0	0	0	0	0	0	0	1	0.38	2

02TY_236	STY236_MERGE_AL513382	ERR1079305	1	2014	Nepal	1	0	0	0	0	0	0	0	0	1	0.38	2
02TY_239	28STY239_AL513382	ERR1079306	1	2014	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
02TY_240	29STY240_AL513382	ERR1079307	1	2014	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
02TY_241	30STY241_AL513382	ERR1079308	1	2014	Nepal	1	0	0	0	0	0	0	0	0	1	0.5	2
02TY_244	12STY244_AL513382	ERR1079309	1	2014	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
01TY072	5886_3#1	ERR119817	0	2008	Nepal	0	0	0	0	0	0	0	0	0	0	0.016	1
01TY075	5886_3#2	ERR119818	0	2008	Nepal	0	0	0	0	0	0	0	0	0	0	0.016	1
01TY090	5886_3#4	ERR119820	0	2009	Nepal	0	0	0	0	0	0	0	0	0	1	0.016	1
01TY098	5886_3#5	ERR119821	0	2009	Nepal	0	0	0	0	0	0	0	0	0	0	0.016	1
01TY101	5886_3#6	ERR119822	0	2009	Nepal	0	0	0	0	0	0	0	0	0	1	0.016	1
01TY103	5886_3#3	ERR119819	0	2009	Nepal	0	0	0	0	0	0	0	0	0	1	0.016	1
01TY104	5886_3#7	ERR119823	0	2009	Nepal	0	0	0	0	0	0	0	0	0	1	0.016	1
01TY112	5886_3#8	ERR119824	0	2009	Nepal	0	0	0	0	0	0	0	0	0	1	0.016	1
01TY122	5886_3#9	ERR119825	0	2009	Nepal	0	0	0	0	0	0	0	0	0	1	0.016	1
CC_04	10561_2#1	ERR357576	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.19	2
CC_05	10561_2#8	ERR357583	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.19	2
CC_08	10561_2#12	ERR357587	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.19	2
CC_11	10561_2#19	ERR357594	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
CC_12	10561_2#25	ERR357600	0	2011	Nepal	0	0	0	0	0	0	0	1	0	0	0.004	1
CC_13	10561_2#31	ERR357606	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.19	2
CC_16	10561_2#37	ERR357612	0	2011	Nepal	0	0	0	0	0	0	0	1	0	0	0.004	1
CC_22	10561_2#44	ERR357619	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
CC_27	10425_1#9	ERR349339	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.38	2
CC_31	10561_2#53	ERR357628	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
CC_39	10561_2#57	ERR357632	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
CC_40	10561_2#62	ERR357637	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.38	2
CC_41	10561_2#2	ERR357577	0	2011	Nepal	1	0	0	0	0	0	0	1	0	0	0.125	2
CC_43	10561_2#9	ERR357584	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.38	2

CC_44	10561_2#13	ERR357588	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
CC_45	10425_1#5	ERR349335	0	2011	Nepal	0	0	0	0	0	0	0	1	0	0	0.008	1
CC_49	10561_2#26	ERR357601	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
CC_50	10561_2#32	ERR357607	0	2011	Nepal	0	1	0	0	0	0	0	0	0	0	0.125	2
CC_51	10561_2#38	ERR357613	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.38	2
CC_57	10561_2#45	ERR357620	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.38	2
CC_60	10561_2#50	ERR357625	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.19	2
CC_61	10426_1#6	ERR349528	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.19	2
CC_65	10561_2#58	ERR357633	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.125	2
CC_67	10561_2#63	ERR357638	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.19	2
CC_69	10425_1#2	ERR349332	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.125	2
CC_74	10425_1#3	ERR349333	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
CC_75	10561_2#14	ERR357589	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
CC_80	10425_1#6	ERR349336	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.19	2
CC_81	10561_2#27	ERR357602	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.19	2
CC_83	10561_2#33	ERR357608	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
CC_87	10561_2#39	ERR357614	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.25	2
CC_88	10561_2#46	ERR357621	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.19	2
CC_89	10561_2#51	ERR357626	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.19	2
CC_90	10561_2#54	ERR357629	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.19	2
CC_92	10561_2#59	ERR357634	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.19	2
CC_93	10561_2#64	ERR357639	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.19	2
CC_94	10561_2#3	ERR357578	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.19	2
CC_95	10561_2#10	ERR357585	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.19	2
CC_97	10561_2#15	ERR357590	0	2011	Nepal	1	0	0	0	0	0	0	0	0	1	0.19	2
CC_99	10561_2#20	ERR357595	0	2011	Nepal	1	0	0	0	0	0	0	1	0	0	0.094	2
Ind082356_2008	10349_1#34	ERR343282	0	2008	India	1	1	0	0	0	1	0	0	0	1	32	3
Ind101104_2010	10349_1#52	ERR343300	0	2010	India	1	1	0	0	0	1	0	0	0	1	32	3

IndBCR175_2011	8616_4#48	ERR420424	0	2011	India	1	1	0	0	0	1	0	0	0	1	32	3
IndBCR191_2011	8616_4#50	ERR420426	0	2011	India	1	1	0	0	0	1	0	0	0	1	32	3
IndBCR211_2011	8616_4#51	ERR420427	0	2011	India	1	1	0	0	0	1	0	0	0	1	32	3
IndBCR62_2009	8616_4#41	ERR420417	0	2009	India	1	1	0	0	0	1	0	0	0	1	32	3
IndMDUST127_2011	10492_1#6	ERR352259	0	2011	India	1	1	0	0	0	1	0	0	0	1	32	3
IndMDUST128_2011	10493_1#9	ERR352434	0	2011	India	1	1	0	0	0	1	0	0	0	1	32	3
IndMDUST147_2012	10492_1#13	ERR352266	0	2012	India	1	1	0	0	0	1	0	0	0	1	32	3
IndMDUST167_2011	10562_2#13	ERR357768	0	2011	India	1	1	0	0	0	1	0	0	0	1	24	3
IndMDUST197_2010	10492_1#25	ERR352278	0	2010	India	1	1	0	0	0	1	0	0	0	1	32	3
IndMDUST216_2011	10492_1#31	ERR352284	0	2011	India	1	1	0	0	0	1	0	0	0	1	32	3
IndMDUST247_2011	10562_2#32	ERR357787	0	2011	India	1	1	0	0	0	1	0	0	0	1	32	3
IndMDUST248_2011	10492_1#43	ERR352296	0	2011	India	1	1	0	0	0	1	0	0	0	1	24	3
IndMDUST254_2010	10562_2#34	ERR357789	0	2010	India	1	1	0	0	0	1	0	0	0	1	32	3
IndMDUST400_2012	10562_2#79	ERR357834	0	2012	India	1	1	0	0	0	1	0	0	0	1	24	3
IndMDUST408_2012	10492_1#82	ERR352335	0	2012	India	1	1	0	0	0	1	0	0	0	1	32	3
IndSP80_2011	8616_4#58	ERR420434	0	2011	India	1	1	0	0	0	1	0	0	0	1	32	3
NEN_2512	13254_5#40	ERR586915	0	2013	Nepal	1	1	0	0	0	1	0	0	0	1	32	3
NEN_2533	2533_AL513382	ERR1079310	0	2013	Nepal	1	0	0	0	0	0	0	0	0	1	0.38	2
NEN_2546	2546_AL513382	ERR1079311	0	2013	Nepal	1	0	1	0	0	1	0	1	0	0	6	3
OTHERS_510	13254_5#42	ERR586917	0	2013	Nepal	1	1	0	0	0	1	0	0	0	1	32	3
OTHERS_514	13254_5#43	ERR586918	0	2013	Nepal	1	0	1	0	0	1	0	1	0	0	8	3
OTHERS_515	13254_5#44	ERR586919	0	2013	Nepal	1	1	0	0	0	1	0	0	0	1	32	3
OTHERS_516	13254_5#45	ERR586920	0	2013	Nepal	1	0	1	0	0	1	0	1	0	0	8	3
OTHERS_521	13254_5#46	ERR586921	0	2013	Nepal	1	1	0	0	0	1	0	0	0	1	8	3
OTHERS_535	535_AL513382	ERR1079312	0	2013	Nepal	1	0	1	0	0	1	0	1	0	0	32	3
UnkMDUST146_2012	10562_2#10	ERR357765	0	2012	Unknown	1	1	0	0	0	1	0	0	0	1	32	3

* ciprofloxacin MIC group: 1; susceptible, 2; intermediate, 3; resistan

Appendix B *Salmonella* Paratyphi A isolates and their corresponding metadata in chapter 4

Strain	Acute/Carrier	Lineage	Genotype	Lat	Long	GyrA	Date	Age	Sex	Main_water	Treat_water	Anorexia	Nausea	Abdominal pain	Diarrhoea	Constipation	Headache
DM035	Acute	A	A1.6	27.69116	85.34435	S83F	17-Jun-05	23	F			Yes	Yes	Yes	No	Yes	Yes
DM038	Acute	A	A1	27.65838	85.32773	S83F	19-Jun-05	25	M			Yes	Yes	Yes	Yes	No	Yes
DM042	Acute	A	A1.6	27.6898	85.34475	S83F	20-Jun-05	30	M			Yes	No	No	No	No	Yes
DM049	Acute	A	A1	27.68463	85.32151	WT	21-Jun-05	22	M			Yes	No	No	No	No	Yes
DM056	Acute	A	A1.1	27.69708	85.33345	S83F	25-Jun-05	11	M			Yes	No	No	No	No	Yes
DM067	Acute	A	A1.2	27.6692	85.32035	S83F	27-Jun-05	16	M			Yes	Yes	Yes	No	No	Yes
DM072	Acute	A	A1.7	27.67321	85.32898	S83F	28-Jun-05	31	M			Yes	Yes	No	No	No	Yes
DM080	Acute	A	A1	27.67098	85.3219	S83F	29-Jun-05	6	M			No	No	Yes	No	No	Yes
DM082	Acute	A	A1.2	27.67166	85.32711	S83F	29-Jun-05	50	F			Yes	No	No	No	Yes	Yes
DM089	Acute	A	A1	27.67448	85.31856	S83F	30-Jun-05	13	M			Yes	No	No	Yes	No	Yes
DM093	Acute	C	C	27.676	85.33893	WT	30-Jun-05	21	F			Yes	Yes	Yes	No	No	Yes
DM098	Acute	A	A1.8	27.67525	85.33198	S83F	3-Jul-05	18	M			Yes	No	Yes	No	No	Yes
DM111	Acute	A	A1.9	27.65938	85.32823	S83F	6-Jul-05	21	M			Yes	No	No	No	No	Yes
DM116	Acute	A	A1.7	27.6421	85.32248	S83F	8-Jul-05	32	M			Yes	No	Yes	No	No	Yes
DM118	Acute	A	A1.9	27.69901	85.35241	S83F	8-Jul-05	30	M			Yes	No	No	Yes	No	Yes
DM158	Acute	A	A1.6	27.68921	85.34438	S83F	16-Jul-05	29	M			Yes	No	No	No	No	Yes
DM175	Acute	A	A1.10	27.67495	85.30568	S83F	17-Jul-05	13	F			No	No	No	No	No	Yes
DM178	Acute	A	A1.2	27.66806	85.3289	S83F	19-Jul-05	8	F			No	No	No	No	No	Yes
DM183	Acute	A	A1	27.67451	85.31931	S83F	19-Jul-05	22	F			No	No	Yes	No	No	Yes
DM186	Acute	A	A1.7	27.67746	85.33521	S83F	20-Jul-05	15	M			Yes	No	No	Yes	No	Yes
DM188	Acute	C	C	27.66326	85.32745	D87Y	20-Jul-05	28	M			No	Yes	No	No	No	Yes
DM192	Acute	A	A1.10	27.65668	85.3298	S83F	21-Jul-05	14	F			Yes	No	Yes	No	No	Yes
DM210	Acute	A	A1.10	27.6751	85.30683	S83F	26-Jul-05	20	M			No	No	No	No	No	Yes
DM211	Acute	A	A1.7	27.67746	85.33521	S83F	26-Jul-05	16	M			Yes	No	Yes	No	No	Yes
DM219	Acute	A	A1.2	27.66865	85.32715	S83F	28-Jul-05	10	M			No	No	No	No	Yes	Yes

DM227	Acute	A	A1.2	27.66311	85.31691	S83F	31-Jul-05	14	M			No	No	No	No	No	Yes
DM233	Acute	A	A1.7	27.66948	85.32518	S83F	31-Jul-05	4	F			Yes	Yes	Yes	Yes	No	Yes
DM247	Acute	A	A1.8	27.67205	85.32166	S83F	2-Aug-05	20	M			Yes	No	No	No	No	Yes
DM253	Acute	A	A1	27.66886	85.32515	S83F	4-Aug-05	21	M			Yes	Yes	No	No	No	Yes
DM254	Acute	A	A1.2	27.66913	85.32666	S83F	4-Aug-05	12	F			No	No	No	No	No	Yes
DM256	Acute	A	A1.2	27.67683	85.3219	S83F	4-Aug-05	17	M			No	No	No	No	No	Yes
DM265	Acute	A	A1.2	27.66198	85.32601	S83F	5-Aug-05	14	M			No	No	No	No	No	Yes
DM274	Acute	A	A1.2	27.68061	85.32708	S83F	8-Aug-05	23	M			Yes	No	No	No	No	Yes
DM295	Acute	A	A1.8	27.68211	85.3234	S83F	12-Aug-05	23	M			Yes	No	No	No	No	Yes
DM305	Acute	A	A1.3	27.67885	85.32426	S83F	17-Aug-05	20	M			Yes	Yes	Yes	No	No	Yes
DM313	Acute	C	C	27.67413	85.30495	WT	19-Aug-05	17	M			Yes	No	No	No	No	Yes
DM323	Acute	A	A1.2	27.67336	85.32105	S83F	22-Aug-05	30	M			Yes	Yes	Yes	Yes	No	Yes
DM329	Acute	A	A1.10	27.67471	85.30631	S83F	25-Aug-05	30	M			Yes	No	No	No	No	No
DM356	Acute	A	A1.2	27.67951	85.32698	S83F	30-Aug-05	24	M			No	Yes	Yes	No	No	Yes
DM358	Acute	A	A1.2	27.64466	85.3157	S83F	31-Aug-05	15	M			No	No	Yes	No	No	Yes
DM363	Acute	C	C	27.67413	85.30495	WT	1-Sep-05	16	M			Yes	No	No	No	No	Yes
DM387	Acute	A	A1	27.68	85.3272	S83F	7-Sep-05	19	M			Yes	No	No	No	No	Yes
DM390	Acute	A	A1.8	27.67256	85.3215	S83F	8-Sep-05	25	M			No	No	No	No	No	Yes
ED027	Acute	A	A1.5	27.66598	85.31043	S83F	20-May-06	22	M	Mix	Untreated	Yes	No	No	No	No	Yes
ED034	Acute	A	A1.11	27.6673	85.33213	S83F	22-May-06	21	M	Municipal	Untreated	No	No	No	No	No	Yes
ED035	Acute	A	A1.3	27.67546	85.3212	S83F	22-May-06	10	F	Municipal	Boil	Yes	No	Yes	No	No	Yes
ED037	Acute	A	A1.3	27.6748	85.32711	S83F	23-May-06	12	M	Well	Untreated	No	No	No	No	No	No
ED039	Acute	A	A1.11	27.67965	85.30196	S83F	23-May-06	26	M	Municipal	Boil	Yes	Yes	No	No	No	Yes
ED043	Acute	A	A1.11	27.67363	85.32805	S83F	24-May-06	8	M	Well	Untreated	Yes	Yes	Yes	Yes	No	Yes
ED044	Acute	A	A1.1	27.67761	85.34775	S83F	24-May-06	21	M	Mix	Boil	No	No	Yes	No	No	Yes
ED051	Acute	A	A1.9	27.67986	85.32321	S83F	28-May-06	39	M	Municipal	Other	No	No	No	No	No	Yes
ED053	Acute	A	A1.9	27.67731	85.33545	S83F	28-May-06	19	M	Well	Chlorine	Yes	Yes	Yes	No	Yes	Yes

ED065	Acute	A	A1.4	27.67181	85.30208	S83F	1-Jun-06	17	M	Municipal	Filter	Yes	No	Yes	No	No	Yes
ED068	Acute	A	A1.11	27.66946	85.31273	S83F	4-Jun-06	35	M	Municipal	Boil	No	No	No	No	No	Yes
ED070	Acute	A	A1.11	27.6352	85.31878	S83F	5-Jun-06	7	F	Municipal	Untreated	No	No	Yes	No	No	Yes
ED071	Acute	A	A1.7	27.68526	85.31778	S83F	5-Jun-06	16	M	Municipal	Untreated	Yes	No	No	No	No	Yes
ED074	Acute	C	C	27.68463	85.32151	WT	6-Jun-06	4	F	Stone spout	Untreated	No	No	Yes	No	No	No
ED079	Acute	A	A1.11	27.67256	85.3323	S83F	6-Jun-06	21	M	Stone spout	Untreated	No	No	Yes	No	No	Yes
ED099	Acute	A	A1	27.66573	85.33275	WT	10-Jun-06	28	M	Municipal	Filter	No	No	No	No	No	Yes
ED161	Acute	A	A1.11	27.65751	85.31991	S83F	25-Jun-06	23	F	Well	Mix	No	Yes	No	No	No	Yes
ED164	Acute	A	A1.11	27.67776	85.32176	S83F	26-Jun-06	34	M	Municipal		Yes	Yes	No	No	Yes	Yes
ED168	Acute	A	A1.3	27.67761	85.32058	S83F	27-Jun-06	13	F	Mix	Boil	Yes	No	No	No	No	Yes
ED179	Acute	A	A1.3	27.67891	85.32911	S83F	2-Jul-06	21	F	Stone spout	Boil	Yes	No	No	No	No	Yes
ED182	Acute	A	A1.7	27.67718	85.32768	S83F	3-Jul-06	24	M	Stone spout	Untreated	Yes	No	No	No	No	Yes
ED191	Acute	A	A1.5	27.67318	85.32173	S83F	5-Jul-06	13	M	Well	Filter	Yes	No	No	Yes	No	Yes
ED192	Acute	A	A1.3	27.67595	85.32538	S83F	5-Jul-06	8	M	Mix	Untreated	No	No	Yes	No	No	Yes
ED193	Acute	A	A1.11	27.68211	85.3234	S83F	5-Jul-06	20	M	Mix	Other	Yes	No	No	No	No	Yes
ED194	Acute	A	A1.3	27.67546	85.3212	S83F	5-Jul-06	19	F			Yes	Yes	Yes	No	No	Yes
ED199	Acute	C	C	27.66366	85.32826	WT	6-Jul-06	35	F	Stone spout	Boil	Yes	No	Yes	No	No	Yes
ED260	Acute	A	A1	27.41299	85.187	S83F	23-Jul-06	7	F	Municipal	Chlorine	Yes	No	No	No	No	Yes
ED261	Acute	A	A1.4	27.66781	85.32068	S83F	23-Jul-06	26	M	Tanker	Untreated	No	Yes	No	No	No	Yes
ED262	Acute	A	A1.11	27.66475	85.32506	S83F	24-Jul-06	9	M	Well	Untreated	Yes	No	Yes	Yes	No	Yes
ED267	Acute	A	A1.11	27.68061	85.32708	S83F	25-Jul-06	18	M	Stone spout	Chlorine	No	No	No	No	Yes	Yes
ED272	Acute	A	A1.5	27.673	85.3224	S83F	26-Jul-06	27	F	Well	Filter	Yes	Yes	No	No	No	Yes
ED274	Acute	A	A1.9	27.67776	85.32176	S83F	26-Jul-06	8	F	Municipal	Untreated	Yes	No	No	No	No	No
ED275	Acute	A	A1.11	27.68531	85.31443	S83F	26-Jul-06	17	F	Municipal	Boil	No	Yes	No	No	No	Yes
ED286	Acute	A	A1.5	27.66928	85.32481	S83F	28-Jul-06	26	M	Municipal	Mix	Yes	Yes	Yes	No	Yes	Yes
ED293	Acute	A	A1.3	27.67653	85.32718	S83F	30-Jul-06	24	M	Stone spout	Untreated	No	No	No	No	No	No
ED317	Acute	A	A1.9	27.66355	85.31371	S83F	6-Aug-06	33	F	Mix	Filter	Yes	Yes	Yes	No	No	Yes

ED318	Acute	A	A1.5	27.67208	85.34118	S83F	6-Aug-06	24	M	Stone spout	Untreated	Yes	No	Yes	Yes	No	Yes
ED320	Acute	A	A1.1	27.65438	85.32065	S83F	6-Aug-06	20	M	Mix	Chlorine	Yes	No	No	No	No	Yes
ED321	Acute	A	A1.11	27.67246	85.32088	S83F	7-Aug-06	26	F	Municipal	Mix	No	Yes	No	Yes	No	Yes
ED327	Acute	A	A1.9	27.67821	85.31538	S83F	9-Aug-06	12	M	Municipal	Untreated	Yes	Yes	Yes	No	Yes	Yes
ED330	Acute	A	A1.11	27.66311	85.31691	S83F	10-Aug-06	17	F	Bottled	Other	Yes	Yes	No	No	No	Yes
ED335	Acute	A	A1.5	27.67105	85.32501	S83F	13-Aug-06	24	F	Well	Untreated	Yes	No	No	No	No	Yes
ED338	Acute	A	A1.1	27.67106	85.3047	S83F	15-Aug-06	33	F	Municipal	Filter	Yes	Yes	Yes	No	No	Yes
ED340	Acute	A	A1.5	27.67303	85.32156	S83F	16-Aug-06	11	M	Municipal	Filter	Yes	No	No	No	Yes	Yes
ED345	Acute	A	A1.4	27.68505	85.30958	S83F	17-Aug-06	25	M	Municipal	Mix	No	No	No	No	No	Yes
ED347	Acute	A	A1.5	27.6657	85.32391	S83F	17-Aug-06	16	M	Mix	Untreated	No	No	No	No	No	Yes
ED348	Acute	A	A1.11	27.66415	85.31873	S83F	18-Aug-06	23	F	Well	Other	Yes	Yes	No	No	No	Yes
ED356	Acute	A	A1.11	27.66821	85.3221	S83F	21-Aug-06	25	M	Stone spout	Untreated	Yes	No	No	Yes	No	Yes
ED361	Acute	A	A1.11	27.67396	85.31955	S83F	22-Aug-06	9	F	Municipal	Untreated	Yes	Yes	Yes	No	Yes	Yes
ED362	Acute	A	A1.11	27.66913	85.32336	S83F	23-Aug-06	15	M	Municipal	Filter	Yes	Yes	No	Yes	No	Yes
ED363	Acute	A	A1.1	27.67048	85.31215	S83F	23-Aug-06	25	M	Mix	Untreated	Yes	Yes	Yes	No	No	Yes
ED366	Acute	A	A1.11	27.67031	85.3251	S83F	23-Aug-06	24	F	Well	Boil	No	No	No	No	No	No
ED375	Acute	A	A1.5	27.6799	85.31151	S83F	28-Aug-06	22	M	Municipal	Chlorine	Yes	Yes	Yes	No	No	Yes
ED383	Acute	A	A1.11	27.66998	85.32285	S83F	31-Aug-06	3	M	Well	Untreated	No	No	Yes	Yes	Yes	Yes
ED408	Acute	A	A1.11	27.65583	85.33271	S83F	17-Dec-06	26	M	Stone spout	Boil	Yes	Yes	Yes	No	No	Yes
ED416	Acute	C	C	27.67343	85.3222	WT	4-Jan-07	19	F	Well	Boil	Yes	Yes	No	No	No	Yes
ED417	Acute	A	A1	27.68121	85.32736	S83F	5-Jan-07	19	F	Stone spout	Mix	Yes	No	Yes	No	Yes	Yes
ED432	Acute	A	A1.9	27.67593	85.31933	S83F	2-Mar-07	25	F	Municipal	Mix	Yes	Yes	No	No	No	Yes
ED443	Acute	A	A1	27.68886	85.31315	S83F	16-Mar-07	24	M	Bottled	Untreated	Yes	No	No	Yes	No	Yes
ED446	Acute	A	A2	27.68018	85.32763	D87Y	19-Mar-07	25	M	Stone spout	Boil	No	No	No	No	No	Yes
ED459	Acute	A	A1.1	27.67735	85.32795	S83F	4-Apr-07	23	M	Mix	Untreated	Yes	Yes	No	No	No	Yes
ED494	Acute	A	A1.11	27.66525	85.33801	S83F	6-May-07	22	F	Municipal	Untreated	Yes	Yes	Yes	No	No	Yes
ED503	Acute	A	A1.9	27.68463	85.32151	S83F	16-May-07	23	M	Stone spout	Untreated	Yes	No	No	No	No	No

ED523	Acute	A	A1.11	27.669	85.32908	S83F	1-Jun-07	17	F	Mix	Filter	Yes	No	No	No	No	Yes
ED579	Acute	A	A1.11	27.6729	85.3215	S83F	11-Jul-07	16	M	Well	Filter	Yes	No	No	No	No	Yes
ED624	Acute	A	A1.11	27.68061	85.32708	S83F	10-Aug-07	20	M	Municipal	Chlorine	Yes	Yes	Yes	Yes	No	Yes
ED632	Acute	A	A1	27.67006	85.32481	S83F	20-Aug-07	33	M	Municipal	Filter	Yes	Yes	No	No	No	Yes
ED646	Acute	A	A1.1	27.6724	85.32131	S83F	5-Sep-07	11	M	Well	Boil	Yes	No	No	No	No	Yes
ED805	Acute	A	A1.11	27.66795	85.29868	S83F	2-Jun-08	20	F	Municipal	Untreated	Yes	No	No	No	No	Yes
ED849	Acute	A	A1.4	27.67208	85.30555	S83F	22-Jul-08	15	F	Stone spout	Mix	Yes	No	Yes	No	No	Yes
GB113	Carrier	C	C	27.13393	85.15166	WT	17-Jan-08	45	F								
GB193	Carrier	A	A2	27.699308	85.290304	WT	10-Mar-08	43	F								
GB214	Carrier	F	F	27.67391	85.3218	WT	24-Mar-08	38	F								
GB224	Carrier	C	C	27.67391	85.3218	WT	27-Mar-08	42	F								
GB245	Carrier	A	A1.11	27.13733	85.00836	S83F	11-Apr-08	33	F								
GB280	Carrier	A	A1	27.15806	85.15575	S83F	4-May-08	34	F								
GB493	Carrier	A	A1	27.670389	85.437631	S83F	3-Oct-08										
GB58	Carrier	A	A1.10	27.672926	85.428279	S83F	20-Dec-07	28	M								
GB624	Carrier	A	A1	27.69179	85.342969	WT	9-Jan-09	33	F								
GB625	Carrier	A	A1.3	27.678325	85.316358	S83F	16-Jan-09	57	M								
GB63	Carrier	A	A1.7	27.693811	85.218566	S83F	21-Dec-07	35	F								
GB637	Carrier	C	C	27.693335	85.281691	S83F	23-Jan-09	24	F								
GB640	Carrier	A	A2	27.720193	85.329767	S83F	30-Jan-09	18	M								
GB672	Carrier	A	A2	27.621328	85.302213	N/A	25-Mar-09	31	F								
GB726	Carrier	A	A2	27.67076	85.32031	S83Y	9-Apr-09	44	F								
GB20	Carrier	A	A1	27.665874	85.328512	S83F	12-Jun-09	40	F								
GB75	Carrier	A	A1.7	27.00845	85.13981	S83F	19-Jun-09	49	M								
SPA010	Acute	A	A1.11	27.6779	85.31366	S83F	4-Oct-11	22	F	Municipal	Untreated	Yes	Yes	No	No	No	Yes
SPA014	Acute	A	A1.1	27.67681	85.31805	S83F	4-Nov-11	28	F	Municipal	Mix	Yes	Yes	No	No	No	Yes
SPA015	Acute	A	A1.11	27.66745	85.30191	S83F	13-Apr-11	15	F	Municipal	Mix	No	No	Yes	No	No	Yes

SPA017	Acute	A	A2.1	27.67895	85.32656	S83F	15-Apr-11	19	M	Stone spout	Mix	Yes	No	No	No	No	Yes
SPA018	Acute	A	A1.1	27.6671	85.3055	S83F	20-Apr-11	9	M	Municipal	Untreated	Yes	Yes	Yes	No	No	Yes
SPA019	Acute	A	A1.11	27.66545	85.32481	S83F	28-Apr-11	25	M	Municipal	Filter	No	Yes	No	No	No	Yes
SPA02	Acute	A	A2	27.66033	85.31735	S83F	21-Mar-11	17	M	Municipal	Filter	No	No	Yes	No	No	Yes
SPA020	Acute	A	A2	27.67618	85.34698	S83F	28-Apr-11	18	M	Bottled	Filter	No	No	No	No	No	Yes
SPA021	Acute	A	A2	27.65503	85.3199	S83F	5-May-11	24	F	Bottled	Mix	No	Yes	No	No	No	Yes
SPA023	Acute	A	A1.1	27.69058	85.31155	S83F	5-Mar-11	20	M	Bottled	Filter	Yes	No	No	No	No	Yes
SPA024	Acute	A	A2	27.67995	85.3481	S83F	5-May-11	15	M	Well	Filter	Yes	No	No	No	No	Yes
SPA025	Acute	A	A1.11	27.6573	85.3115	S83F	5-Aug-11	9	F	Bottled	Filter	No	Yes	No	No	No	Yes
SPA026	Acute	A	A2	27.66845	85.34206	S83F	15-May-11	22	M	Bottled	Untreated	No	No	No	No	No	Yes
SPA028	Acute	A	A1.2	27.6661	85.3171	S83F	22-May-11	10	M	Municipal	Filter	No	Yes	Yes	No	Yes	Yes
SPA029	Acute	A	A2	27.67298	85.31961	S83F	20-May-11	18	M	Municipal	Filter	Yes	Yes	Yes	No	No	Yes
SPA030	Acute	A	A1.2	27.66778	85.3242	S83F	22-May-11	20	M	Municipal	Boil	No	Yes	Yes	No	No	Yes
SPA032	Acute	A	A1.11	27.69856	85.31235	S83F	26-May-11	23	F	Municipal	Filter	Yes	No	No	No	No	Yes
SPA033	Acute	A	A1.11	27.67223	85.03216	S83F	24-May-11	20	M	Municipal	Filter	No	No	No	No	No	Yes
SPA034	Acute	A	A2	27.67246	85.322	S83F	26-May-11	22	M	Municipal	Filter	Yes	No	No	No	No	Yes
SPA037	Acute	A	A2	27.67661	85.32435	S83F	30-May-11	8	M	Stone spout	Untreated	No	No	No	No	No	No
SPA038	Acute	A	A2	27.67661	85.32435	S83F	30-May-11	5	M	Stone spout	Untreated	Yes	No	Yes	No	No	Yes
SPA042	Acute	A	A1.2	27.67038	85.32515	S83F	6-Jun-11	16	M	Municipal	Filter	Yes	No	No	No	No	Yes
SPA046	Acute	A	A2	27.67661	85.32435	S83F	14-Jun-11	6	F	Stone spout	Untreated	Yes	No	No	No	No	No
SPA047	Acute	A	A2	27.67758	85.32498	S83F	19-Jun-11	21	M	Municipal	Filter	Yes	Yes	Yes	Yes	No	Yes
SPA048	Acute	A	A1.11	27.67373	85.31835	S83F	20-Jun-11	25	M	Municipal	Filter	No	No	Yes	No	No	Yes
SPA052	Acute	A	A2	27.67781	85.32286	S83F	7-Oct-11	7	M	Municipal	Boil	No	No	No	No	No	No
SPA053	Acute	A	A1.1	27.68108	85.32763	S83F	14-Jul-11	25	M	Mix	Untreated	Yes	Yes	No	No	Yes	Yes
SPA055	Acute	A	A2	27.67088	85.33695	S83F	18-Jul-11	8	M	Municipal	Filter	Yes	No	No	No	No	Yes
SPA056	Acute	A	A2	27.66813	85.30933	S83F	17-Jul-11	8	F	Municipal	Untreated	Yes	Yes	Yes	No	No	Yes
SPA057	Acute	A	A2.1	27.67536	85.34578	S83F	22-Jul-11	55	M	Mix	Filter	Yes	Yes	No	Yes	No	Yes

SPA058	Acute	A	A1.1	27.6792	85.34925	S83F	22-Jul-11	24	F	Mix	Untreated	Yes	Yes	No	No	No	Yes
SPA06	Acute	A	A1.2	27.68125	85.30543	S83F	3-Mar-11	20	F	Well	Filter	Yes	No	No	No	No	Yes
SPA062	Acute	A	A2.1	27.65415	85.32045	S83F	24-Jul-11	10	M	Municipal	Boil	Yes	No	No	Yes	No	No
SPA063	Acute	A	A2	27.66641	85.34001	S83F	24-Jul-11	18	M	Bottled	Filter	Yes	No	No	No	No	Yes
SPA064	Acute	A	A2.1	27.65408	85.32073	S83F	25-Jul-11	25	M	Mix	Untreated	Yes	Yes	No	No	No	Yes
SPA07	Acute	A	A2	27.68031	85.3365	S83F	3-Mar-11	18	M	Bottled	Filter	Yes	No	No	No	No	Yes
SPA072	Acute	A	A2.1	27.65323	85.3214	S83F	8-Feb-11	19	M	Municipal	Untreated	No	No	No	No	No	Yes
SPA076	Acute	A	A2	27.65968	85.31146	S83F	8-Jul-11	6	M	Municipal	Filter	No	No	Yes	Yes	No	Yes
SPA077	Acute	A	A1.5	27.66098	85.30935	S83F	14-Aug-11	7	M	Municipal	Filter	Yes	Yes	Yes	No	No	Yes
SPA078	Acute	A	A2.1	27.65855	85.32275	S83F	14-Aug-11	10	M	Municipal	Filter	Yes	No	No	No	No	Yes
SPA079	Acute	C	C	27.64206	85.33405	S83F	15-Aug-11	33	M	Municipal	Filter	Yes	Yes	Yes	Yes	No	Yes
SPA082	Acute	A	A2	27.66486	85.33193	S83F	17-Aug-11	21	M	Bottled	Filter	No	No	No	No	No	Yes
SPA085	Acute	A	A2.1	27.65855	85.32275	S83F	15-Aug-11	21	F	Municipal	Untreated	Yes	Yes	No	No	No	Yes
SPA086	Acute	A	A2.1	27.65051	85.32133	S83F	25-Aug-11	10	M	Municipal	Mix	Yes	No	No	No	No	Yes
SPA09	Acute	A	A1.2	27.6666	85.3295	S83F	4-Nov-11	9	M	Municipal	Filter	Yes	Yes	No	No	No	Yes
SPA096	Acute	A	A2.1	27.65415	85.31986	S83F	9-Sep-11	12	F	Municipal	Mix	Yes	Yes	Yes	No	No	Yes
SPA098	Acute	A	A2.1	27.65886	85.3134	S83F	13-Sep-11	25	F	Municipal	Filter	Yes	Yes	Yes	No	No	Yes
SPA100	Acute	A	A2	27.67816	85.32581	S83F	15-Sep-11	25	F	Municipal	Untreated	No	Yes	Yes	No	No	Yes
SPA102	Acute	A	A2	27.68808	85.31613	S83F	23-Sep-11	11	F	Municipal	Mix	Yes	Yes	No	No	No	Yes
TY018	Acute	A	A2	27.6762	85.32461	S83F	6-Aug-08	18	M	Stone spout	Untreated	No	Yes	No	No	No	Yes
TY035	Acute	A	A1.3	27.66776	85.3051	S83F	4-Sep-08	17	M	Stone spout	Untreated	Yes	Yes	Yes	No	Yes	Yes
TY037	Acute	A	A1.11	27.66083	85.32318	S83F	7-Sep-08	5	M	Municipal	Filter	Yes	No	Yes	No	Yes	Yes
TY051	Acute	A	A1.11	27.67546	85.29706	S83F	31-Oct-08	14	F	Municipal	Filter	Yes	Yes	No	No	No	Yes
02TY004	Acute	A	A2	27.68808	85.31613	S83F	23/9/2011	11	F	Municipal	Boil	Yes	Yes	Yes	No	No	Yes
02TY013	Acute	A	A2.1	27.66751	85.32836	S83F	25/12/2011	22	M	Stone sprout	Filter	No	No	No	No	No	Yes
02TY036	Acute	A	A2.1	27.66438	85.30686	S83F	22/4/2012	30	M	Municipal	Untreated	Yes	Yes	No	Yes	No	Yes
02TY041	Acute	A	A2	27.68046	85.33966	S83F	29/4/2012	23	F	Municipal	Filter	Yes	Yes	No	Yes	No	Yes

02TY044	Acute	A	A2	27.68915	85.32486	S83F	8/5/12	4	M	Municipal	Filter	Yes	No	Yes	No	No	No
02TY046	Acute	A	A2	27.68053	85.33798	S83F	9/5/12	32	M	Municipal	Boil	No	No	No	No	Yes	Yes
02TY049	Acute	A	A2.1	27.67411	85.32641	S83F	15/5/2012	16	M	Bottled	Untreated	YesYes	No	YesYes	No	No	Yes
02TY052	Acute	A	A2.1	27.66096	85.31985	S83F	17/5/2012	21	M	Municipal	Filter	Yes	Yes	No	Yes	No	Yes
02TY053	Acute	A	A2	27.66735	85.31911	S83F	22/5/2012	12	M	Municipal	Boil	Yes	No	No	No	No	No
02TY062	Acute	A	A2	27.68858	85.32641	S83F	10/6/12	22	M	Bottled	Untreated	No	Yes	No	No	No	Yes
02TY063	Acute	A	A2	27.68858	85.32641	S83F	10/6/12	19	M	Bottled	Untreated	Yes	Yes	No	No	No	No
02TY067	Acute	A	A1.9	27.68263	85.3381	S83F	14/6/2012	24	M	Bottled	Boil	No	No	Yes	No	Yes	Yes
02TY068	Acute	A	A2	27.693	85.34703	S83F	15/6/2012	24	M	Bottled	Untreated	Yes	No	No	No	No	Yes
02TY072	Acute	A	A2.1	27.67253	85.3352	S83F	18/6/2012	21	M	Bottled	Untreated	No	No	No	No	No	Yes
02TY074	Acute	A	A2	27.68193	85.32373	S83F	18/6/2012	22	M	Municipal	Filter	Yes	No	No	No	No	Yes
02TY075	Acute	A	A2	27.67121	85.3166	S83F	11/7/12	24	M	Municipal	Boil	Yes	Yes	No	No	Yes	Yes
02TY076	Acute	A	A2.1	27.66033	85.33945	S83F	12/7/12	12	F	Municipal	Filter	Yes	No	No	No	No	No
02TY077	Acute	A	A2	27.6673	85.34081	S83F	13/7/2012	12	F	Municipal	Filter	Yes	No	No	No	No	Yes
02TY081	Acute	A	A2	27.67988	85.33608	S83F	22/7/2012	22	M	Well	Boil	No	Yes	No	No	No	Yes
02TY084	Acute	A	A2	27.68456	85.34125	S83F	30/7/2012	15	F	Municipal	Filter	Yes	No	Yes	No	No	Yes
02TY085	Acute	A	A2	27.67305	85.33271	S83F	2/8/12	13	M	Tanker	Untreated	Yes	Yes	Yes	No	No	Yes
02TY086	Acute	A	A2	27.6698	85.3161	S83F	5/8/12	10	M	Municipal	Untreated	Yes	No	No	No	No	Yes
02TY091	Acute	A	A2	27.67656	85.32735	S83F	19/8/2012	27	M	Stone sprout	Filter	Yes	No	No	No	No	Yes
02TY094	Acute	A	A2	27.68203	85.34038	S83F	21/8/2012	19	M	Bottled	Untreated	Yes		No	No	Yes	Yes
02TY100	Acute	A	A2	27.68236	85.31561	S83F	5/9/12	9	M	Tanker	Filter	Yes	Yes	Yes	No	Yes	Yes
02TY101	Acute	A	A2.1	27.67696	85.32135	S83F	1/11/12	14	M	Municipal	Filter	No	No	No	No	No	Yes
02TY109	Acute	A	A2	27.66198	85.30661	S83F	8/9/12	10	M	Municipal	Filter	Yes	Yes	No	No	Yes	Yes
02TY129	Acute	A	A2.1	27.65771	85.31621	S83F	16/04/2013	6	M	Municipal	Untreated	Yes	No	No	No	No	Yes
02TY152	Acute	A	A2	27.69551	85.33811	S83F	4/6/13	35	M	Municipal	Boil	No	No	No	No	Yes	Yes
02TY153	Acute	A	A2	27.69343	85.34256	S83F	4/6/13	20	M	Municipal	Untreated	Yes	No	No	No	No	Yes
02TY155	Acute	A	A2	27.69343	85.34256	S83F	16/06/2013	18	M	Municipal	Untreated	Yes	Yes	Yes	Yes	No	Yes

02TY180	Acute	C	C	27.66913	85.32658	S83F	14/08/2013	10	F	Municipal	Filter	No	No	No	Yes	No	Yes
02TY187	Acute	A	A2.1	27.68955	85.33551	S83F	28/88/2013	25	M	Bottled	Untreated	Yes	No	No	No	No	Yes
02TY212	Acute	A	A2	27.68063	85.31225	S83F	12/1/14	34	M	Municipal	Untreated	No	No	No	No	No	Yes
02TY215	Acute	A	A2.1	27.66568	85.32441	S83F	29/01/2014	32	M	Municipal	Filter	No	Yes	No	No	Yes	Yes
02TY221	Acute	A	A2	27.67211	85.31981	S83F	31/03/2014	24	M	Municipal	Filter	Yes	Yes	No	No	No	Yes
02TY223	Acute	A	A2	27.69096	85.33345	S83F	4/3/14	17	M	Municipal	Filter	No	No	No	Yes	No	Yes
02TY224	Acute	A	A1.1	27.69075	85.3488	S83F	17/03/2014	27	M	Stone sprout	Filter	No	No	No	No	No	Yes
02TY231	Acute	A	A2	27.66741	85.35096	S83F	1/5/14	18	M	Bottled	Untreated	Yes	No	Yes	No	Yes	Yes
02TY242	Acute	A	A2.1	27.68543	85.3415	S83F	16/06/2014	15	F	Municipal	Filter	No	No	No	No	No	Yes

Appendix C *Salmonella* Typhi isolates and their corresponding metadata in chapter 5

ID	Age_ years	Sex	Year	Province	District	Commune	Longtitude	Latitude	Lineages	Sublineages	IncH1 plasmid	AMP	CHL	SXT	Na	AZI_MIC	CIP_MIC	Acc No
1	9.3	F	2007	Siem Reap	Siem Reab	Sala Kamraeuk	103.867133	13.345557	I	I	0	S	S	S	I	8	0.094	ERR340759
2	13.6	F	2010	Siem Reap	Angkor Chum	Nokor Pheas	103.719479	13.624206	II	II	1	R	R	R	S	12	0.016	ERR360845
3	14.8	M	2007	Siem Reap	Angkor Chum	Ta Saom	103.677526	13.671199	II	II	1	R	R	R	S	8	0.012	ERR340758
4	12.0	F	2007	Siem Reap	Siem Reab	Sala Kamraeuk	103.867133	13.345557	II	II	0	S	S	S	R	16	0.19	ERR340774
5	7.1	F	2007	Banteaymea nchey	O Chrauv	Poipet	102.659645	13.642996	III	IIIa	1	R	R	R	R	8	0.19	ERR340760
6	1.5	F	2011	Battambang	Samlaut	Chamlorng Kuoy	102.859385	12.583148	III	IIIa	1	R	R	R	0	0	0	ERR360770
7	14.7	F	2012	Kampong Cham	Cheung Prey	Khnaor Dambang	105.014179	12.018816	III	IIIa	1	R	R	R	0	0	0	ERR319473
8	3.8	F	2011	Kampong Chhnang	Kampong Chhnan	Khsam	104.660713	12.279773	III	IIIa	1	R	R	R	0	12	0.25	ERR319445
9	12.7	M	2011	Kampong Thom	Stoeng Sen	Ou Kanthor	104.816132	12.667063	III	IIIa	1	R	R	R	0	0	0	ERR360773
10	13.7	M	2010	Kampong Thom	Stoeng Sen	Srayov	104.858972	12.59656	III	IIIa	1	R	R	R	R	12	0.25	ERR360952
11	4.2	F	2009	Siem Reap	Siem Reab	Siem Reab	103.843962	13.277306	III	IIIa	0	S	S	S	R	12	0.25	ERR319416
12	13.2	F	2011	Siem Reap	Siem Reab	Sla Kram	103.885703	13.367605	III	IIIa	0	S	S	S	S	8	0.008	ERR319459
13	#N/A	#N/A	2010	Banteaymea nchey	Preah Netr Preah	Prasat	103.365501	13.457949	III	IIIb	1	R	R	R	R	12	0.38	ERR360754
14	5.9	F	2010	Banteaymea nchey	Preah Netr Preah	Rohal	103.237876	13.593116	III	IIIb	1	R	R	R	R	16	0.38	ERR319425
15	9.8	F	2010	Banteaymea nchey	Preah Netr Preah	Tuek Chour	103.369654	13.620483	III	IIIb	1	R	R	R	R	12	0.25	ERR360956
16	11.7	F	2010	Siem Reap	Angkor Chum	Kouk Doung	103.792493	13.701174	III	IIIb	1	R	R	R	R	8	0.25	ERR319439
17	8.1	M	2010	Siem Reap	Banteay Srei	Tbaeng	104.038638	13.552495	III	IIIb	1	R	R	R	R	12	0.25	ERR360846
18	14.1	F	2010	Siem Reap	Banteay Srei	Tbaeng	104.038638	13.552495	III	IIIb	1	R	R	R	R	12	0.25	ERR319434
19	4.6	M	2010	Siem Reap	Kralanh	Krouch Kor	103.587557	13.703525	III	IIIb	1	R	R	R	R	8	0.25	ERR319437
20	13.3	F	2010	Siem Reap	Kralanh	Krouch Kor	103.587557	13.703525	III	IIIb	1	R	R	R	R	12	0.25	ERR319441
21	12.2	F	2010	Siem Reap	Prasat Bakong	Bakong	103.991452	13.355008	III	IIIb	1	R	R	R	R	12	0.25	ERR319429

22	1.6	F	2012	Siem Reap	Siem Reab	Kouk Chak	103.841705	13.429301	III	IIIb	0	S	S	S	R	12	0.5	ERR360792
23	1.3	F	2010	Siem Reap	Siem Reab	Kouk Chak	103.841705	13.429301	III	IIIb	0	S	S	S	R	0	0.5	ERR319423
24	12.5	F	2010	Siem Reap	Siem Reab	Nokor Thum	103.889397	13.423837	III	IIIb	1	R	R	R	R	8	0.25	ERR319422
25	10.2	M	2010	Siem Reap	Siem Reab	Siem Reab	103.843962	13.277306	III	IIIb	1	R	R	R	R	16	0.25	ERR319427
26	7.8	M	2010	Siem Reap	Siem Reab	Siem Reab	103.843962	13.277306	III	IIIb	0	S	S	S	R	12	0.38	ERR319432
27	7.4	M	2010	Siem Reap	Svay Leu	Boeng Mealea	104.280929	13.45211	III	IIIb	0	R	R	R	R	12	0.19	ERR319438
28	10.7	M	2011	Banteaymeanchey	Mongkol Borei	Koy Maeng	103.101314	13.550598	III	IIIc	1	R	R	R	0	0	0	ERR319463
29	11.8	M	2011	Banteaymeanchey	Preah Netr Preah	Chob Veari	103.167669	13.66311	III	IIIc	1	R	R	R	0	0	0	ERR319454
30	7.9	F	2011	Banteaymeanchey	Preah Netr Preah	Phnum Lieb	103.297435	13.520053	III	IIIc	1	R	R	R	0	0	0	ERR319466
31	1.9	M	2011	Banteaymeanchey	Preah Netr Preah	Phnum Lieb	103.297435	13.520053	III	IIIc	1	R	R	R	0	0	0	ERR319470
32	7.1	M	2011	Battambang	Phnom Preuk	Pech Chenda	102.538466	13.268541	III	IIIc	1	R	R	R	0	0	0	ERR360758
33	12.3	F	2011	Siem Reap	Angkor Chum	Char Chhuk	103.623954	13.663726	III	IIIc	0	S	S	S	S	6	0.006	ERR319457
34	5.1	F	2012	Siem Reap	Chi Kreng	Anlong Samnar	104.229051	12.94545	III	IIIc	1	R	R	R	R	12	0.38	ERR361000
35	9.1	F	2011	Siem Reap	Kralanh	Chonloas Dai	103.46967	13.641287	III	IIIc	1	R	R	R	0	0	0	ERR319458
36	16.2	M	2011	Siem Reap	Kralanh	Kralanh	103.432035	13.55668	III	IIIc	1	R	R	R	0	0	0	ERR319452
37	7.7	M	2011	Siem Reap	Puok	Doun Kaev	103.774015	13.478241	III	IIIc	1	R	R	R	0	0	0	ERR360766
38	6.9	M	2011	Siem Reap	Puok	Kaev Poar	103.716096	13.321008	III	IIIc	1	R	R	R	R	12	0.38	ERR319460
39	14.2	M	2011	Siem Reap	Puok	Prey Chruk	103.542856	13.503103	III	IIIc	1	R	R	R	0	0	0	ERR319471
40	5.6	M	2011	Siem Reap	Puok	Trei Nhoar	103.706054	13.574489	III	IIIc	1	R	R	R	0	0	0	ERR360962
41	11.2	F	2011	Siem Reap	Puok	Tuek Vil	103.807153	13.413196	III	IIIc	1	R	R	R	0	12	0.38	ERR319446
42	5.5	M	2011	Siem Reap	Puok	Yeang	103.624304	13.563253	III	IIIc	1	R	R	R	0	12	0.38	ERR319448
43	7.0	F	2012	Siem Reap	Siem Reab	Chong Khnies	103.829966	13.175937	III	IIIc	1	R	R	R	R	6	0.38	ERR360988
44	8.7	F	2011	Siem Reap	Siem Reab	Sla Kram	103.885703	13.367605	III	IIIc	1	R	R	R	R	12	0.5	ERR319447
45	11.1	F	2009	Siem Reap	Siem Reab	Svay Dangkum	103.822374	13.311014	III	IIIc	0	S	S	S	R	12	0.38	ERR360746
46	5.7	M	2011	Siem Reap	Sotr Nikom	Popel	104.163519	13.39285	III	IIIc	1	R	R	R	0	0	0	ERR319455
47	5.1	F	2008	Banteaymeanchey	O Chrauv	Nimitt	102.659645	13.642996	IV	IVa	0	S	S	S	R	16	0.5	ERR360741

48	14.1	F	2008	Oddar Meanchey	Anlong Veng	Anlong Veng	104.153127	14.240485	IV	IVa	0	S	S	S	R	12	0.38	ERR319410
49	12.2	F	2008	Siem Reap	Siem Reab	Kouk Chak	103.841705	13.429301	IV	IVa	0	S	S	S	R	12	0.25	ERR340767
50	9.1	F	2008	Siem Reap	Siem Reab	Kouk Chak	103.841705	13.429301	IV	IVa	0	S	S	S	S	8	0.008	ERR319411
51	5.7	M	2008	Siem Reap	Siem Reab	Svay Dangkum	103.822374	13.311014	IV	IVa	1	R	R	R	R	12	0.19	ERR360890
52	8.5	M	2008	Siem Reap	Siem Reab	Svay Dangkum	103.822374	13.311014	IV	IVa	0	S	R	S	R	16	0.5	ERR319406
53	10.4	M	2010	Banteaymea nchey	O Chrauv	Changha	102.892727	13.647435	IV	IV	1	R	R	R	R	12	0.38	ERR360955
54	13.2	M	2008	Banteaymea nchey	Preah Netr Preah	Prasat	103.365501	13.457949	IV	IV	1	R	R	R	R	12	0.38	ERR360742
55	1.8	M	2008	Banteaymea nchey	Preah Netr Preah	Prasat	103.365501	13.457949	IV	IV	1	R	R	R	R	12	0.38	ERR360889
56	5.5	F	2008	Battambang	ek Phnum	Kaoh Chiveang	103.610762	13.21135	IV	IV	1	R	R	R	R	16	0.25	ERR360743
57	11.7	F	2009	Kampong Cham	Chamkar Leu	Bos Khnaor	105.335766	12.170203	IV	IV	1	R	R	R	R	16	0.5	ERR319415
58	2.0	M	2007	Kampong Thom	Stoung	Banteay Stoung	104.6337	12.943935	IV	IV	1	R	R	R	R	12	0.5	ERR340757
59	12.2	M	2008	Kampong Thom	Stoeng Sen	Achar Leak	104.903907	12.727871	IV	IV	1	R	R	R	R	12	0.5	ERR360950
60	5.0	F	2007	Pursat	Bakan	Me Tuek	103.782131	12.803292	IV	IV	1	R	R	R	R	16	0.38	ERR340770
61	4.5	M	2012	Pursat	Bakan	Me Tuek	103.782131	12.803292	IV	IV	1	R	R	R	R	0	0.38	ERR360999
62	12.8	M	2008	Siem Reap	Angkor Chum	Srae Khvav	103.672335	13.793712	IV	IV	1	R	R	R	R	12	0.5	ERR360739
63	6.0	F	2007	Siem Reap	Chi Kreng	Anlong Samnar	104.229051	12.94545	IV	IV	1	R	R	R	R	12	0.38	ERR340766
64	13.0	F	2007	Siem Reap	Chi Kreng	Anlong Samnar	104.229051	12.94545	IV	IV	1	R	R	R	R	12	0.38	ERR340761
65	11.9	M	2009	Siem Reap	Chi Kreng	Anlong Samnar	104.229051	12.94545	IV	IV	1	R	R	R	R	12	0.5	ERR319419
66	14.8	M	2010	Siem Reap	Chi Kreng	Chi Kraeng	104.254404	13.074933	IV	IV	0	S	S	S	0	12	0.5	ERR319431
67	6.0	F	2007	Siem Reap	Kralanh	Sambuor	103.445813	13.433224	IV	IV	1	R	R	R	R	12	0.38	ERR340772
68	4.9	M	2008	Siem Reap	Kralanh	Sambuor	103.445813	13.433224	IV	IV	1	R	R	R	R	12	0.38	ERR360738
69	2.4	M	2008	Siem Reap	Kralanh	Sambuor	103.445813	13.433224	IV	IV	1	R	R	R	R	12	0.38	ERR360740
70	8.0	M	2008	Siem Reap	Kralanh	Sambuor	103.445813	13.433224	IV	IV	1	R	R	R	R	16	0.38	ERR319405
71	10.0	F	2008	Siem Reap	Kralanh	Sambuor	103.445813	13.433224	IV	IV	1	R	R	R	R	16	0.5	ERR319409
72	4.3	M	2009	Siem Reap	Siem Reab	Chong Khnies	103.829966	13.175937	IV	IV	1	R	R	R	R	0	0.25	ERR360744
73	5.1	F	2008	Siem Reap	Siem Reab	Sambuor	103.7935	13.30069	IV	IV	1	R	R	R	R	12	0.38	ERR319403

74	15.2	F	2010	Siem Reap	Siem Reab	Srangae	103.776959	13.303244	IV	IV	1	R	R	R	R	8	0.38	ERR319428
75	12.0	M	2007	Siem Reap	Sotr Nikom	Dan Run	104.087793	13.182907	IV	IV	1	R	R	R	R	12	0.38	ERR340768
76	15.1	F	2008	Siem Reap	Sotr Nikom	Dan Run	104.087793	13.182907	IV	IV	1	R	R	R	R	12	0.38	ERR360948
77	2.4	F	2007	Siem Reap	Sotr Nikom	Kampong Khleang	104.142061	13.046753	IV	IV	1	R	R	R	R	16	0.38	ERR340771
78	4.5	F	2007	Siem Reap	Sotr Nikom	Kampong Khleang	104.142061	13.046753	IV	IV	1	R	R	R	R	12	0.38	ERR340773
79	#N/A	#N/A	2007	Siem Reap	Sotr Nikom	Kampong Khleang	104.142061	13.046753	IV	IV	1	R	R	R	R	12	0.5	ERR340775
80	0.6	M	2007	Siem Reap	Sotr Nikom	Kampong Khleang	104.142061	13.046753	IV	IV	1	R	R	R	R	12	0.38	ERR340763
81	3.2	F	2008	Siem Reap	Sotr Nikom	Kampong Khleang	104.142061	13.046753	IV	IV	1	R	R	R	R	16	0.38	ERR360891
82	3.2	M	2008	Siem Reap	Sotr Nikom	Kampong Khleang	104.142061	13.046753	IV	IV	1	R	R	R	R	12	0.5	ERR319402
83	6.1	M	2008	Banteaymeanchey	Preah Netr Preah	Prasat	103.365501	13.457949	IV	IV	1	R	R	R	R	16	0.38	ERR319404
84	5.2	F	2010	Banteaymeanchey	Preah Netr Preah	Prasat	103.365501	13.457949	IV	IV	1	NA	NA	NA	NA	NA	NA	ERR319433
85	4.1	M	2009	Siem Reap	Kralanh	Sambuor	103.445813	13.433224	IV	IV	1	R	R	R	R	16	0.38	ERR319412
86	6.6	F	2009	Siem Reap	Kralanh	Sambuor	103.445813	13.433224	IV	IV	1	R	R	R	R	16	0.38	ERR319413
87	14.7	M	2010	Banteaymeanchey	Preah Netr Preah	Tuek Chour	103.369654	13.620483	IV	IVb	1	R	R	R	R	16	0.5	ERR319426
88	3.3	F	2010	Battambang	ek Phnum	Kaoh Chiveang	103.610762	13.21135	IV	IVb	1	R	R	R	R	8	0.38	ERR360748
89	2.6	F	2011	Battambang	ek Phnum	Kaoh Chiveang	103.610762	13.21135	IV	IVb	1	R	R	R	0	0	0	ERR360761
90	5.6	F	2011	Battambang	ek Phnum	Kaoh Chiveang	103.610762	13.21135	IV	IVb	1	R	R	R	0	0	0	ERR360772
91	6.8	F	2008	Battambang	ek Phnum	Kaoh Chiveang	103.610762	13.21135	IV	IVb	1	R	R	R	R	16	0.38	ERR360949
92	1.9	F	2009	Battambang	ek Phnum	Kaoh Chiveang	103.610762	13.21135	IV	IVb	0	S	S	S	R	16	0.38	ERR319420
93	1.2	M	2010	Battambang	ek Phnum	Kaoh Chiveang	103.610762	13.21135	IV	IVb	1	R	R	R	R	8	0.38	ERR319436
94	4.1	M	2011	Battambang	ek Phnum	Kaoh Chiveang	103.610762	13.21135	IV	IVb	1	R	R	R	0	0	0	ERR319469
95	5.7	M	2008	Battambang	ek Phnum	Kaoh Chiveang	103.610762	13.21135	IV	IVb	1	R	R	R	R	12	0.5	ERR319408
96	#N/A	#N/A	2012	Battambang	ek Phnum	Kaoh Chiveang	103.610762	13.21135	IV	IVb	1	NA	NA	NA	NA	NA	NA	ERR319476
97	11.6	F	2012	Battambang	ek Phnum	Kaoh Chiveang	103.610762	13.21135	IV	IVb	1	R	R	R	R	0	0	ERR319484
98	5.9	M	2011	Battambang	ek Phnum	Prey Chas	103.39292	13.304805	IV	IVb	0	R	R	R	0	0	0	ERR360771
99	12.1	M	2012	Preah Vihear	Kulen	Srayang	104.511339	13.76033	IV	IVb	1	R	R	R	0	0	0	ERR319475

100	6.6	F	2012	Siem Reap	Angkor Chum	Nokor Pheas	103.719479	13.624206	IV	IVb	0	S	S	S	0	0	0	ERR360964
101	12.7	M	2012	Siem Reap	Chi Kreng	Chi Kraeng	104.254404	13.074933	IV	IVb	1	R	R	R	R	0	0.38	ERR360993
102	5.3	M	2009	Siem Reap	Kralanh	Sambuor	103.445813	13.433224	IV	IVb	1	R	R	R	R	12	0.38	ERR319417
103	4.1	M	2011	Siem Reap	Prasat Bakong	Kampong Phluk	103.986709	13.235758	IV	IVb	1	R	R	R	0	0	0	ERR360769
104	4.6	M	2011	Siem Reap	Prasat Bakong	Kampong Phluk	103.986709	13.235758	IV	IVb	1	R	R	R	0	0	0	ERR319467
105	5.2	M	2009	Siem Reap	Puok	Kaev Poar	103.716096	13.321008	IV	IVb	1	R	R	R	R	8	0.38	ERR319421
106	14.2	F	2010	Siem Reap	Puok	Kaev Poar	103.716096	13.321008	IV	IVb	1	R	R	R	0	0	0	ERR319440
107	12.2	M	2010	Siem Reap	Puok	Lvea	103.659697	13.416587	IV	IVb	1	R	R	R	R	12	0.38	ERR319424
108	13.0	M	2012	Siem Reap	Puok	Pou Treay	103.53096	13.374667	IV	IVb	1	R	R	R	R	0	0.38	ERR360997
109	8.1	M	2011	Banteaymeanchey	Serei Saophonj	Mkak	102.965143	13.674269	IV	IVc	1	R	R	R	0	0	0	ERR360760
110	8.5	M	2012	Banteaymeanchey	Serei Saophonj	Preah Ponlea	102.990584	13.579188	IV	IVc	1	R	R	R	R	0	0.38	ERR360787
111	10.3	M	2012	Kampong Thom	Kampong Svay	Kampong Svay	104.980724	12.761543	IV	IVc	1	R	R	R	0	0	0	ERR319472
112	4.0	M	2012	Oddar Meanchey	Chong Kal	Pongro	103.611356	14.057113	IV	IVc	1	R	R	R	R	0	0.38	ERR360782
113	7.9	M	2012	Oddar Meanchey	Samraong	Osmach	103.662372	14.270904	IV	IVc	1	R	R	R	R	0	0.38	ERR360972
114	8.3	F	2012	Oddar Meanchey	Samraong	Osmach	103.662372	14.270904	IV	IVc	0	R	R	R	R	0	0.38	ERR360987
115	10.1	M	2012	Oddar Meanchey	Samraong	Osmach	103.662372	14.270904	IV	IVc	1	R	R	R	R	0	0.25	ERR319495
116	8.5	F	2011	Oddar Meanchey	Samraong	Samraong	103.543846	14.154499	IV	IVc	1	R	R	R	0	0	0	ERR360765
117	5.3	M	2011	Preah Vihear	Choam Ksant	Kantuot	104.6417	14.280355	IV	IVc	1	R	R	R	0	0	0	ERR360756
118	12.2	F	2012	Preah Vihear	Kulen	Srayang	104.511339	13.76033	IV	IVc	1	R	R	R	R	0	0.25	ERR360986
119	9.9	F	2012	Siem Reap	Angkor Chum	Char Chhuk	103.623954	13.663726	IV	IVc	1	R	R	R	0	0	0	ERR360779
120	13.0	F	2012	Siem Reap	Angkor Chum	Char Chhuk	103.623954	13.663726	IV	IVc	1	R	R	R	R	0	0.38	ERR360981
121	9.6	F	2011	Siem Reap	Angkor Chum	Char Chhuk	103.623954	13.663726	IV	IVc	1	R	R	R	0	0	0	ERR319465
122	9.1	F	2011	Siem Reap	Angkor Chum	Kouk Doung	103.792493	13.701174	IV	IVc	1	R	R	R	0	0	0	ERR319464
123	10.3	M	2011	Siem Reap	Angkor Chum	Nokor Pheas	103.719479	13.624206	IV	IVc	1	R	R	R	R	0	0.38	ERR360762
124	6.8	F	2012	Siem Reap	Angkor Chum	Ta Saom	103.677526	13.671199	IV	IVc	1	R	R	R	R	0	0	ERR360973
125	1.5	M	2012	Siem Reap	Angkor Chum	Ta Saom	103.677526	13.671199	IV	IVc	1	R	R	R	R	0	0	ERR360977

126	6.7	M	2012	Siem Reap	Angkor Chum	Ta Saom	103.677526	13.671199	IV	IVc	1	R	R	R	R	0	0.38	ERR319486
127	12.3	M	2012	Siem Reap	Angkor Chum	Ta Saom	103.677526	13.671199	IV	IVc	1	R	R	R	R	0	0.38	ERR319487
128	12.3	M	2012	Siem Reap	Angkor Thum	Leang Dai	103.876266	13.502826	IV	IVc	1	R	R	R	R	0	0.25	ERR360775
129	12.9	F	2012	Siem Reap	Angkor Thum	Leang Dai	103.876266	13.502826	IV	IVc	1	R	R	R	R	0	0.38	ERR360979
130	1.8	M	2010	Siem Reap	Angkor Thum	Peak Snaeng	103.885243	13.610438	IV	IVc	1	R	R	R	R	8	0.25	ERR319435
131	4.7	M	2012	Siem Reap	Banteay Srei	Khnar Sanday	103.952295	13.569677	IV	IVc	1	R	R	R	R	0	0.38	ERR360966
132	3.9	M	2012	Siem Reap	Banteay Srei	Khun Ream	104.032137	13.668959	IV	IVc	1	R	R	R	R	0	0.25	ERR360778
133	9.3	F	2012	Siem Reap	Banteay Srei	Khun Ream	104.032137	13.668959	IV	IVc	1	R	R	R	R	0	0	ERR360982
134	11.3	M	2012	Siem Reap	Banteay Srei	Khun Ream	104.032137	13.668959	IV	IVc	1	R	R	R	R	0	0.25	ERR319494
135	2.0	F	2012	Siem Reap	Banteay Srei	Rumchek	103.981952	13.508645	IV	IVc	1	R	R	R	R	0	0.38	ERR360783
136	13.1	M	2011	Siem Reap	Banteay Srei	Run Ta Aek	104.039073	13.465833	IV	IVc	1	R	R	R	0	12	0.38	ERR319444
137	13.0	F	2012	Siem Reap	Banteay Srei	Run Ta Aek	104.039073	13.465833	IV	IVc	1	R	R	R	R	0	0	ERR319479
138	10.0	F	2012	Siem Reap	Banteay Srei	Tbaeng	104.038638	13.552495	IV	IVc	1	R	R	R	R	0	0.38	ERR360992
139	3.8	F	2010	Siem Reap	Chi Kreng	Anlong Samnar	104.229051	12.94545	IV	IVc	1	R	R	R	R	0	0.25	ERR360959
140	15.4	M	2012	Siem Reap	Chi Kreng	Ruessei Lok	104.227712	13.251327	IV	IVc	1	R	R	R	R	0	0	ERR319482
141	14.2	F	2012	Siem Reap	Kralanh	Chonloas Dai	103.46967	13.641287	IV	IVc	1	R	R	R	R	0	0.38	ERR360965
142	#N/A	#N/A	2009	Siem Reap	Prasat Bakong	Kandaek	103.931669	13.299934	IV	IVc	1	R	R	R	R	0	0.38	ERR360745
143	7.8	F	2012	Siem Reap	Prasat Bakong	Kandaek	103.931669	13.299934	IV	IVc	1	R	R	R	R	0	0	ERR360781
144	6.6	F	2012	Siem Reap	Prasat Bakong	Kandaek	103.931669	13.299934	IV	IVc	1	NA	NA	NA	NA	NA	NA	ERR360968
145	7.0	F	2012	Siem Reap	Prasat Bakong	Kandaek	103.931669	13.299934	IV	IVc	1	R	R	R	R	0	0.19	ERR319485
146	13.4	M	2012	Siem Reap	Prasat Bakong	Kantreang	104.03968	13.347416	IV	IVc	1	R	R	R	R	0	0.25	ERR360990
147	9.0	F	2011	Siem Reap	Prasat Bakong	Roluos	103.98127	13.310897	IV	IVc	1	R	R	R	0	0	0	ERR360759
148	5.8	F	2011	Siem Reap	Puok	Reul	103.73586	13.536268	IV	IVc	1	R	R	R	0	0	0	ERR319453
149	3.0	F	2012	Siem Reap	Puok	Samraong Yea	103.69054	13.361494	IV	IVc	1	R	R	R	R	0	0.38	ERR360984
150	11.8	F	2009	Siem Reap	Puok	Sasar Sdam	103.630882	13.509808	IV	IVc	1	R	R	R	R	12	0.38	ERR319414
151	10.9	M	2012	Siem Reap	Puok	Trei Nhoar	103.706054	13.574489	IV	IVc	1	R	R	R	R	0	0.25	ERR331208

152	10.9	F	2012	Siem Reap	Puok	Trei Nhoar	103.706054	13.574489	IV	IVc	1	R	R	R	R	0	0.25	ERR360788
153	5.9	F	2012	Siem Reap	Puok	Trei Nhoar	103.706054	13.574489	IV	IVc	1	R	R	R	R	0	0.38	ERR360776
154	11.5	M	2012	Siem Reap	Puok	Trei Nhoar	103.706054	13.574489	IV	IVc	1	R	R	R	R	0	0.38	ERR360998
155	9.3	F	2012	Siem Reap	Siem Reab	Chreav	103.888751	13.283577	IV	IVc	1	R	R	R	R	0	0.38	ERR360777
156	8.2	F	2011	Siem Reap	Siem Reab	Kouk Chak	103.841705	13.429301	IV	IVc	1	R	R	R	0	0	0	ERR360763
157	1.6	M	2012	Siem Reap	Siem Reab	Kouk Chak	103.841705	13.429301	IV	IVc	1	R	R	R	R	0	0	ERR360971
158	6.7	M	2012	Siem Reap	Siem Reab	Kouk Chak	103.841705	13.429301	IV	IVc	1	R	R	R	R	0	0.38	ERR319483
159	9.0	M	2012	Siem Reap	Siem Reab	Nokor Thum	103.889397	13.423837	IV	IVc	1	R	R	R	R	0	0.38	ERR360976
160	7.2	F	2012	Siem Reap	Siem Reab	Sala Kamraeuk	103.867133	13.345557	IV	IVc	0	R	R	R	R	0	0.38	ERR360791
161	9.8	F	2012	Siem Reap	Siem Reab	Sambuor	103.7935	13.30069	IV	IVc	1	R	R	R	R	0	0.25	ERR360963
162	10.4	F	2011	Siem Reap	Siem Reab	Sambuor	103.7935	13.30069	IV	IVc	1	R	R	R	0	0	0	ERR319461
163	11.0	M	2012	Siem Reap	Siem Reab	Siem Reab	103.843962	13.277306	IV	IVc	1	R	R	R	R	0	0.25	ERR360991
164	15.7	F	2012	Siem Reap	Siem Reab	Sla Kram	103.885703	13.367605	IV	IVc	1	R	R	R	R	0	0.38	ERR360789
165	3.7	F	2010	Siem Reap	Siem Reab	Sla Kram	103.885703	13.367605	IV	IVc	1	R	R	R	R	12	0.25	ERR360847
166	1.7	F	2010	Siem Reap	Siem Reab	Sla Kram	103.885703	13.367605	IV	IVc	1	R	R	R	R	12	0.38	ERR360848
167	6.3	M	2010	Siem Reap	Siem Reab	Sla Kram	103.885703	13.367605	IV	IVc	1	R	R	R	0	12	0.25	ERR360960
168	11.4	M	2012	Siem Reap	Siem Reab	Sla Kram	103.885703	13.367605	IV	IVc	0	S	S	S	0	0	0	ERR360970
169	6.4	F	2012	Siem Reap	Siem Reab	Svay Dangkum	103.822374	13.311014	IV	IVc	1	R	R	R	R	0	0.25	ERR331205
170	2.9	F	2012	Siem Reap	Siem Reab	Svay Dangkum	103.822374	13.311014	IV	IVc	1	R	R	R	R	0	0	ERR360780
171	14.1	M	2012	Siem Reap	Siem Reab	Svay Dangkum	103.822374	13.311014	IV	IVc	1	R	R	R	R	0	0.25	ERR360785
172	6.2	F	2012	Siem Reap	Siem Reab	Svay Dangkum	103.822374	13.311014	IV	IVc	1	R	R	R	R	0	0.38	ERR360983
173	14.1	M	2012	Siem Reap	Siem Reab	Svay Dangkum	103.822374	13.311014	IV	IVc	1	R	R	R	R	0	0.25	ERR319489
174	13.2	F	2012	Siem Reap	Soth Nikom	Dam Daek	104.120345	13.255388	IV	IVc	1	R	R	R	R	0	0.19	ERR360967
175	11.6	F	2012	Siem Reap	Soth Nikom	Dam Daek	104.120345	13.255388	IV	IVc	1	R	R	R	R	0	0.25	ERR360975
176	15.1	F	2012	Siem Reap	Soth Nikom	Dan Run	104.087793	13.182907	IV	IVc	1	R	R	R	R	0	0.25	ERR331206
177	3.8	F	2012	Siem Reap	Soth Nikom	Dan Run	104.087793	13.182907	IV	IVc	1	R	R	R	R	0	0.38	ERR360786
178	6.4	M	2010	Siem Reap	Sotr Nikom	Dan Run	104.087793	13.182907	IV	IVc	1	S	R	S	R	12	0.25	ERR360954

179	10.0	M	2011	Siem Reap	Soth Nikom	Dan Run	104.087793	13.182907	IV	IVc	1	R	R	R	0	0	0	ERR319456
180	4.0	M	2011	Siem Reap	Sotr Nikom	Kampong Khleang	104.142061	13.046753	IV	IVc	1	R	R	R	0	0	0	ERR360755
181	7.6	M	2011	Siem Reap	Soth Nikom	Kampong Khleang	104.142061	13.046753	IV	IVc	1	R	R	R	R	0	0.38	ERR360774
182	8.6	F	2010	Siem Reap	Soth Nikom	Kampong Khleang	104.142061	13.046753	IV	IVc	1	R	R	R	0	0	0	ERR360958
183	2.2	F	2011	Siem Reap	Soth Nikom	Kampong Khleang	104.142061	13.046753	IV	IVc	1	R	R	R	0	0	0	ERR360961
184	3.0	F	2012	Siem Reap	Sotr Nikom	Kampong Khleang	104.142061	13.046753	IV	IVc	1	R	R	R	0	0	0	ERR360980
185	#N/A	#N/A	2012	Siem Reap	Soth Nikom	Kampong Khleang	104.142061	13.046753	IV	IVc	1	NA	NA	NA	NA	NA	NA	ERR360985
186	4.8	F	2012	Siem Reap	Soth Nikom	Kampong Khleang	104.142061	13.046753	IV	IVc	1	NA	NA	NA	NA	NA	NA	ERR360994
187	9.3	M	2011	Siem Reap	Soth Nikom	Kampong Khleang	104.142061	13.046753	IV	IVc	1	R	R	R	0	0	0	ERR319442
188	5.1	F	2011	Siem Reap	Soth Nikom	Kampong Khleang	104.142061	13.046753	IV	IVc	1	R	R	R	0	12	0.38	ERR319443
189	2.0	F	2011	Siem Reap	Soth Nikom	Kampong Khleang	104.142061	13.046753	IV	IVc	1	R	R	R	R	0	0	ERR319449
190	5.7	M	2011	Siem Reap	Soth Nikom	Kampong Khleang	104.142061	13.046753	IV	IVc	1	R	R	R	R	12	0.38	ERR319450
191	7.0	M	2011	Siem Reap	Soth Nikom	Kampong Khleang	104.142061	13.046753	IV	IVc	1	R	R	R	0	12	0.38	ERR319451
192	7.4	M	2011	Siem Reap	Soth Nikom	Kampong Khleang	104.142061	13.046753	IV	IVc	1	R	R	R	0	0	0	ERR319468
193	11.8	F	2012	Siem Reap	Soth Nikom	Kampong Khleang	104.142061	13.046753	IV	IVc	1	R	R	R	R	0	0.38	ERR319474
194	7.7	M	2012	Siem Reap	Soth Nikom	Kampong Khleang	104.142061	13.046753	IV	IVc	1	R	R	R	R	0	0	ERR319477
195	3.8	M	2012	Siem Reap	Soth Nikom	Kampong Khleang	104.142061	13.046753	IV	IVc	1	R	R	R	R	0	0	ERR319480
196	10.4	M	2012	Siem Reap	Soth Nikom	Kampong Khleang	104.142061	13.046753	IV	IVc	1	R	R	R	R	0	0	ERR319481
197	12.9	F	2012	Siem Reap	Soth Nikom	Kampong Khleang	104.142061	13.046753	IV	IVc	1	R	R	R	R	0	0	ERR319490
198	6.5	M	2012	Siem Reap	Sotr Nikom	Kampong Khleang	104.142061	13.046753	IV	IVc	1	R	R	R	R	0	0	ERR319492
199	9.5	M	2012	Siem Reap	Sotr Nikom	Khchas	104.06283	13.254239	IV	IVc	1	R	R	R	R	0	0.38	ERR360996
200	4.8	F	2012	Siem Reap	Sotr Nikom	Khnar Pou	104.089546	13.423928	IV	IVc	1	R	R	R	R	0	0.38	ERR331207
201	7.0	M	2012	Siem Reap	Soth Nikom	Khnar Pou	104.089546	13.423928	IV	IVc	0	R	R	R	R	0	0	ERR319488
202	3.1	F	2012	Siem Reap	Soth Nikom	Kien Sangkae	104.160906	13.208966	IV	IVc	1	R	R	R	R	0	0.25	ERR360978
203	10.8	F	2012	Siem Reap	Varin	Lvea Krang	103.912669	13.900364	IV	IVc	1	R	R	R	R	0	0	ERR319478
204	8.7	F	2010	Banteaymeanchey	O Chrauv	Nimitt	102.659645	13.642996	non_H58	non_H58	0	S	S	S	S	6	0.006	ERR360747

205	11.1	F	2007	Siem Reap	Angkor Chum	Char Chhuk	103.623954	13.663726	non_H58	non_H58	1	R	R	R	R	0	0.25	ERR340765
206	7.7	F	2012	Siem Reap	Angkor Chum	Doun Peaeng	103.705808	13.728858	non_H58	non_H58	0	S	S	S	S	0	0.006	ERR360995
207	7.6	F	2011	Siem Reap	Puok	Sasar Sdam	103.630882	13.509808	non_H58	non_H58	0	S	S	S	0	0	0	ERR319462
208	3.8	M	2009	Siem Reap	Siem Reab	Kouk Chak	103.841705	13.429301	non_H58	non_H58	0	S	S	S	R	0	0.25	ERR319418
209	8.4	F	2007	Siem Reap	Siem Reab	Sla Kram	103.885703	13.367605	non_H58	non_H58	1	R	R	R	R	0	0.38	ERR340769
210	12.3	F	2014	Siem Reap	Prasat Bak™ng	Ampil	103.9468022	13.39642561										
211	9.4	F	2013	Siem Reap	Chi Kr□ng	Anlong Samnar	104.2274314	12.94743614										
212	5.2	F	2013	Siem Reap	Chi Kr□ng	Anlong Samnar	104.2274314	12.94743614										
213	6.2	M	2013	Siem Reap	Chi Kr□ng	Anlong Samnar	104.2274314	12.94743614										
214	2.2	M	2013	Siem Reap	Chi Kr□ng	Anlong Samnar	104.2274314	12.94743614										
215	4.8	M	2013	Siem Reap	Chi Kr□ng	Anlong Samnar	104.2274314	12.94743614										
216	9.6	M	2014	Siem Reap	Prasat Bak™ng	Bakong	103.9894232	13.35739817										
217	11.0	M	2014	Siem Reap	Sotr Nik™m	Chan Sar	104.1103659	13.37891805										
218	0.3	F	2014	Banteay Meanchey	Preah NŽtr Preah	Chob Veari	103.1649315	13.66601257										
219	6.6	M	2013	Siem Reap	Siem RŽab	Chong Knies	103.8401794	13.25630554										
220	1.3	F	2014	Siem Reap	Siem RŽab	Chong Knies	103.8401794	13.25630554										
221	12.3	M	2013	Siem Reap	Srei Sn‰m	Chrouy Neang Nguon	103.5517513	13.8336277										
222	8.3	M	2014	Siem Reap	Soth Nikom	Dam Daek	104.1184591	13.25764664										
223	10.3	M	2013	Siem Reap	Soth Nikom	Dan Run	104.0859347	13.18512897										
224	10.7	F	2014	Siem Reap	Soth Nikom	Dan Run	104.0859347	13.18512897										
225	10.0	F	2014	Siem Reap	Soth Nikom	Dan Run	104.0859347	13.18512897										
226	10.7	M	2014	Kampong Thom	Stoung	Kampong Chen Tboung	104.5560628	12.93187723										
227	1.2	M	2013	Siem Reap	Soth Nikom	Kampong Khleang	104.1403332	13.04884688										
228	7.2	F	2013	Siem Reap	Soth Nikom	Kampong Khleang	104.1403332	13.04884688										
229	1.4	M	2013	Siem Reap	Soth Nikom	Kampong Khleang	104.1403332	13.04884688										

230	2.7	M	2013	Siem Reap	Soth Nikom	Kampong Khleang	104.1403332	13.04884688										
231	4.7	F	2013	Siem Reap	Soth Nikom	Kampong Khleang	104.1403332	13.04884688										
232	6.9	F	2013	Siem Reap	Soth Nikom	Kampong Khleang	104.1403332	13.04884688										
233	1.9	F	2013	Siem Reap	Soth Nikom	Kampong Khleang	104.1403332	13.04884688										
234	6.5	F	2013	Siem Reap	Soth Nikom	Kampong Khleang	104.1403332	13.04884688										
235	7.9	F	2013	Siem Reap	Soth Nikom	Kampong Khleang	104.1403332	13.04884688										
236	2.6	F	2014	Siem Reap	Soth Nikom	Kampong Khleang	104.1403332	13.04884688										
237	8.4	M	2014	Siem Reap	Soth Nikom	Kampong Khleang	104.1403332	13.04884688										
238	11.9	F	2014	Siem Reap	Sotr Nik™m	Kampong Khleang	104.1403332	13.04884688										
239	14.1	M	2014	Siem Reap	Prasat Bak™ng	Kandaek	103.9296375	13.30229341										
240	10.4	M	2014	Siem Reap	Prasat Bak™ng	Kandaek	103.9296375	13.30229341										
241	7.7	F	2014	Siem Reap	Prasat Bak™ng	Kandaek	103.9296375	13.30229341										
242	9.1	F	2013	Battambang	æk Phnum	Kaoh Chiveang	103.5919063	13.21889273										
243	1.7	F	2013	Battambang	æk Phnum	Kaoh Chiveang	103.5919063	13.21889273										
244	12.3	F	2013	Battambang	æk Phnum	Kaoh Chiveang	103.5919063	13.21889273										
245	1.7	F	2013	Battambang	æk Phnum	Kaoh Chiveang	103.5919063	13.21889273										
246	3.0	F	2014	Battambang	æk Phnum	Kaoh Chiveang	103.5919063	13.21889273										
247	7.7	M	2014	Siem Reap	Sotr Nik™m	Khchas	104.0609096	13.25652359										
248	12.0	F	2014	Siem Reap	Sotr Nik™m	Khchas	104.0609096	13.25652359										
249	11.4	F	2014	Siem Reap	Svay Leu	Khnang Phnum	104.1521159	13.55969797										
250	6.2	M	2014	Siem Reap	Sotr Nik™m	Kien Sangkae	104.1590757	13.21118358										
251	8.6	F	2014	Siem Reap	Siem RŽab	Kouk Chak	103.8395359	13.43179055										
252	4.2	M	2014	Siem Reap	Siem RŽab	Kouk Chak	103.8395359	13.43179055										
253	12.4	F	2014	Siem Reap	Siem RŽab	Kouk Chak	103.8395359	13.43179055										
254	13.0	M	2014	Siem Reap	ångk™r Chum	Kouk Doung	103.790108	13.7038741										
255	12.0	F	2014	Siem Reap	Puok	Lvea	103.6574108	13.41912907										

256	9.0	M	2014	Banteay Meanchey	O Chrauv	Poipet	102.659645	13.642996										
257	13.0	F	2013	Siem Reap	Sotr Nik™m	Popel	104.1615677	13.3951848										
258	14.3	M	2014	Siem Reap	Sotr Nik™m	Popel	104.1615677	13.3951848										
259	5.5	F	2014	Siem Reap	Puok	Pou Treay	103.5286413	13.37723486										
260	7.8	F	2014	Battambang	Battambang	Preaek Preah Sdach	103.2117684	13.08410463										
261	0.0	F	2014	Preah Vihear	R™vieng	Robieb	105.0841422	13.38869225										
262	11.2	M	2013	Battambang	Battambang	Rotanak	103.2150136	13.09952554										
263	6.6	F	2013	Siem Reap	Chi Kr□ng	Ruessei Lok	104.2258928	13.25354713										
264	3.1	M	2014	Siem Reap	Chi Kr□ng	Ruessei Lok	104.2258928	13.25354713										
265	6.7	F	2013	Siem Reap	Siem RŽab	Sala Kamraeuk	103.8650309	13.347982										
266	4.4	F	2014	Siem Reap	Siem RŽab	Sala Kamraeuk	103.8650309	13.347982										
267	6.8	M	2014	Siem Reap	Siem RŽab	Sala Kamraeuk	103.8650309	13.347982										
268	11.3	F	2014	Siem Reap	Siem RŽab	Siem Reab	103.8437332	13.30476315										
269	11.5	F	2014	Siem Reap	Siem RŽab	Siem Reab	103.8437332	13.30476315										
270	10.5	F	2014	Siem Reap	Siem RŽab	Siem Reab	103.8437332	13.30476315										
271	11.1	M	2013	Siem Reap	Siem RŽab	Sla Kram	103.8836083	13.37003502										
272	11.7	M	2014	Siem Reap	Siem RŽab	Sla Kram	103.8836083	13.37003502										
273	8.1	F	2014	Siem Reap	Siem RŽab	Sla Kram	103.8836083	13.37003502										
274	13.3	F	2014	Siem Reap	Siem RŽab	Sla Kram	103.8836083	13.37003502										
275	5.5	M	2014	Siem Reap	Siem RŽab	Sla Kram	103.8836083	13.37003502										
276	10.0	F	2014	Siem Reap	Varin	Srae Nouy	104.0130835	13.83855219										
277	11.0	M	2013	Siem Reap	Siem RŽab	Svay Dangkum	103.8202681	13.313428										
278	10.0	F	2014	Siem Reap	Siem RŽab	Svay Dangkum	103.8202681	13.313428										
279	9.6	F	2014	Siem Reap	Siem RŽab	Svay Dangkum	103.8202681	13.313428										
280	9.5	M	2014	Siem Reap	Siem RŽab	Svay Dangkum	103.8202681	13.313428										
281	6.1	M	2014	Siem Reap	Siem RŽab	Svay Dangkum	103.8202681	13.313428										

282	9.4	F	2014	Siem Reap	Sotr Nik™m	Ta Yaek	104.1847881	13.30881913										
283	15.3	M	2013	Siem Reap	Prasat Bak™ng	Trapeang Thum	104.0448328	13.30570753										
284	7.5	F	2014	Siem Reap	Puok	Tuek Vil	103.8049637	13.41569467										

Appendix D *Salmonella* Typhi isolates and their corresponding metadata in chapter 6

ID	Study_code	Date	Sex	Age_year	Year	Acute/Carrier	Genotype	H58 status	gyrA_mutation
01TY052	01TY	4-Nov-08	Male	4	2008	Acute	4.3.1	H58	WT
01TY061	01TY	30-Nov-08	Male	16	2008	Acute	4.3.1	H58	S83F
01TY065	01TY	2-Dec-08	Female	23	2008	Acute	4.3.1	H58	S83F
01TY066	01TY	9-Dec-08	Female	11	2008	Acute	4.3.1	H58	WT
01TY069	01TY	11-Dec-08	Male	24	2008	Acute	4.3.1	H58	S83F
01TY072	01TY	8-Dec	Male	18	2008	Acute	4.1	non-H58	WT
01TY073	01TY	16-Dec-08	Female	24	2008	Acute	4.3.1	H58	S83F
01TY075	01TY	8-Dec	Male	14	2008	Acute	3.2.1	non-H58	WT
01TY077	01TY	26-Dec-08	Female	6	2008	Acute	4.3.1	H58	S83F
01TY080	01TY	9-Jan-09	Male	0	2009	Acute	2.2	non-H58	WT
01TY083	01TY	14-Jan-09	Male	6	2009	Acute	4.3.1	H58	WT
01TY086	01TY	30-Jan-09	Male	13	2009	Acute	3.3	non-H58	WT
01TY088	01TY	9-Feb-09	Male	19	2009	Acute	4.3.1	H58	S83F
01TY090	01TY	9-Feb	Female	29	2009	Acute	4.3.1	H58	WT
01TY095	01TY	19-Feb-09	Male	16	2009	Acute	4.3.1	H58	S83F
01TY096	01TY	20-Feb-09	Male	15	2009	Acute	4.3.1	H58	S83F
01TY098	01TY	9-Feb	Female	21	2009	Acute	3.2.1	non-H58	WT
01TY099	01TY	24-Feb-09	Female	19	2009	Acute	2.2	non-H58	WT
01TY101	01TY	9-Feb	Male	24	2009	Acute	4.3.1	H58	WT
01TY102	01TY	2-Mar-09	Male	18	2009	Acute	4.3.1	H58	S83F
01TY103	01TY	9-Mar	Female	14	2009	Acute	4.3.1	H58	WT
01TY104	01TY	9-Mar	Male	15	2009	Acute	4.3.1	H58	WT
01TY105	01TY	9-Mar-09	Male	22	2009	Acute	3.3	non-H58	WT

01TY107	01TY	11-Mar-09	Male	21	2009	Acute	4.3.1	H58	S83F
01TY112	01TY	9-Mar	Female	15	2009	Acute	4.3.1	H58	WT
01TY122	01TY	9-Apr	Female	11	2009	Acute	4.3.1	H58	WT
01TY128	01TY	3-May-09	Male	21	2009	Acute	4.3.1	H58	S83F
01TY132	01TY	17-May-09	Female	28	2009	Acute	4.3.1	H58	S83F
01TY134	01TY	18-May-09	Female	14	2009	Acute	3.0.1	non-H58	WT
01TY139	01TY	14-Jun-09	Male	19	2009	Acute	2.2	non-H58	WT
01TY142	01TY	19-Jun-09	Male	15	2009	Acute	4.3.1	H58	S83F
01TY150	01TY	25-Jun-09	Female	29	2009	Acute	2.2.1	non-H58	WT
01TY151	01TY	25-Jun-09	Female	6	2009	Acute	3.3	non-H58	WT
01TY152	01TY	25-Jun-09	Male	28	2009	Acute	4.3.1	H58	S83F
01TY154	01TY	8-Jul-09	Female	17	2009	Acute	3.3	non-H58	WT
01TY157	01TY	12-Jul-09	Female	24	2009	Acute	4.3.1	H58	S83F
01TY163	01TY	17-Jul-09	Female	27	2009	Acute	3.0.1	non-H58	WT
01TY170	01TY	24-Jul-09	Female	17	2009	Acute	4.3.1	H58	S83F
01TY172	01TY	26-Jul-09	Male	10	2009	Acute	4.3.1	H58	S83F
01TY191	01TY	11-Aug-09	Male	18	2009	Acute	4.3.1	H58	S83F
01TY205	01TY	28-Aug-09	Male	6	2009	Acute	4.3.1	H58	S83F
01TY208	01TY	30-Aug-09	Female	28	2009	Acute	4.3.1	H58	S83F
01TY211	01TY	30-Aug-09	Female	4	2009	Acute	4.3.1	H58	S83F
01TY216	01TY	1-Sep-09	Male	25	2009	Acute	4.3.1	H58	S83F
01TY218	01TY	3-Sep-09	Male	11	2009	Acute	4.3.1	H58	S83F
01TY226	01TY	10-Sep-09	Female	7	2009	Acute	4.3.1	H58	S83F
01TY230	01TY	1-Oct-09	Female	11	2009	Acute	4.3.1	H58	S83F
01TY233	01TY	4-Oct-09	Male	7	2009	Acute	4.3.1	H58	S83F
01TY242	01TY	22-Oct-09	Male	3	2009	Acute	4.3.1	H58	S83F
01TY256	01TY	26-Nov-09	Male	8	2009	Acute	4.3.1	H58	S83F

01TY257	01TY	30-Nov-09	Male	19	2009	Acute	3.3	non-H58	WT
01TY258	01TY	2-Dec-09	Male	19	2009	Acute	3.3	non-H58	WT
01TY261	01TY	6-Dec-09	Male	7	2009	Acute	3.3	non-H58	WT
01TY264	01TY	10-Dec-09	Male	21	2009	Acute	3.3	non-H58	WT
01TY267	01TY	14-Dec-09	Female	4	2009	Acute	3.3	non-H58	WT
01TY274	01TY	18-Jan-10	Male	9	2010	Acute	4.3.1	H58	S83F
01TY283	01TY	5-Mar-10	Male	17	2010	Acute	3.3	non-H58	WT
01TY286	01TY	14-Mar-10	Male	16	2010	Acute	3.3	non-H58	WT
01TY301	01TY	5-Apr-10	Male	16	2010	Acute	4.3.1	H58	S83F
01TY305	01TY	11-Apr-10	Female	22	2010	Acute	4.3.1	H58	S83F
01TY311	01TY	18-Apr-10	Female	9	2010	Acute	3.3	non-H58	WT
01TY312	01TY	18-Apr-10	Male	18	2010	Acute	2.2	non-H58	WT
01TY313	01TY	18-Apr-10	Male	13	2010	Acute	2.2	non-H58	WT
01TY315	01TY	20-Apr-10	Female	11	2010	Acute	2.2	non-H58	WT
01TY327	01TY	4-May-10	Male	0	2010	Acute	4.3.1	H58	S83F
01TY332	01TY	16-May-10	Female	14	2010	Acute	2.2	non-H58	WT
01TY337	01TY	21-May-10	Male	17	2010	Acute	2.2	non-H58	WT
01TY342	01TY	2-Jun-10	Male	0	2010	Acute	4.3.1	H58	S83F
01TY347	01TY	10-Jun-10	Female	20	2010	Acute	4.3.1	H58	S83F
01TY349	01TY	14-Jun-10	Male	14	2010	Acute	2.2	non-H58	WT
01TY366	01TY	6-Jul-10	Female	9	2010	Acute	3.2.2	non-H58	WT
01TY370	01TY	8-Jul-10	Male	18	2010	Acute	4.3.1	H58	S83F
01TY371	01TY	8-Jul-10	Male	12	2010	Acute	4.3.1	H58	S83F
01TY374	01TY	11-Jul-10	Female	8	2010	Acute	4.3.1	H58	S83Y
01TY375	01TY	12-Jul-10	Male	8	2010	Acute	4.3.1	H58	S83Y
01TY391	01TY	21-Jul-10	Male	13	2010	Acute	3.1	non-H58	WT
01TY394	01TY	23-Jul-10	Male	8	2010	Acute	4.3.1	H58	S83F

01TY397	01TY	23-Jul-10	Male	20	2010	Acute	4.3.1	H58	S83Y
01TY405	01TY	1-Aug-10	Female	4	2010	Acute	3.3	non-H58	WT
01TY408	01TY	2-Aug-10	Male	9	2010	Acute	2.2	non-H58	WT
01TY410	01TY	3-Aug-10	Female	31	2010	Acute	4.3.1	H58	WT
01TY421	01TY	9-Aug-10	Male	17	2010	Acute	4.3.1	H58	S83F
01TY426	01TY	10-Aug-10	Male	17	2010	Acute	4.3.1	H58	S83Y
01TY432	01TY	13-Aug-10	Male	19	2010	Acute	4.3.1	H58	S83F
01TY433	01TY	15-Aug-10	Male	19	2010	Acute	4.3.1	H58	S83F
01TY441	01TY	18-Aug-10	Male	28	2010	Acute	4.3.1	H58	S83F
01TY443	01TY	19-Aug-10	Male	18	2010	Acute	4.3.1	H58	S83F
01TY444	01TY	20-Aug-10	Female	17	2010	Acute	4.3.1	H58	S83F
01TY445	01TY	20-Aug-10	Male	18	2010	Acute	4.3.1	H58	S83F
01TY447	01TY	22-Aug-10	Female	8	2010	Acute	4.3.1	H58	S83F
01TY448	01TY	22-Aug-10	Female	13	2010	Acute	3.2.2	non-H58	WT
01TY472	01TY	9-Sep-10	Male	13	2010	Acute	4.3.1	H58	S83Y
01TY476	01TY	13-Sep-10	Male	22	2010	Acute	4.3.1	H58	S83F
01TY481	01TY	19-Sep-10	Male	17	2010	Acute	4.3.1	H58	S83F
01TY484	01TY	22-Sep-10	Male	16	2010	Acute	3.3	non-H58	WT
01TY492	01TY	3-Oct-10	Male	13	2010	Acute	4.3.1	H58	S83F
GB003	GB	12-Jun-09	Female	40	2009	Carrier	4.3.1	H58	S83Y
GB005	GB	5-Nov-07	Female	34	2007	Carrier	3.2.2	non-H58	D87N
GB026	GB	23-Dec-07	Female	27	2007	Carrier	2	non-H58	WT
GB031	GB	26-Jun-09	Female	53	2009	Carrier	4.3.1	H58	S83F
GB044	GB	6-Dec-07	Female	30	2007	Carrier	4.3.1	H58	S83Y
GB071	GB	22-Mar-08	Female	42	2008	Carrier	4.3.1	H58	S83F
GB076	GB	17-Jul-09	Female	35	2009	Carrier	4.3.1	H58	WT
GB125	GB	21-Aug-09	Female	25	2009	Carrier	4.3.1	H58	S83F

GB169	GB	14-Sep-09	Female	27	2009	Carrier	4.3.1	H58	S83F
GB199	GB	29-Sep-07	Female	15	2007	Carrier	4.3.1	H58	S83F
GB266	GB	14-Mar-08	Male	50	2008	Carrier	4.3.1	H58	WT
GB276	GB	1-May-08	Female	27	2008	Carrier	3.3	non-H58	WT
GB281	GB	4-May-08	Female	27	2008	Carrier	2.2.2	non-H58	WT
GB335	GB	6-Jun-08	Female	33	2008	Carrier	3.3	non-H58	WT
GB336	GB	10-Jun-08	Female	30	2008	Carrier	3.3	non-H58	WT
GB368	GB	30-Jun-08	Female	30	2008	Carrier	4.3.1	H58	S83F
GB387	GB	8-Feb-10	Female	60	2010	Carrier	4.3.1	H58	WT
GB428	GB	11-Mar-10	Female	34	2010	Carrier	2.2	non-H58	WT
GB441	GB	19-Mar-10	Male	38	2010	Carrier	4.3.1	H58	WT
GB496	GB	22-Apr-10	Female	20	2010	Carrier	4.3.1	H58	S83F
GB580	GB	12-Dec-08	Female	54	2008	Carrier	2.2	non-H58	WT
GB650	GB	6-Aug-10	Female	34	2010	Carrier	4.3.1	H58	S83F
GB705	GB	16-Sep-10	Male	23	2010	Carrier	3.2.2	non-H58	WT
GB710	GB	20-Sep-10	Female	67	2010	Carrier	4.3.1	H58	S83F

Appendix E Acute-specific nonsynonymous mutations and their functional classes

SNP_Site	S/NS	Ref_ base	SNP _base	ancestral Codon	derived Codon	Ancestral AA	Derived AA	CDS_name	product	functional_classes	Number_acute_isolate	Frequency
2952	N	G	T	CGA	CTA	R	L	STY0003	homoserine kinase	1.D.2 Aspartate Family	2	2.083333333
25425	N	G	A	GAT	AAT	D	N	bcfB	fimbrial chaperone	3.C.3 Surface structures	2	2.083333333
47769	N	T	C	GTG	GCG	V	A	STY0048	transcriptional activator protein NhaR	2 Broad regulatory function	1	1.041666667
50775	N	A	G	ATG	ACG	M	T	STY0050	putative transport protein	4.A Transport/binding proteins	2	2.083333333
54513	N	C	T	GCG	GTG	A	V	STY0055	isoleucyl-tRNA synthetase	3.A.5 Aminoacyl tRNA synthetase	1	1.041666667
58654	N	G	A	GGC	AGC	G	S	STY0058	LytB protein	2 Broad regulatory function	2	2.083333333
62410	N	C	T	CGG	CAG	R	Q	citA	sensor kinase cita	2 Broad regulatory function	1	1.041666667
63682	N	G	T	CGC	AGC	R	S	oadB	oxaloacetate decarboxylase beta chain	4.A Transport/binding proteins	1	1.041666667
71918	N	C	T	GCG	GTG	A	V	citF2	citrate lyase alpha chain	1.C Central Intermediary metabolism	1	1.041666667
83223	N	A	G	TCC	CCC	S	P	STY0081	probable crotonobetaine/carnitine-CoA ligase	1.A.1 Degradation of carbohydrates	1	1.041666667
92753	N	G	A	AGC	AAC	S	N	STY0089	putative metabolite transport protein	4.A Transport/binding proteins	1	1.041666667
101269	N	G	A	CGC	TGC	R	C	STY0103	bis(5'-nucleosyl)-tetraphosphatase	1.F.4 Salvage of nucleosides and nucleotides	1	1.041666667
103365	N	C	T	GCG	ACG	A	T	STY0106	pyridoxal phosphate biosynthetic protein PdxA	1.G.6 Pyridoxine	3	3.125
103932	N	C	T	GCG	ACG	A	T	STY0106	pyridoxal phosphate biosynthetic protein PdxA	1.G.6 Pyridoxine	1	1.041666667
113304	N	G	A	CCG	CTG	P	L	STY0112	DNA polymerase II	3.A.7 DNA - replication, repair, restriction./modification	5	5.208333333
118304	N	C	T	CCT	CTT	P	L	STY0117	hypothetical protein	5.I Unknown	1	1.041666667
144484	N	G	A	GTT	ATT	V	I	STY0142	penicillin-binding protein 3 precursor	3.C.4 Murein sacculus, peptidoglycan	1	1.041666667
151468	N	G	A	GGC	GAC	G	D	STY0148	UDP-N-acetylglucosamine:N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase	3.C.4 Murein sacculus, peptidoglycan	1	1.041666667

162346	N	A	G	TAT	TGT	Y	C	STY0156	preprotein translocase SecA subunit	4.A.6 Transport Other	1	1.041666667
176718	N	T	G	ACC	CCC	T	P	STY0173	aromatic amino acid transport protein AroP	4.A.1 Transport amino acid and amines	1	1.041666667
204893	N	G	T	ATG	ATT	M	I	STY0194	hypothetical ABC transporter ATP-binding protein	4.A Transport/binding proteins	1	1.041666667
237022	N	A	G	ACT	GCT	T	A	STY0225	putative membrane protein	3.C.1 Membranes lipoprotein	1	1.041666667
249957	N	T	G	ATC	CTC	I	L	STY0238	methionine aminopeptidase	3.A.8 Protein translation and modification	5	5.208333333
256630	N	G	A	GCC	ACC	A	T	STY0244	undecaprenyl pyrophosphate synthetase	1.C.3 Sugar-nucleotide biosynthesis, conversions	6	6.25
268918	N	A	G	GAG	GGG	E	G	STY0254	DNA polymerase III, alpha chain	3.A.7 DNA - replication, repair, restriction./modification	2	2.083333333
272059	N	T	C	TGG	CGG	W	R	STY0257	putative secreted chitinase	3.B.4 Degradations of polysaccharides	1	1.041666667
274703	N	C	T	TCC	TTC	S	F	STY0259	lysine decarboxylase	1.B.2 Pyruvate dehydrogenase	1	1.041666667
304517	N	G	A	GCC	GTC	A	V	STY0288	hypothetical protein	SPI-6	1	1.041666667
306763	N	C	T	GGT	GAT	G	D	STY0290	conserved hypothetical protein	SPI-6	1	1.041666667
314369	N	G	A	GGG	AGG	G	R	STY0300	hypothetical protein	SPI-6	1	1.041666667
320218	N	C	T	CCG	CTG	P	L	STY0307	hypothetical protein	SPI-6	2	2.083333333
330250	N	G	A	GGC	AGC	G	S	STY0319	Rhs-family protein	SPI-6	2	2.083333333
337459	N	G	A	GGC	AGC	G	S	STY0324	Rhs-family protein	SPI-6	1	1.041666667
340105	N	C	T	GTT	ATT	V	I	STY0328	hypothetical protein	SPI-6	1	1.041666667
343912	N	C	T	CGC	TGC	R	C	safC	outer-membrane fimbrial usher protein	SPI-6	2	2.083333333
365122	N	G	A	GGT	AGT	G	S	STY0355	phosphoheptose isomerase	3.C.2 Surface polysaccharides & antigens	1	1.041666667
396216	N	G	A	GGG	AGG	G	R	mod	type III restriction-modification system StyLTI enzyme mod	3.A.7 DNA - replication, repair, restriction./modification	1	1.041666667
406182	N	G	A	ATG	ATA	M	I	STY0396	ferrioxamine B receptor precursor	4.A.6 Transport Other	1	1.041666667
406738	2	G	A	CAG	TAG	Q	*	STY0397	RhtC-like transporter	4.A Transport/binding proteins	1	1.041666667
439088	N	T	G	GAA	GAC	E	D	STY0429	exonuclease SbcC	3.B.2 Degradation of DNA	1	1.041666667
472038	N	G	A	GCG	GTG	A	V	phnV	probable membrane component of 2-aminoethylphosphonate transporter	4.A.1 Transport amino acid and amines	1	1.041666667
485521	N	A	G	CTG	CCG	L	P	STY0481	cytochrome o ubiquinol oxidase C subunit	1.B.7.a Aerobic Respiration	2	2.083333333

490825	N	C	T	GTA	ATA	V	I	STY0486	AmpG protein	4.A Transport/binding proteins	1	1.041666667
490875	N	G	A	GCG	GTG	A	V	STY0486	AmpG protein	4.A Transport/binding proteins	1	1.041666667
502112	N	G	A	GTA	ATA	V	I	ppiD	peptidyl-prolyl cis-trans isomerase D	3.A.8 Protein translation and modification	1	1.041666667
518336	N	G	A	CCC	CTC	P	L	ylaB	hypothetical 58.9 kDa protein in tesb-hha intergenic region. Contains probable N-terminal signal sequence	3.C.1 Membranes lipoprotein	2	2.083333333
520037	N	C	A	GGG	TGG	G	W	STY0514	putative membrane protein	3.C.1 Membranes lipoprotein	1	1.041666667
540170	N	G	A	ATG	ATA	M	I	STY0533	ferrochelatase	1.G.12 Heme and porphyrin	1	1.041666667
554496	N	C	T	GGC	AGC	G	S	STY0546	putative membrane protein	3.C.1 Membranes lipoprotein	7	7.291666667
554650	N	G	A	GCA	GTA	A	V	STY0547	putative membrane protein	3.C.1 Membranes lipoprotein	1	1.041666667
556290	N	T	G	GTA	GGA	V	G	STY0548	putative ABC transporter ATP-binding protein	3.A.8 Protein translation and modification	2	2.083333333
557966	N	G	A	CCA	CTA	P	L	ybbN	thioredoxin-like protein	5.I Unknown	1	1.041666667
589822	N	C	A	GCT	TCT	A	S	STY0581	phosphoribosylaminoimidazole carboxylase ATPase subunit	1.F.1 Purine ribonucleotide biosynthesis	2	2.083333333
596512	N	A	C	AAG	CAG	K	Q	fimA	type-1 fimbrial protein, a chain precursor	3.C.3 Surface structures	1	1.041666667
608472	N	C	T	GGG	GAG	G	E	gtrB	bactoprenol glucosyl transferase	3.C.2 Surface polysaccharides & antigens	1	1.041666667
617891	2	C	T	TGG	TAG	W	*	STY0618	putative membrane protein	3.C.1 Membranes lipoprotein	6	6.25
618849	N	A	G	TAC	TGC	Y	C	apeE	outer membrane esterase	4.I Pathogenicity	1	1.041666667
621564	N	G	A	TCC	TTC	S	F	STY0620	oxygen-insensitive NAD(P)H nitroreductase	5.D Drug/Analogue sensitivity	1	1.041666667
624836	N	C	T	GGC	AGC	G	S	STY0626	conserved hypothetical protein	5.H.b Hypothetical protein	1	1.041666667
625772	N	C	T	GAT	AAT	D	N	STY0628	ferrienterobactin receptor precursor	4.A.6 Transport Other	1	1.041666667
639491	N	T	C	CAG	CGG	Q	R	STY0638	ferrienterobactin-binding periplasmic protein precursor	4.A.6 Transport Other	2	2.083333333
641681	N	G	A	GTC	ATC	V	I	STY0640	2,3-dihydroxybenzoate-AMP ligase	1.G.14 Iron uptake and storage	5	5.208333333
643446	N	G	T	GCG	TCG	A	S	STY0641	isochorismatase	1.G.14 Iron uptake and storage	2	2.083333333
648137	N	C	T	CGC	CAC	R	H	STY0646	putative oxidoreductase	5.I Unknown	1	1.041666667
664501	N	A	G	CTG	CCG	L	P	STY0664	regulator of nucleoside diphosphate kinase	1.F.5 Miscellaneous	1	1.041666667

670557	N	A	G	CTG	CCG	L	P	citE	citrate lyase beta chain	1.C Central Intermediary metabolism	7	7.291666667
676249	N	C	T	GCA	ACA	A	T	dcuC	C4-dicarboxylate anaerobic carrier	4.A.3 Transport Carbohydrates, organic acids and alcohols	1	1.041666667
678229	N	G	A	CCC	CTC	P	L	crcB	putative membrane protein	3.C.1 Membranes lipoprotein	2	2.083333333
691764	N	G	A	CCA	CTA	P	L	STY0696	conserved hypothetical protein	5.H.b Hypothetical protein	2	2.083333333
692954	N	C	T	GAC	AAC	D	N	STY0697	DNA polymerase III, delta subunit	3.A.7 DNA - replication, repair, restriction./modification	1	1.041666667
698331	N	T	A	ATC	AAC	I	N	STY0702	probable permease	4.A Transport/binding proteins	1	1.041666667
715716	N	G	A	CAC	TAC	H	Y	STY0718	asparagine synthetase B	1.D.2 Aspartate Family	1	1.041666667
722451	N	A	G	AAC	AGC	N	S	STY0723	pts system, N-acetylglucosamine-specific IIABC component	4.A.3 Transport Carbohydrates, organic acids and alcohols	1	1.041666667
724837	N	C	A	AGC	AGA	S	R	STY0724	glutaminyl-tRNA synthetase	3.A.5 Aminoacyl tRNA synthetase	1	1.041666667
725967	N	G	A	GGT	AGT	G	S	ybfM	putative outer membrane protein	3.C Cell envelope	1	1.041666667
737776	N	G	A	CCG	CTG	P	L	STY0737	possible 5-nitroimidazole antibiotic resistance	5.D Drug/Analogue sensitivity	1	1.041666667
750451	N	C	T	CAT	TAT	H	Y	STY0749	deoxyribodipyrimidine photolyase	3.A.7 DNA - replication, repair, restriction./modification	1	1.041666667
752359	N	C	A	CGC	CTC	R	L	STY0750	PTR2-family transport protein	4.A Transport/binding proteins	1	1.041666667
752996	N	C	T	GGT	AGT	G	S	STY0750	PTR2-family transport protein	4.A Transport/binding proteins	1	1.041666667
761027	N	T	C	GTA	GCA	V	A	STY0760	putative glycosyl transferase	3.C.2 Surface polysaccharides & antigens	1	1.041666667
761098	N	T	G	TGT	GGT	C	G	STY0760	putative glycosyl transferase	3.C.2 Surface polysaccharides & antigens	1	1.041666667
770444	N	C	T	GGT	AGT	G	S	STY0772	putative membrane protein	3.C.1 Membranes lipoprotein	2	2.083333333
771204	N	C	T	GCG	ACG	A	T	STY0773	citrate synthase	1.B.3 Tricarboxylic acid cycle	5	5.208333333
848432	N	A	G	GTG	GCG	V	A	STY0853	HlyD-family secretion protein	4.A Transport/binding proteins	1	1.041666667
862862	N	C	T	GCG	ACG	A	T	STY0871	putative membrane protein	3.C.1 Membranes lipoprotein	1	1.041666667
883410	N	A	G	AGC	GGC	S	G	yliB	putative ABC transporter periplasmic binding protein	4.A Transport/binding proteins	2	2.083333333
884918	N	C	A	CCC	ACC	P	T	yliD	hypothetical ABC transporter permease protein	4.A Transport/binding proteins	2	2.083333333
888699	N	A	G	AAC	AGC	N	S	yliI	putative oxidoreductase	5.I Unknown	1	1.041666667
901044	N	G	A	CGG	CAG	R	Q	STY0908	ribosomal protein S6 modification protein	3.A.2 Ribosomal proteins - synthesis, modification	1	1.041666667

903783	N	T	C	GTG	GCG	V	A	STY0910	putrescine-binding periplasmic protein precursor	4.A.1 Transport amino acid and amines	1	1.041666667
905124	N	C	T	ACG	ATG	T	M	STY0911	putrescine transport ATP-binding protein PotG	4.A.1 Transport amino acid and amines	2	2.083333333
913871	N	G	A	CCG	CTG	P	L	STY0921	arginine transport system permease protein ArtM	4.A.1 Transport amino acid and amines	1	1.041666667
923357	N	C	T	GCC	ACC	A	T	STY0931	pyruvate dehydrogenase	1.A.1 Degradation of carbohydrates	2	2.083333333
942558	N	C	T	GAC	AAC	D	N	STY0954	transport ATP-binding protein CydC	4.A Transport/binding proteins	1	1.041666667
949130	N	G	A	GCC	ACC	A	T	STY0958	cell division protein FtsK	4.C Cell division	1	1.041666667
960373	N	G	A	GGT	AGT	G	S	ycaD	probable transport protein	4.A Transport/binding proteins	1	1.041666667
974754	N	A	G	AAT	GAT	N	D	STY0979	putative lipoprotein	3.C.1 Membranes lipoprotein	1	1.041666667
982823	N	C	T	CCG	TCG	P	S	lpxK	tetraacyldisaccharide 4'-kinase	3.C.2 Surface polysaccharides & antigens	1	1.041666667
1000363	N	C	T	GAC	AAC	D	N	STY1002	outer membrane protein F precursor	3.C Cell envelope	1	1.041666667
1054897	N	G	A	GCC	ACC	A	T	STY1078	aminopeptidase N	3.B.3 Degradation of proteins, peptides, glycoproteins	6	6.25
1086577	N	T	G	AAC	ACC	N	T	pipA	conserved hypothetical protein	SPI-5	1	1.041666667
1092031	2	C	T	TGG	TAG	W	*	pipD	putative secreted peptidase	SPI-5	1	1.041666667
1106048	N	C	T	CAT	TAT	H	Y	hpaA	4-hydroxyphenylacetate 3-monooxygenase operon regulatory protein	1.A.1 Degradation of carbohydrates	1	1.041666667
1130628	N	G	A	GGT	AGT	G	S	STY1167	conserved hypothetical protein	5.H.b Hypothetical protein	2	2.083333333
1133453	N	T	C	GTT	GCT	V	A	STY1170	putative oxidoreductase	5.I Unknown	1	1.041666667
1135518	N	G	A	GTC	ATC	V	I	STY1172	putative 2-hydroxyacid dehydrogenase in phoh-csgg intergenic region	5.I Unknown	1	1.041666667
1137867	N	T	C	TGG	CGG	W	R	STY1175	putative membrane protein	3.C.1 Membranes lipoprotein	2	2.083333333
1139129	N	G	A	CCG	CTG	P	L	STY1176	assembly/transport component in curli production	3.C.3 Surface structures	2	2.083333333
1145923	N	C	T	AGC	AAC	S	N	STY1186	glucans biosynthesis protein	5.F Adaptions and atypical conditions	1	1.041666667
1157933	N	G	A	CAT	TAT	H	Y	STY1199	conserved hypothetical protein	5.H.a Hypothetical protein	2	2.083333333
1190420	N	G	A	GGC	AGC	G	S	STY1236	3-oxoacyl-[acyl-carrier-protein] synthase II	1.H Fatty acid biosynthesis	1	1.041666667
1194025	N	G	A	GGC	AGC	G	S	STY1240	DNA polymerase III, delta' subunit	3.A.7 DNA - replication, repair, restriction./modification	1	1.041666667

1194365	N	A	G	GAG	GGG	E	G	STY1240	DNA polymerase III, delta' subunit	3.A.7 DNA - replication, repair, restriction./modification	1	1.041666667
1195346	N	G	A	GCC	ACC	A	T	STY1241	conserved hypothetical protein	5.H.b Hypothetical protein	4	4.166666667
1208580	N	C	T	GGT	AGT	G	S	STY1256	transcription-repair coupling factor (TrcF)	3.A.7 DNA - replication, repair, restriction./modification	2	2.083333333
1210415	N	C	T	CGT	CAT	R	H	STY1256	transcription-repair coupling factor (TrcF)	3.A.7 DNA - replication, repair, restriction./modification	1	1.041666667
1216818	N	G	A	GTG	ATG	V	M	cobB	putative regulatory protein	2 Broad regulatory function	1	1.041666667
1221269	2	C	T	TGG	TAG	W	*	STY1265	spermidine/putrescine transport system permease protein PotB	4.A.1 Transport amino acid and amines	1	1.041666667
1230675	N	C	T	GCG	ACG	A	T	STY1273	conserved hypothetical protein	5.H.a Hypothetical protein	3	3.125
1233131	N	G	A	CCC	TCC	P	S	STY1277	putative ribosomal large subunit pseudouridine synthase B	3.A.3 Ribosomes - maturation and modification	1	1.041666667
1239073	N	A	G	AGT	GGT	S	G	STY1284	putative invasin	4.I Pathogenicity	6	6.25
1239209	N	C	T	CCT	CTT	P	L	STY1284	putative invasin	4.I Pathogenicity	1	1.041666667
1283954	N	C	A	TGA	TTA	*	L	STY1326	indole-3-glycerol phosphate synthase	1.D.4 Aromatic Amino Acid	1	1.041666667
1285353	N	C	T	GGC	AGC	G	S	STY1327	anthranilate synthase component II; anthranilate phosphoribosyltransferase	1.D.4 Aromatic Amino Acid	1	1.041666667
1288407	N	A	G	ACC	GCC	T	A	STY1330	conserved hypothetical protein	5.H.b Hypothetical protein	1	1.041666667
1306213	N	C	T	GCG	ACG	A	T	STY1347	putative regulatory protein	2 Broad regulatory function	6	6.25
1316960	N	T	C	CAA	CGA	Q	R	STY1355	peptide transport system ATP-binding protein SapF	4.A.6 Transport Other	1	1.041666667
1318086	N	G	A	CCC	CTC	P	L	STY1356	peptide transport system ATP-binding protein SapD	4.A.6 Transport Other	2	2.083333333
1332408	N	C	T	GCC	GTC	A	V	STY1378	transcriptional regulatory protein TyrR	2 Broad regulatory function	1	1.041666667
1332544	N	G	A	ATG	ATA	M	I	STY1378	transcriptional regulatory protein TyrR	2 Broad regulatory function	1	1.041666667
1333454	N	G	A	GAC	AAC	D	N	STY1378	transcriptional regulatory protein TyrR	2 Broad regulatory function	1	1.041666667
1335943	2	G	A	TGG	TAG	W	*	STY1382	putative mandelate racemase / muconate lactonizing enzyme family protein	5.I Unknown	1	1.041666667
1342083	N	C	T	GGA	GAA	G	E	STY1389	putative oxidoreductase	5.I Unknown	5	5.208333333
1342916	N	G	A	GGA	GAA	G	E	STY1390	putative transcriptional regulator	2 Broad regulatory function	1	1.041666667

1351519	N	C	T	GGC	AGC	G	S	STY1399	conserved hypothetical protein	5.H.b Hypothetical protein	1	1.041666667
1358778	N	G	A	GGA	AGA	G	R	STY1408	putative chemo-receptor protein	4.D Chemotaxis and mobility	1	1.041666667
1378120	N	T	C	AAC	AGC	N	S	STY1427	acyl carrier protein phosphodiesterase	1.H Fatty acid biosynthesis	1	1.041666667
1379147	N	A	G	ACC	GCC	T	A	STY1428	ATP-dependent helicase HrpA	3.A.7 DNA - replication, repair, restriction./modification	1	1.041666667
1397389	N	C	T	ACG	ATG	T	M	STY1444	putative glycolate oxidase	5.I Unknown	1	1.041666667
1399474	N	C	T	GAT	AAT	D	N	STY1447	putative ribulose-5-phosphate 3-epimerase	5.I Unknown	1	1.041666667
1402707	N	T	C	GAC	GGC	D	G	STY1451	putative phosphotransferase enzyme	4.A Transport/binding proteins	1	1.041666667
1403493	N	C	T	GAT	AAT	D	N	STY1452	conserved hypothetical protein	5.H.a Hypothetical protein	1	1.041666667
1410199	N	C	T	CCC	TCC	P	S	STY1460	putative peptidase	3.B.3 Degradation of proteins, peptides, glycoproteins	2	2.083333333
1424538	N	C	T	GTC	ATC	V	I	srfA	putative virulence effector protein	4.I Pathogenicity	1	1.041666667
1433892	N	A	C	TCC	GCC	S	A	ansP	L-asparagine permease	4.A.1 Transport amino acid and amines	1	1.041666667
1446984	N	G	A	CAC	TAC	H	Y	STY1490	putative regulatory protein	2 Broad regulatory function	1	1.041666667
1447910	N	G	T	TGG	TTG	W	L	smvA	methyl viologen resistance protein SmvA	5.I Unknown	1	1.041666667
1453048	N	G	A	ACC	ATC	T	I	STY1494	NAD-linked malic enzyme; malate oxidoreductase	1.C.2 Gluconeogenesis	2	2.083333333
1458457	2	G	A	TGG	TAG	W	*	STY1502	putative secreted protein	3.C.1 Membranes lipoprotein	1	1.041666667
1462308	N	G	T	CAG	CAT	Q	H	STY1504	putative hydrolase	5.I Unknown	1	1.041666667
1464345	2	C	T	CAG	TAG	Q	*	STY1505	putative glycogen debranching protein	3.B.4 Degradations of polysaccharides	1	1.041666667
1467534	N	G	T	TGG	TTG	W	L	STY1509	hypothetical protein	5.I Unknown	1	1.041666667
1485774	N	A	C	AAT	ACT	N	T	hyaF2	hydrogenase-1 operon protein HyaF2	1.B.7.a Aerobic Respiration	1	1.041666667
1499896	N	A	G	AAC	GAC	N	D	STY1547	conserved hypothetical protein	5.H.b Hypothetical protein	1	1.041666667
1507830	N	T	C	AAG	GAG	K	E	STY1554	putative membrane transport protein	4.A Transport/binding proteins	2	2.083333333
1516980	N	C	T	TCC	TTC	S	F	STY1566	putative dimethyl sulphoxide reductase subunit	1.A.1 Degradation of carbohydrates	2	2.083333333
1519220	N	T	C	GTA	GCA	V	A	STY1568	putative dimethyl sulphoxide reductase subunit	1.A.1 Degradation of carbohydrates	5	5.208333333
1521052	N	T	C	TTC	TCC	F	S	STY1571	putative ABC transporter periplasmic binding protein	4.A Transport/binding proteins	1	1.041666667
1527528	N	C	T	GGC	GAC	G	D	STY1578	putative regulatory protein	2 Broad regulatory function	6	6.25

1530677	N	G	A	GGT	AGT	G	S	STY1582	putative secreted stress response protein	5.I Unknown	1	1.041666667
1582360	N	G	A	CCG	CTG	P	L	STY1653	fumarate hydratase class II	1.B.3 Tricarboxylic acid cycle	1	1.041666667
1582393	N	G	A	GCG	GTG	A	V	STY1653	fumarate hydratase class II	1.B.3 Tricarboxylic acid cycle	1	1.041666667
1592955	N	G	A	GGC	AGC	G	S	STY1664	putative ferredoxin-like protein, cytoplasmic membrane	1.C Central Intermediary metabolism	2	2.083333333
1601552	N	C	T	GGT	AGT	G	S	pdxY	pyridoxamine kinase	1.C Central Intermediary metabolism	2	2.083333333
1615120	N	G	A	CCG	CTG	P	L	STY1689	conserved hypothetical protein	5.H.b Hypothetical protein	2	2.083333333
1622704	N	C	T	GTG	ATG	V	M	STY1696	riboflavin synthase alpha chain	1.G.9 Riboflavin	1	1.041666667
1656084	N	A	G	ACC	GCC	T	A	ttrB	tetrathionate reductase subunit B	SPI-2	2	2.083333333
1657492	N	A	G	ACA	GCA	T	A	ttrA	tetrathionate reductase subunit A	SPI-2	1	1.041666667
1658116	N	G	A	GGC	AGC	G	S	ttrA	tetrathionate reductase subunit A	SPI-2	1	1.041666667
1658494	N	A	C	ACC	CCC	T	P	ttrA	tetrathionate reductase subunit A	SPI-2	2	2.083333333
1659856	N	T	C	TTT	CTT	F	L	ttrA	tetrathionate reductase subunit A	SPI-2	4	4.166666667
1676627	N	G	A	CAT	TAT	H	Y	STY1757	conserved hypothetical protein	5.H.b Hypothetical protein	1	1.041666667
1707915	N	A	C	AAA	ACA	K	T	STY1786	conserved hypothetical protein	5.H.a Hypothetical protein	1	1.041666667
1717530	N	A	G	TAT	CAT	Y	H	STY1797	phospho-beta-glucosidase B	1.A.1 Degradation of carbohydrates	5	5.208333333
1720562	N	C	A	GAG	GAT	E	D	celB	PTS system, cellobiose-specific IIC component	4.A.3 Transport Carbohydrates, organic acids and alcohols	1	1.041666667
1725400	N	G	A	CAT	TAT	H	Y	astE	succinylglutamate desuccinylase	1.A.2 Degradation of amino acids	2	2.083333333
1726378	N	A	G	GTG	GCG	V	A	astB	succinylarginine dihydrolase	1.A.2 Degradation of amino acids	1	1.041666667
1749721	N	G	A	GGC	AGC	G	S	STY1831	conserved hypothetical protein	5.H.b Hypothetical protein	1	1.041666667
1751929	N	T	C	TTT	TCT	F	S	STY1832	conserved hypothetical protein	5.H.b Hypothetical protein	3	3.125
1757283	N	C	T	CGT	TGT	R	C	STY1840	putative membrane protein	3.C.1 Membranes lipoprotein	1	1.041666667
1771368	N	G	A	CCG	CTG	P	L	STY1862	putative ABC transport ATP-binding protein	4.A Transport/binding proteins	4	4.166666667
1782708	N	G	A	AGG	AAG	R	K	pagD	putative outer membrane virulence protein	4.I Pathogenicity	1	1.041666667
1812529	N	G	A	GTG	ATG	V	M	appC	cytochrome bd-II oxidase subunit I (pseudogene)	1.B.7.c Electron Transport	1	1.041666667
1821293	N	G	A	GGT	GAT	G	D	STY1928	conserved hypothetical protein	5.H.b Hypothetical protein	2	2.083333333

1836701	N	T	C	GAG	GGG	E	G	STY1947	ribonuclease D	3.B.1 Degradation of RNA	1	1.041666667
1843733	N	G	A	GGT	AGT	G	S	STY1954	para-aminobenzoate synthase component I	1.G.2 Folic Acid	1	1.041666667
1856446	N	G	A	CCT	TCT	P	S	ftsI2	penicillin-binding protein	4.C Cell division	1	1.041666667
1863982	N	C	T	GGT	AGT	G	S	STY1976	tail-specific protease precursor	3.B.3 Degradation of proteins, peptides, glycoproteins	2	2.083333333
1884909	N	G	A	CGT	TGT	R	C	STY2004	putative hydrolase	5.I Unknown	1	1.041666667
1938130	N	C	T	GGT	AGT	G	S	STY2085	oligopeptidase	3.B.3 Degradation of proteins, peptides, glycoproteins	1	1.041666667
1940521	N	T	A	ATC	TTC	I	F	STY2087	putative exported protein	3.C.1 Membranes lipoprotein	2	2.083333333
1951699	N	C	T	GGC	GAC	G	D	STY2098	conserved hypothetical protein	5.H.b Hypothetical protein	2	2.083333333
1965470	N	G	A	GAC	AAC	D	N	STY2114	conserved hypothetical protein	5.H.b Hypothetical protein	1	1.041666667
1996037	2	C	T	TGG	TAG	W	*	STY2146	conserved hypothetical protein	5.H.b Hypothetical protein	1	1.041666667
1996313	N	C	T	CGG	CAG	R	Q	STY2146	conserved hypothetical protein	5.H.b Hypothetical protein	6	6.25
1997612	N	T	C	ACC	GCC	T	A	STY2149	putative exported protein	3.C.1 Membranes lipoprotein	1	1.041666667
2001948	N	G	T	CCC	ACC	P	T	STY2154	excinuclease ABC subunit C	3.B.2 Degradation of DNA	1	1.041666667
2013806	N	T	A	CTG	CAG	L	Q	fliD	flagellar hook associated protein 2	3.C.3 Surface structures	1	1.041666667
2018745	N	T	C	TCC	CCC	S	P	STY2174	conserved hypothetical protein	5.H.b Hypothetical protein	1	1.041666667
2053440	N	G	A	CCG	TCG	P	S	STY2218	putative exported protein	3.C.1 Membranes lipoprotein	1	1.041666667
2054602	N	G	A	ACC	ATC	T	I	STY2219	nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyl transferase	1.G.13 Cobalamin	1	1.041666667
2077040	N	A	G	AAA	GAA	K	E	pduE	diol dehydratase small subunit	1.A.1 Degradation of carbohydrates	1	1.041666667
2102317	N	G	A	GCA	ACA	A	T	STY2276	exodeoxyribonuclease I	3.B.2 Degradation of DNA	1	1.041666667
2126184	N	G	A	TCT	TTT	S	F	rfbX	putative O-antigen transporter	3.C.2 Surface polysaccharides & antigens	1	1.041666667
2128153	N	T	C	AAT	GAT	N	D	rfbS	paratose synthase	3.C.2 Surface polysaccharides & antigens	2	2.083333333
2139553	N	A	G	TTC	CTC	F	L	STY2310	putative glycosyltransferase	3.C.2 Surface polysaccharides & antigens	1	1.041666667
2149716	N	G	A	CCG	CTG	P	L	STY2320	GDP-fucose synthetase	1.C.3 Sugar-nucleotide biosynthesis, conversions	1	1.041666667
2156901	N	C	T	GGC	GAC	G	D	wzc	putative tyrosine-protein kinase	3.A.8 Protein translation and modification	1	1.041666667

2157615	N	G	A	CCG	CTG	P	L	wzc	putative tyrosine-protein kinase	3.A.8 Protein translation and modification	1	1.041666667
2158497	N	C	T	GGC	GAC	G	D	wzc	putative tyrosine-protein kinase	3.A.8 Protein translation and modification	1	1.041666667
2169314	N	A	G	CAG	CGG	Q	R	STY2336	putative membrane protein	3.C.1 Membranes lipoprotein	6	6.25
2175898	N	C	T	CCG	CTG	P	L	STY2340	putative RND-family transporter protein	4.A Transport/binding proteins	1	1.041666667
2177044	N	C	T	ACC	ATC	T	I	STY2341	putative RND-family transporter protein	4.A Transport/binding proteins	1	1.041666667
2184586	N	C	T	GCC	GTC	A	V	STY2346	hypothetical protein	5.I Unknown	1	1.041666667
2186873	N	A	G	GAA	GGA	E	G	STY2348	hypothetical protein	5.I Unknown	7	7.291666667
2188648	N	T	G	AGT	CGT	S	R	STY2350	putative exported protein	3.C.1 Membranes lipoprotein	1	1.041666667
2190440	N	C	T	GCT	ACT	A	T	STY2353	putative exported protein	3.C.1 Membranes lipoprotein	1	1.041666667
2190646	N	A	G	GTG	GCG	V	A	STY2353	putative exported protein	3.C.1 Membranes lipoprotein	1	1.041666667
2193219	N	C	T	GAA	AAA	E	K	STY2359	hypothetical protein	5.I Unknown	1	1.041666667
2204169	N	G	A	CGG	CAG	R	Q	STY2373	putative sugar kinase	1.C Central Intermediary metabolism	1	1.041666667
2205213	N	G	A	CCG	CTG	P	L	STY2374	putative gntR-family transcriptional regulator	2 Broad regulatory function	1	1.041666667
2212052	N	G	A	CCA	CTA	P	L	stcB	putative fimbrial chaperone protein	3.C.3 Surface structures	1	1.041666667
2218346	N	T	C	ATC	ACC	I	T	STY2386	putative lipoprotein	3.C.1 Membranes lipoprotein	1	1.041666667
2219526	N	A	C	TTT	GTT	F	V	STY2388	putative two-component system response regulator	2 Broad regulatory function	1	1.041666667
2219654	N	T	G	CAG	CCG	Q	P	STY2388	putative two-component system response regulator	2 Broad regulatory function	1	1.041666667
2219655	2	G	A	CAG	TAG	Q	*	STY2388	putative two-component system response regulator	2 Broad regulatory function	1	1.041666667
2220514	N	A	T	CTG	CAG	L	Q	STY2389	putative two-component system sensor kinase	2 Broad regulatory function	1	1.041666667
2220658	N	G	A	GCC	GTC	A	V	STY2389	putative two-component system sensor kinase	2 Broad regulatory function	1	1.041666667
2221033	N	C	T	AGT	AAT	S	N	STY2389	putative two-component system sensor kinase	2 Broad regulatory function	1	1.041666667
2221417	N	C	T	GGC	GAC	G	D	STY2389	putative two-component system sensor kinase	2 Broad regulatory function	1	1.041666667
2235092	N	T	C	AAA	GAA	K	E	STY2402	putative lipoprotein	3.C.1 Membranes lipoprotein	2	2.083333333
2237767	2	G	A	CAG	TAG	Q	*	STY2405	putative n-hydroxybenzoate hydroxylase	1.A.1 Degradation of carbohydrates	1	1.041666667
2244495	N	A	T	CAG	CTG	Q	L	STY2412	puatative membrane protein	3.C.1 Membranes lipoprotein	2	2.083333333
2256327	N	T	G	AGT	CGT	S	R	STY2426	putative membrane protein	3.C.1 Membranes lipoprotein	1	1.041666667

2258325	N	G	A	CGC	CAC	R	H	STY2428	putative esterase	5.I Unknown	2	2.083333333
2267741	N	C	T	GTG	ATG	V	M	STY2436	putative transcriptional regulator	2 Broad regulatory function	1	1.041666667
2267997	N	C	T	CCT	TCT	P	S	STY2437	putative membrane protein	3.C.1 Membranes lipoprotein	1	1.041666667
2272180	N	C	T	GTC	ATC	V	I	STY2441	1-phosphofructokinase	1.B.1 Glycolysis	1	1.041666667
2272181	N	C	T	ATG	ATA	M	I	STY2441	1-phosphofructokinase	1.B.1 Glycolysis	1	1.041666667
2274674	N	G	A	CGC	CAC	R	H	STY2443	sugar efflux transporter	4.A Transport/binding proteins	1	1.041666667
2326118	N	G	A	GAT	AAT	D	N	STY2494	putative two-component system sensor kinase	2 Broad regulatory function	1	1.041666667
2329159	N	G	A	GCG	GTG	A	V	rcsC	sensor protein RcsC	2 Broad regulatory function	1	1.041666667
2344308	N	C	T	GCG	GTG	A	V	STY2510	putative transcriptional regulator	2 Broad regulatory function	1	1.041666667
2349126	N	G	A	AGT	AAT	S	N	STY2513	anaerobic glycerol-3-phosphate dehydrogenase subunit A	1.B.7.b Anaerobic Respiration	2	2.083333333
2363792	N	G	A	GCG	ACG	A	T	STY2529	putative lipopolysaccharide modification protein	3.C.2 Surface polysaccharides & antigens	1	1.041666667
2384852	N	C	T	GCG	ACG	A	T	STY2550	NADH dehydrogenase I chain J	1.B.7.a Aerobic Respiration	1	1.041666667
2388486	N	G	A	GCG	GTG	A	V	nuoG	NADH dehydrogenase I chain G	1.B.7.a Aerobic Respiration	1	1.041666667
2426730	N	A	G	GTT	GCT	V	A	STY2590	putative amino acid decarboxylase	1.D Amino Acid Biosynthesis	1	1.041666667
2450980	N	T	G	GAT	GCT	D	A	STY2617	conserved hypothetical protein	5.I Unknown	1	1.041666667
2458075	N	G	A	GCG	ACG	A	T	STY2623	long-chain fatty acid transport protein precursor	4.A.6 Transport Other	1	1.041666667
2479104	N	G	A	CAT	TAT	H	Y	STY2644	glucokinase	1.A.1 Degradation of carbohydrates	1	1.041666667
2483383	N	T	C	CTC	CCC	L	P	STY2647	putative ion-channel protein	4.A.2 Transport Cations	2	2.083333333
2486342	N	A	C	ATC	CTC	I	L	STY2650	nucleoside permease NupC	4.A.6 Transport Other	1	1.041666667
2496221	2	G	A	CGA	TGA	R	*	STY2658	xanthosine phosphorylase	1.F.5 Miscellaneous	1	1.041666667
2502367	N	G	C	CCG	GCG	P	A	STY2664	cell division protein	4.C Cell division	1	1.041666667
2512520	N	C	T	GGC	GAC	G	D	STY2677	cysteine synthase B	1.D.3 Serine Family	1	1.041666667
2512550	N	G	A	CCG	CTG	P	L	STY2677	cysteine synthase B	1.D.3 Serine Family	2	2.083333333
2513837	N	G	A	CCG	CTG	P	L	STY2678	sulphate transport ATP-binding protein CysA	4.A.5 Transport Anions	1	1.041666667
2514286	2	G	A	CAG	TAG	Q	*	STY2679	sulphate transport system permease protein CysW	4.A.5 Transport Anions	1	1.041666667
2524516	N	C	T	CGC	CAC	R	H	STY2691	ethanolamine operon transcriptional regulator	2 Broad regulatory function	1	1.041666667

2527158	N	C	T	CGC	CAC	R	H	eutC	ethanolamine ammonia-lyase light chain	1.A Degradation	1	1.041666667
2531257	N	G	A	CCG	CTG	P	L	STY2697	putative membrane protein	3.C.1 Membranes lipoprotein	2	2.083333333
2532336	N	G	A	CGC	TGC	R	C	eutG	putative alchohol dehydrogenase	1.A Degradation	1	1.041666667
2535319	N	C	T	GCG	ACG	A	T	eutN	putative ethanolamine utilization protein EutN	1.A Degradation	1	1.041666667
2535661	N	A	C	TCA	GCA	S	A	eutN	putative ethanolamine utilization protein EutN	1.A Degradation	1	1.041666667
2545292	N	C	T	CCG	CTG	P	L	STY2711	transketolase 2	1.B.5.b Non-oxydative branch	1	1.041666667
2546678	2	C	T	TGG	TAG	W	*	STY2713	putative exported protein	3.C.1 Membranes lipoprotein	1	1.041666667
2560281	N	G	A	GCG	GTG	A	V	STY2723	conserved hypothetical protein	5.H.b Hypothetical protein	1	1.041666667
2578401	N	G	A	GAT	AAT	D	N	STY2741	phosphoribosylglycinamidine myltransferase	1.F.1 Purine ribonucleotide biosynthesis	1	1.041666667
2579223	N	G	A	CGC	CAC	R	H	STY2742	polyphosphate kinase	1.C Central Intermediary metabolism	1	1.041666667
2606065	N	G	T	AAC	AAA	N	K	STY2760	putative exported protein	4.I Pathogenicity	7	7.291666667
2616492	N	C	T	GTA	ATA	V	I	STY2764	putative GTP-binding protein	5.I Unknown	2	2.083333333
2616845	N	G	A	CCG	CTG	P	L	STY2764	putative GTP-binding protein	5.I Unknown	1	1.041666667
2621661	N	G	A	CCG	CTG	P	L	STY2769	putative DNA-binding protein	5.I Unknown	1	1.041666667
2623817	N	T	C	CAG	CGG	Q	R	STY2770	conserved hypothetical protein	5.H.b Hypothetical protein	1	1.041666667
2624579	N	C	T	GGC	AGC	G	S	STY2771	nucleoside diphosphate kinase (ndk)	1.F.1 Purine ribonucleotide biosynthesis	1	1.041666667
2671118	N	G	A	CTT	TTT	L	F	STY2811	putative sensor kinase protein	2 Broad regulatory function	5	5.208333333
2678943	2	C	T	TGG	TGA	W	*	STY2815	putative membrane protein	3.C.1 Membranes lipoprotein	1	1.041666667
2683796	N	A	G	GTA	GCA	V	A	STY2820	putative transmembrane transport protein	4.A Transport/binding proteins	1	1.041666667
2684372	N	A	G	TTG	TCG	L	S	STY2820	putative transmembrane transport protein	4.A Transport/binding proteins	2	2.083333333
2694835	N	T	C	ATA	ATG	I	M	STY2832	sigma-E factor negative regulatory protein	2 Broad regulatory function	1	1.041666667
2727753	2	G	A	CAG	TAG	Q	*	STY2857	phospho-2-dehydro-3-deoxyheptonate aldolase, tyr-sensitive	1.D.4 Aromatic Amino Acid	1	1.041666667
2744407	N	A	G	ACC	GCC	T	A	STY2875	large repetitive protein	SPI-9	2	2.083333333
2746073	N	G	A	GGC	GAC	G	D	STY2875	large repetitive protein	SPI-9	1	1.041666667
2747171	N	C	T	GCG	GTG	A	V	STY2875	large repetitive protein	SPI-9	1	1.041666667

2790994	N	C	T	CTC	TTC	L	F	STY2909	gab protein homolog	5.I Unknown	1	1.041666667
2792931	N	C	T	GCA	GTA	A	V	STY2910	putative GAB DTP gene cluster repressor	5.I Unknown	5	5.208333333
2795907	N	G	A	ATG	ATA	M	I	STY2913	GabA permease (4-amino butyrate transport carrier)	4.A.3 Transport Carbohydrates, organic acids and alcohols	1	1.041666667
2820645	N	T	C	AAT	GAT	N	D	STY2944	gamma-glutamylcysteine synthetase	1.G.10 Thioredoxin	6	6.25
2873905	N	T	C	TCG	CCG	S	P	iagB	cell invasion protein	SPI-1	2	2.083333333
2882909	N	C	T	GAA	AAA	E	K	spaT	unknown function	SPI-1	1	1.041666667
2898283	N	G	A	GAC	AAC	D	N	STY3027	hypothetical protein	SPI-1	5	5.208333333
2909874	N	C	T	ATG	ATA	M	I	STY3041	possible sugar aldolase	1.A.1 Degradation of carbohydrates	6	6.25
2915639	N	G	A	ACA	ATA	T	I	rpoS	RNA polymerase sigma subunit RpoS (sigma-38)	2 Broad regulatory function	1	1.041666667
2920444	N	G	C	CTG	GTG	L	V	ygbP	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	5.I Unknown	2	2.083333333
2921062	N	C	T	GGG	AGG	G	R	STY3056	conserved hypothetical protein	5.H.b Hypothetical protein	2	2.083333333
2922117	N	G	A	GCT	GTT	A	V	cysC	adenosine 5-phosphosulfate kinase	1.C.5 Sulphur Metabolism	1	1.041666667
2922180	N	T	C	CAA	CGA	Q	R	cysC	adenosine 5-phosphosulfate kinase	1.C.5 Sulphur Metabolism	1	1.041666667
2924510	2	G	A	CAG	TAG	Q	*	cysD	ATP sulfurylase (ATP:sulfate adenyltransferase)	1.C.5 Sulphur Metabolism	1	1.041666667
2924810	N	G	A	CGT	TGT	R	C	cysD	ATP sulfurylase (ATP:sulfate adenyltransferase)	1.C.5 Sulphur Metabolism	1	1.041666667
2933498	2	G	A	TGG	TAG	W	*	STY3071	conserved hypothetical protein	5.H.a Hypothetical protein	1	1.041666667
2944324	2	G	A	CAG	TAG	Q	*	STY3081	enolase	1.B.1 Glycolysis	1	1.041666667
2951288	N	T	C	TCG	CCG	S	P	steB	outer membrane usher protein	3.C Cell envelope	1	1.041666667
2954233	N	G	A	GGT	AGT	G	S	steF	fimbrial subunit	3.C.3 Surface structures	1	1.041666667
2974429	N	T	G	CTG	CGG	L	R	STY3108	conserved hypothetical protein	5.H.a Hypothetical protein	6	6.25
2983229	2	C	T	CAA	TAA	Q	*	STY3115	L-fucose permease	4.A.3 Transport Carbohydrates, organic acids and alcohols	2	2.083333333
2987279	N	G	A	GTA	ATA	V	I	STY3118	fuscose operon fucU protein	1.A.1 Degradation of carbohydrates	5	5.208333333
2987784	N	C	T	GCC	GTC	A	V	STY3119	l-fucose operon activator	1.A.1 Degradation of carbohydrates	1	1.041666667
2988676	N	G	T	CCG	CAG	P	Q	STY3120	conserved hypothetical	5.H.b Hypothetical protein	1	1.041666667

									protein			
3002057	N	G	A	CCG	CTG	P	L	recB	exonuclease V subunit	3.B.2 Degradation of DNA	1	1.041666667
3006403	N	G	A	CCG	TCG	P	S	STY3133	protease III precursor (pitrilysin)	3.B.3 Degradation of proteins, peptides, glycoproteins	1	1.041666667
3021088	N	G	A	GTT	ATT	V	I	STY3149	putative membrane protein	3.C.1 Membranes lipoprotein	2	2.083333333
3025188	N	G	A	CCG	TCG	P	S	STY3153	2-acylglycerophosphoethanolamine acyl transferase/acyl carrier protein synthetase	1.H Fatty acid biosynthesis	1	1.041666667
3035138	N	G	A	CGC	TGC	R	C	STY3162	2-keto-3-deoxygluconate oxidoreductase	1.H Fatty acid biosynthesis	1	1.041666667
3039476	N	G	A	ATG	ATA	M	I	STY3167	probable amino acid transport protein	4.A.1 Transport amino acid and amines	1	1.041666667
3046278	2	G	A	CAG	TAG	Q	*	stdB	probable outer membrane fimbrial usher protein	3.C.3 Surface structures	1	1.041666667
3068338	N	G	A	AGC	AAC	S	N	STY3201	flavodoxin II	1.B.7.a Aerobic Respiration	1	1.041666667
3072213	N	A	C	ACC	CCC	T	P	STY3207	6-phospho-beta-glucosidase	1.A.1 Degradation of carbohydrates	1	1.041666667
3088543	N	G	A	GTG	ATG	V	M	STY3220	chromosome intitiation inhibitor	3.A.7 DNA - replication, repair, restriction./modification	1	1.041666667
3098424	N	T	A	CTC	CAC	L	H	STY3232	possible ABC-transport protein, ATP-binding component	4.A Transport/binding proteins	1	1.041666667
3111928	N	T	C	TCG	CCG	S	P	STY3247	conserved hypothetical protein	5.H.b Hypothetical protein	2	2.083333333
3128126	N	G	A	GGG	AGG	G	R	STY3268	nucleoside permease	4.A.6 Transport Other	1	1.041666667
3128789	N	C	T	CCG	TCG	P	S	STY3268	nucleoside permease	4.A.6 Transport Other	1	1.041666667
3136486	N	G	A	GTA	ATA	V	I	STY3285	putative exported protein	SPI-8	1	1.041666667
3138646	N	G	A	GCA	GTA	A	V	STY3290	hypothetical protein	5.H.a Hypothetical protein	2	2.083333333
3140066	N	G	A	GCC	ACC	A	T	STY3293	LysR-family transcriptional regulator	2 Broad regulatory function	1	1.041666667
3147262	2	G	A	CAG	TAG	Q	*	STY3300	conserved hypothetical protein	5.H.b Hypothetical protein	1	1.041666667
3151151	N	G	A	GCG	GTG	A	V	STY3305	hexuronate transporter	4.A.3 Transport Carbohydrates, organic acids and alcohols	1	1.041666667
3183175	2	G	A	TGG	TAG	W	*	STY3340	possible membrane transport protein	4.A Transport/binding proteins	2	2.083333333
3195738	N	C	T	AGC	AAC	S	N	STY3351	topoisomerase IV subunit A	3.A.7 DNA - replication, repair, restriction./modification	1	1.041666667
3217295	N	C	T	GCG	GTG	A	V	STY3377	putative membrane protein	3.C.1 Membranes lipoprotein	1	1.041666667

3219040	N	G	A	GCG	ACG	A	T	STY3378	putative exported protein	3.C.1 Membranes lipoprotein	1	1.041666667
3224368	N	G	A	TCA	TTA	S	L	STY3380	adenyl-transferase	3.A.8 Protein translation and modification	1	1.041666667
3246839	N	T	C	AAT	GAT	N	D	STY3400	conserved hypothetical protein	5.H.a Hypothetical protein	1	1.041666667
3247366	N	G	A	GCC	ACC	A	T	STY3401	conserved hypothetical protein	5.H.a Hypothetical protein	1	1.041666667
3254191	N	G	A	CGT	CAT	R	H	STY3408	putative exported protein	3.C.1 Membranes lipoprotein	1	1.041666667
3258933	2	C	T	CAA	TAA	Q	*	STY3416	conserved hypothetical protein	5.H.b Hypothetical protein	1	1.041666667
3259259	N	G	A	GCG	ACG	A	T	STY3417	conserved hypothetical protein	5.H.a Hypothetical protein	1	1.041666667
3288370	N	G	A	GAA	AAA	E	K	STY3444	galactitol-1-phosphate dehydrogenase	1.A.1 Degradation of carbohydrates	2	2.083333333
3292861	N	T	C	TTT	TCT	F	S	yraN	conserved hypothetical protein	5.H.b Hypothetical protein	1	1.041666667
3361086	N	G	T	CCG	ACG	P	T	STY3519	putative sialic acid transporter	4.A.3 Transport Carbohydrates, organic acids and alcohols	1	1.041666667
3371173	N	G	A	GAA	AAA	E	K	degS	serine protease	3.B.3 Degradation of proteins, peptides, glycoproteins	1	1.041666667
3382903	2	C	T	CAG	TAG	Q	*	STY3543	conserved hypothetical protein	5.H.a Hypothetical protein	1	1.041666667
3444980	N	T	G	ATT	AGT	I	S	STY3592	putative hydrolase	5.I Unknown	5	5.208333333
3455162	N	G	A	CGC	TGC	R	C	recQ	ATP-dependent DNA helicase	3.A.7 DNA - replication, repair, restriction./modification	1	1.041666667
3506652	N	C	T	CAT	TAT	H	Y	STY3649	LysR-family regulatory protein for ilvC expression	1.D.6 Pyruvate Family	1	1.041666667
3507253	N	C	T	CCG	CTG	P	L	STY3649	LysR-family regulatory protein for ilvC expression	1.D.6 Pyruvate Family	2	2.083333333
3508762	N	A	C	TCT	GCT	S	A	STY3652	threonine deaminase	1.D.6 Pyruvate Family	1	1.041666667
3566895	N	G	A	CCA	TCA	P	S	nfi	putative endonuclease	3.B.2 Degradation of DNA	1	1.041666667
3570766	N	G	A	GTC	ATC	V	I	STY3721	thiamine biosynthesis protein	1.G.8 Thiamine	1	1.041666667
3574272	N	C	T	CGT	TGT	R	C	STY3725	thiamine biosynthesis protein	1.G.8 Thiamine	1	1.041666667
3603732	N	G	A	GCG	ACG	A	T	STY3745	tRNA (uracil-5)-methyltransferase	3.A.5 Aminoacyl tRNA synthetase	5	5.208333333
3614668	N	G	A	TGC	TAC	C	Y	STY3754	phosphoenolpyruvate carboxylase	1.B.8 Fermentation	1	1.041666667
3616596	N	C	T	CCC	TCC	P	S	STY3755	putative membrane protein	3.C.1 Membranes lipoprotein	1	1.041666667
3634106	N	C	T	GGG	GAG	G	E	STY3768	bifunctional aspartokinase II/homoserine dehydrogenase IIcan I write	1.D.2 Aspartate Family	7	7.291666667

3642604	N	G	A	CGA	CAA	R	Q	STY3776	transcriptional repressor	2 Broad regulatory function	1	1.041666667
3653696	N	G	A	GGA	AGA	G	R	STY3788	conserved hypothetical protein	5.H.a Hypothetical protein	1	1.041666667
3656589	N	A	T	CTG	CAG	L	Q	STY3792	putative aldolase	5.I Unknown	1	1.041666667
3669388	N	G	A	CGC	TGC	R	C	STY3806	putative membrane protein	3.C.1 Membranes lipoprotein	1	1.041666667
3700304	N	A	G	AAG	AGG	K	R	fdoG	formate dehydrogenase-O, major subunit	1.B.7.b Anaerobic Respiration	2	2.083333333
3702097	N	G	A	CGC	CAC	R	H	STY3842	formate dehydrogenase-O gamma subunit	1.B.7.b Anaerobic Respiration	1	1.041666667
3714939	N	T	C	TTG	TCG	L	S	STY3859	putative glycosyl hydrolase	5.I Unknown	2	2.083333333
3719502	N	G	A	GGC	GAC	G	D	STY3861	putative membrane permease	4.A Transport/binding proteins	1	1.041666667
3719775	N	G	A	GCC	ACC	A	T	STY3862	conserved hypothetical protein	5.H.a Hypothetical protein	1	1.041666667
3729652	N	C	T	GCG	GTG	A	V	STY3874	glutamine synthetase	1.D.1 Glutamate Family	1	1.041666667
3756565	N	G	A	CGC	TGC	R	C	STY3896	high affinity ribose transport protein	4.A.3 Transport Carbohydrates, organic acids and alcohols	2	2.083333333
3757030	N	C	T	GGC	AGC	G	S	STY3896	high affinity ribose transport protein	4.A.3 Transport Carbohydrates, organic acids and alcohols	1	1.041666667
3770081	N	G	A	GTT	ATT	V	I	STY3908	ATP synthase subunit C	1.A Degradation	1	1.041666667
3786020	N	C	T	ATG	ATA	M	I	STY3923	conserved hypothetical protein	5.H.b Hypothetical protein	1	1.041666667
3810322	N	C	T	TCC	TTC	S	F	STY3943	DNA gyrase subunit B	3.A.7 DNA - replication, repair, restriction./modification	3	3.125
3812138	N	C	T	GAA	AAA	E	K	STY3944	putative LysR-family transcriptioanl regulator	2 Broad regulatory function	1	1.041666667
3828144	N	G	T	GAT	TAT	D	Y	STY3957	TorD protein	1.B.7.b Anaerobic Respiration	1	1.041666667
3828146	N	T	A	GAT	GAA	D	E	STY3957	TorD protein	1.B.7.b Anaerobic Respiration	1	1.041666667
3828148	N	T	C	TTT	TCT	F	S	STY3957	TorD protein	1.B.7.b Anaerobic Respiration	1	1.041666667
3828149	N	T	A	TTT	TTA	F	L	STY3957	TorD protein	1.B.7.b Anaerobic Respiration	1	1.041666667
3848849	N	C	T	GGC	GAC	G	D	STY3981	multidrug resistance protein D	5.D Drug/Analogue sensitivity	1	1.041666667
3865841	N	C	T	CGT	TGT	R	C	STY4000	putative PTS system IIC component	4.A.3 Transport Carbohydrates, organic acids and alcohols	1	1.041666667
3872668	N	T	C	AAG	AGG	K	R	STY4009	putative glycosyl hydrolase	5.I Unknown	1	1.041666667
3886919	N	T	C	TAC	CAC	Y	H	mgtB	Magnesium transport ATPase, P-type 2	SPI-3	2	2.083333333

3901514	N	A	G	GAC	GGC	D	G	STY4043	sodium:galactoside family symporter	4.A.3 Transport Carbohydrates, organic acids and alcohols	2	2.083333333
3918863	N	C	T	GTA	ATA	V	I	STY4057	putative beta-lactamase	5.I Unknown	1	1.041666667
3967115	2	C	T	TGG	TAG	W	*	STY4106	putative lipoprotein	3.C.1 Membranes lipoprotein	2	2.083333333
3978351	N	C	T	GGC	AGC	G	S	STY4115	putative sugar kinase	5.I Unknown	1	1.041666667
3982726	N	C	T	GCG	GTG	A	V	STY4118	putative transcriptional regulator	2 Broad regulatory function	6	6.25
3999907	N	A	G	GTC	GCC	V	A	STY4136	xylose operon regulatory protein	2 Broad regulatory function	1	1.041666667
4002684	N	G	A	CGG	CAG	R	Q	STY4138	xylulose kinase	1.A.1 Degradation of carbohydrates	1	1.041666667
4010228	N	C	T	CGT	TGT	R	C	STY4148	putative acetyltransferase	5.I Unknown	1	1.041666667
4019716	N	G	T	GCT	GAT	A	D	STY4160	3-methyladenine DNA glycosylase I, constitutive	3.A.7 DNA - replication, repair, restriction./modification	1	1.041666667
4033249	N	A	G	TGG	CGG	W	R	STY4173	putative amino acid permease	4.A.1 Transport amino acid and amines	1	1.041666667
4042727	N	C	T	GCG	GTG	A	V	STY4182	putative polysaccharide biosynthesis protein subunit B	3.C.2 Surface polysaccharides & antigens	1	1.041666667
4057031	N	A	G	GAC	GGC	D	G	STY4193	putative membrane protein	3.C.1 Membranes lipoprotein	5	5.208333333
4058591	N	A	G	AAC	AGC	N	S	STY4193	putative membrane protein	3.C.1 Membranes lipoprotein	1	1.041666667
4061821	N	T	C	GAG	GGG	E	G	STY4196	hypothetical lysR-family transcriptional regulator	2 Broad regulatory function	2	2.083333333
4071260	N	T	C	AAC	AGC	N	S	STY4205	glutathione reductase	1.G.10 Thioredoxin	1	1.041666667
4072370	N	C	T	GAA	AAA	E	K	STY4206	conserved hypothetical protein	5.H.b Hypothetical protein	1	1.041666667
4087366	N	C	T	GGA	GAA	G	E	STY4222	putative membrane protein	3.C.1 Membranes lipoprotein	1	1.041666667
4087979	N	C	T	GCA	ACA	A	T	STY4222	putative membrane protein	3.C.1 Membranes lipoprotein	5	5.208333333
4090514	N	G	A	GAT	AAT	D	N	STY4224	hypothetical ABC transporter ATP-binding protein	4.A Transport/binding proteins	2	2.083333333
4092099	N	C	T	GCA	GTA	A	V	STY4224	hypothetical ABC transporter ATP-binding protein	4.A Transport/binding proteins	1	1.041666667
4095057	N	C	T	GAT	AAT	D	N	STY4228	conserved hypothetical protein	5.H.b Hypothetical protein	1	1.041666667
4108285	N	C	A	GCG	GAG	A	E	ftsX	cell division protein	4.C Cell division	4	4.166666667
4134238	N	G	A	AGC	AAC	S	N	gntK	putative gluconokinase	1.C Central Intermediary metabolism	1	1.041666667

4137928	2	C	T	CAG	TAG	Q	*	glgB	1,4-alpha-glucan branching enzyme	3.A.10 Polysaccharides - (cytoplasmic)	2	2.083333333
4137929	N	A	G	CAG	CGG	Q	R	glgB	1,4-alpha-glucan branching enzyme	3.A.10 Polysaccharides - (cytoplasmic)	2	2.083333333
4141884	N	C	T	GCC	GTC	A	V	glgC	glucose-1-phosphate adenylyltransferase	3.A.10 Polysaccharides - (cytoplasmic)	1	1.041666667
4150649	N	G	A	GGT	AGT	G	S	glpR	glycerol-3-phosphate regulon repressor	1.B.7.b Anaerobic Respiration	1	1.041666667
4155227	N	C	T	CGC	TGC	R	C	STY4282	maltodextrin phosphorylase	1.A.1 Degradation of carbohydrates	2	2.083333333
4178175	N	C	T	GTT	ATT	V	I	yrfG	putative hydrolase	5.I Unknown	1	1.041666667
4190284	N	C	T	ACC	ATC	T	I	damX	DamX protein	4.C Cell division	6	6.25
4192170	N	C	T	GCC	GTC	A	V	rpe	ribulose-phosphate 3-epimerase	1.A.1 Degradation of carbohydrates	1	1.041666667
4219044	N	T	G	GAC	GCC	D	A	yheS	probable ABC transporter ATP-binding protein	4.A Transport/binding proteins	1	1.041666667
4225205	2	C	A	TGC	TGA	C	*	yheO	conserved hypothetical protein	5.H.b Hypothetical protein	1	1.041666667
4249038	N	A	G	TAC	CAC	Y	H	trkA	potassium transport protein	4.A.2 Transport Cations	1	1.041666667
4253546	N	G	T	TTG	TTT	L	F	STY4392	conserved hypothetical protein	5.H.b Hypothetical protein	2	2.083333333
4282989	N	C	T	GCC	ACC	A	T	STY4415	putative membrane protein	3.C.1 Membranes lipoprotein	1	1.041666667
4289522	N	G	A	GGC	AGC	G	S	STY4421	putative lipoprotein	3.C.1 Membranes lipoprotein	2	2.083333333
4293291	N	C	T	AGC	AAC	S	N	STY4424	maltose transport inner membrane protein	4.A.3 Transport Carbohydrates, organic acids and alcohols	2	2.083333333
4296632	N	C	T	CGC	TGC	R	C	STY4426	maltose/maltodextrin transport ATP-binding protein	4.A.3 Transport Carbohydrates, organic acids and alcohols	1	1.041666667
4296800	N	C	T	CCG	TCG	P	S	STY4426	maltose/maltodextrin transport ATP-binding protein	4.A.3 Transport Carbohydrates, organic acids and alcohols	1	1.041666667
4301966	N	C	A	GAG	GAT	E	D	STY4431	glycerol-3-phosphate acyltransferase	3.A.11 Phospholipids	1	1.041666667
4308235	N	G	A	GAT	AAT	D	N	STY4438	putative exported protein	3.C.1 Membranes lipoprotein	1	1.041666667
4337777	N	G	A	GGC	AGC	G	S	STY4459	large repetitive protein	SPI-4	2	2.083333333
4355566	N	T	A	GTA	GAA	V	E	STY4469	putative membrane protein	3.C.1 Membranes lipoprotein	1	1.041666667
4358445	2	C	T	TGG	TAG	W	*	STY4472	putative membrane protein	3.C.1 Membranes lipoprotein	1	1.041666667
4365926	N	C	T	GCG	GTG	A	V	STY4479	cytochrome c-type biogenesis protein	1.B.7.b Anaerobic Respiration	1	1.041666667
4385697	N	C	A	GCC	TCC	A	S	adi	arginine decarboxylase	1.A.2 Degradation of amino acids	1	1.041666667

4402460	N	C	T	GGG	AGG	G	R	STY4510	putative exported protein	3.C.1 Membranes lipoprotein	1	1.041666667
4409661	N	G	A	GTA	ATA	V	I	STY4521	hypothetical protein	SPI-7	1	1.041666667
4418337	N	G	A	CGG	CAG	R	Q	STY4529	hypothetical protein	SPI-7	1	1.041666667
4421765	N	G	A	GCA	ACA	A	T	STY4534	hypothetical protein	SPI-7	1	1.041666667
4422155	N	C	T	GCT	GTT	A	V	ssB	single strand binding protein	SPI-7	1	1.041666667
4427699	N	T	C	TCC	CCC	S	P	pilO	putative pilus assembly protein	SPI-7	1	1.041666667
4430073	N	G	A	GCC	ACC	A	T	pilQ	nucleotide-binding protein	SPI-7	2	2.083333333
4430715	N	C	T	CTT	TTT	L	F	pilQ	nucleotide-binding protein	SPI-7	1	1.041666667
4440280	N	C	T	CTT	TTT	L	F	STY4558	putative exported protein	SPI-7	6	6.25
4457340	N	G	A	GGG	GAG	G	E	STY4579	putative membrane protein	SPI-7	1	1.041666667
4460065	N	A	G	ATC	GTC	I	V	STY4583	conserved hypothetical protein	SPI-7	1	1.041666667
4461142	N	C	T	GCA	GTA	A	V	STY4584	conserved hypothetical protein	SPI-7	1	1.041666667
4465215	2	G	A	TGG	TAG	W	*	STY4587	hypothetical protein	SPI-7	2	2.083333333
4471515	N	G	A	GAG	AAG	E	K	STY4594	hypothetical protein	SPI-7	1	1.041666667
4516839	N	A	G	TTT	CTT	F	L	tviE	Vi polysaccharide biosynthesis protein TviE, Glycosyl transferases group 1	ViaB operon	6	6.25
4516889	N	G	A	GCC	GTC	A	V	tviE	Vi polysaccharide biosynthesis protein TviE, Glycosyl transferases group 1	ViaB operon	1	1.041666667
4517865	N	A	G	TGT	CGT	C	R	tviE	Vi polysaccharide biosynthesis protein TviE, Glycosyl transferases group 1	ViaB operon	1	1.041666667
4518191	N	A	G	GTT	GCT	V	A	tviE	Vi polysaccharide biosynthesis protein TviE, Glycosyl transferases group 1	ViaB operon	1	1.041666667
4519127	2	C	A	GGA	TGA	G	*	tviD	Vi polysaccharide biosynthesis protein	ViaB operon	1	1.041666667
4521049	N	T	G	AAT	ACT	N	T	tviD	Vi polysaccharide biosynthesis protein	ViaB operon	2	2.083333333
4521241	N	G	A	GCC	GTC	A	V	tviD	Vi polysaccharide biosynthesis protein	ViaB operon	2	2.083333333
4553988	N	G	A	CTT	TTT	L	F	yjeJ	hypothetical protein	5.H.a Hypothetical protein	1	1.041666667

4563879	N	C	T	GCG	GTG	A	V	yjeM	putative amino acid permease	4.A Transport/binding proteins	1	1.041666667
4574488	N	C	T	AGC	AAC	S	N	yjeS	putative 4Fe-4S binding protein	5.I Unknown	1	1.041666667
4575322	N	G	A	GGC	AGC	G	S	yjeF	conserved hypothetical protein	5.H.b Hypothetical protein	6	6.25
4580561	N	C	T	TCA	TTA	S	L	miaA	tRNA delta-2-isopentenylpyrophosphate (IPP) transferase	3.A.5 Aminoacyl tRNA synthetase	2	2.083333333
4580600	N	G	A	GGA	GAA	G	E	miaA	tRNA delta-2-isopentenylpyrophosphate (IPP) transferase	3.A.5 Aminoacyl tRNA synthetase	1	1.041666667
4595445	N	C	T	GCT	GTT	A	V	aidB	probable acyl Co-A dehydrogenase	1.A.1 Degradation of carbohydrates	1	1.041666667
4596506	2	G	T	GAA	TAA	E	*	aidB	probable acyl Co-A dehydrogenase	1.A.1 Degradation of carbohydrates	1	1.041666667
4600253	N	C	T	GGC	AGC	G	S	yjfR	conserved hypothetical protein	5.H.a Hypothetical protein	1	1.041666667
4621008	2	C	T	TGG	TAG	W	*	ytfJ	putative exported protein	3.C.1 Membranes lipoprotein	2	2.083333333
4627402	N	C	T	CCG	TCG	P	S	ytfN	putative exported protein	3.C.1 Membranes lipoprotein	1	1.041666667
4650675	N	C	T	GGA	AGA	G	R	nrdD	anaerobic ribonucleoside-triphosphate reductase	1.F.3 2'-Deoxyribonucleotide metabolism	2	2.083333333
4669035	N	C	T	ACG	ATG	T	M	STY4808	conserved hypothetical protein	5.H.b Hypothetical protein	2	2.083333333
4681321	N	C	T	GAT	AAT	D	N	STY4819	regulatory protein	2 Broad regulatory function	1	1.041666667
4716667	N	C	T	CGC	TGC	R	C	STY4853	hypothetical protein	SPI-10	1	1.041666667
4724452	N	C	G	GCC	CCC	A	P	STY4861	hypothetical protein	5.H.b Hypothetical protein	1	1.041666667
4728683	N	C	T	TCG	TTG	S	L	STY4866	probable aspartate racemase	1.C Central Intermediary metabolism	1	1.041666667
4759459	N	C	T	ATG	ATA	M	I	dnaC	probable DNA replication protein	3.A.7 DNA - replication, repair, restriction./modification	1	1.041666667
4773929	N	C	T	ACA	ATA	T	I	STY4914	conserved hypothetical protein	5.H.b Hypothetical protein	1	1.041666667
4786646	N	A	T	AGC	TGC	S	C	radA	putative DNA repair protein	3.A.7 DNA - replication, repair, restriction./modification	1	1.041666667
4788023	N	G	A	GCC	ACC	A	T	nadR	conserved hypothetical transcriptional regulator	2 Broad regulatory function	4	4.166666667
4798548	N	A	G	AGA	GGA	R	G	STY4937	inner membrane protein CreD	5.B Colicin-related function	1	1.041666667

Appendix F Carrier-specific nonsynonymous mutations and their functional classes

SNP_Site	SNP_Site	S/NS	Ref_base	SNP_base	ancestral Codon	derived Codon	ancestral AA	derived AA	CDS_name	product	Functional_classes	Number_carrier_isolate	Frequency
2725	2725	N	G	T	GCA	TCA	A	S	STY0002	aspartokinase I/homoserine dehydrogenase I	1.D.2 Aspartate Family	1	4.17
122281	122281	N	C	A	CGC	CTC	R	L	STY0120	L-ribulokinase	1.A.1 Degradation of carbohydrate	1	4.17
135985	135985	N	G	A	ATG	ATA	M	I	STY0134	probable activator protein in leuABCD operon	1.D.6 Pyruvate Family	1	4.17
204001	204001	N	G	A	GCT	GTT	A	V	STY0193	carbonic anhydrase	5.I Unknown	1	4.17
241338	241338	2	G	T	GAA	TAA	E	*	STY0230	deoxyguanosinetriphosphate triphosphohydrolase	1.C Central intermediary metabolism	1	4.17
266063	266063	N	A	T	CAC	CTC	H	L	STY0253	ribonuclease HII	3.B.1 Degradation of RNA	1	4.17
275632	275632	N	G	A	GAT	AAT	D	N	STY0259	lysine decarboxylase	1.B.2 Pyruvate dehydrogenase	1	4.17
281721	281721	N	C	T	GTC	ATC	V	I	STY0269	prolyl-tRNA synthetase	3.A.5 Amino acyl tRNA synthesis; tRNA modification	1	4.17
285329	285329	N	A	G	TAT	CAT	Y	H	STY0274	putative ABC transporter ATP-binding protein	4.A Transport/binding proteins	1	4.17
308726	308726	N	C	T	ATG	ATA	M	I	STY0293	hypothetical protein	SPI-6	1	4.17
341744	341744	N	G	A	GTG	ATG	V	M	safA	probable lipoprotein	SPI-6	1	4.17
343595	343595	N	T	C	TAC	CAC	Y	H	safB	periplasmic fimbrial chaperone protein	SPI-6	1	4.17
355324	355324	N	G	A	GGG	AGG	G	R	tsaC	outer membrane fimbrial usher protein	SPI-6	1	4.17
378535	378535	2	C	A	TGC	TGA	C	*	STY0368	probable secreted protein	3.C.1 Membranes lipoprotein	1	4.17
438939	438939	N	A	G	CTG	CCG	L	P	STY0429	exonuclease SbcC	3.B.2 Degradation of DNA	1	4.17
439695	439695	N	C	T	CGT	CAT	R	H	STY0429	exonuclease SbcC	3.B.2 Degradation of DNA	1	4.17
464238	464238	N	C	T	GCA	GTA	A	V	STY0458	thiamine-monophosphate kinase	1.G.8 Thiamine	1	4.17
526144	526144	N	C	T	GAT	AAT	D	N	STY0520	acriflavin resistance protein A precursor	5.D Drug/Analogue sensitivity	1	4.17
531394	531394	N	C	T	GCC	ACC	A	T	STY0523	conserved hypothetical protein	5.H.b Hypothetical protein	1	4.17
582703	582703	N	C	T	GGC	AGC	G	S	allC	allantoate amidohydrolase	1.A.1 Degradation of carbohydrate	1	4.17
588547	588547	N	G	A	GTG	ATG	V	M	arcC	carbamate kinase	1.A.1 Degradation of carbohydrate	1	4.17
599161	599161	2	C	T	CAG	TAG	Q	*	fimD	outer membrane usher protein FimD precursor	3.C.3 Surface structure	1	4.17

661378	661378	N	G	A	GAA	AAA	E	K	STY0661	molybdopterin-containing oxidoreductase membrane anchor subunit	5.I Unknown	1	4.17
661390	661390	N	G	A	GTG	ATG	V	M	STY0661	molybdopterin-containing oxidoreductase membrane anchor subunit	5.I Unknown	1	4.17
697519	697519	N	G	T	GGC	TGC	G	C	STY0701	putative hydrolase C-terminus	5.I Unknown	1	4.17
698447	698447	N	G	A	GCG	ACG	A	T	STY0702	probable permease	4.A Transport/binding proteins	1	4.17
699681	699681	N	A	G	ACC	GCC	T	A	STY0703	putative sigma-54 dependent transcriptional regulator	2 Broad regulatory function	1	4.17
707587	707587	N	C	T	GAA	AAA	E	K	STY0711	apolipoprotein N-acyltransferase	3.C.1 Membranes lipoprotein	1	4.17
725712	725712	N	A	G	ACT	GCT	T	A	ybfM	putative outer membrane protein	3.C Cell envelope	1	4.17
745075	745075	N	T	G	GAT	GCT	D	A	STY0744	sensor protein KdpD	2 Broad regulatory function	1	4.17
760747	760747	N	T	C	TCA	CCA	S	P	STY0760	putative glycosyl transferase	3.C.2 Surface polysaccharides & antigens	1	4.17
773094	773094	N	G	A	CGC	CAC	R	H	STY0775	succinate dehydrogenase cytochrome b-556 subunit	1.B.3 Tricarboxylic acid cycle	1	4.17
796312	796312	N	C	T	CGC	CAC	R	H	ybgR	probable cation transport protein	4.A.2 Transport Cations	1	4.17
825652	825652	N	C	T	CCG	CTG	P	L	STY0827	biotin synthetase	1.G.1 Biotin	1	4.17
834261	834261	N	C	T	GCG	ACG	A	T	STY0835	conserved hypothetical protein	5.H.b Hypothetical protein	1	4.17
836054	836054	N	G	A	CGC	CAC	R	H	STY0836	molybdenum cofactor biosynthesis protein A	1.G.4 Molybdopterin	1	4.17
866031	866031	N	G	A	CCT	CTT	P	L	ybiP	putative membrane protein	3.C.1 Membranes lipoprotein	1	4.17
884217	884217	N	C	T	CGC	TGC	R	C	yliC	hypothetical ABC transporter permease protein	4.A Transport/binding proteins	2	8.33
919324	919324	2	C	T	TGA	TAA	*	*	STY0929	conserved hypothetical protein	5.H.b Hypothetical protein	1	4.17
938733	938733	N	G	A	ACG	ATG	T	M	STY0948	conserved hypothetical protein	5.H.b Hypothetical protein	1	4.17
962365	962365	N	G	A	ATG	ATA	M	I	ycaM	probable transport protein	4.A Transport/binding proteins	1	4.17
969891	969891	N	C	T	GCC	ACC	A	T	STY0975	conserved hypothetical protein	5.H.b Hypothetical protein	1	4.17
987013	987013	N	C	T	GGG	AGG	G	R	STY0992	putative membrane protein	3.C.1 Membranes lipoprotein	1	4.17
990266	990266	N	C	T	CCG	CTG	P	L	STY0995	KicA protein	4.H Cell Killing	1	4.17
993768	993768	N	T	C	TAC	CAC	Y	H	mukB	cell division protein	4.C Cell division	1	4.17
1051566	1051566	N	C	T	GGT	GAT	G	D	STY1076	conserved hypothetical protein	5.H.b Hypothetical protein	1	4.17
1158627	1158627	N	G	A	CAT	TAT	H	Y	STY1200	damage-inducible protein	5.I Unknown	1	4.17

1187285	1187285	N	G	A	GGT	AGT	G	S	STY1232	3-oxoacyl-[acyl-carrier-protein] synthase III	1.H Fatty acid biosynthesis	1	4.17
1196367	1196367	N	A	G	CAC	CGC	H	R	STY1242	PTS system, glucose-specific IIBC component	4.A.3 Transport Carbohydrates, organic acids and alcohols	1	4.17
1196978	1196978	N	G	A	GCG	ACG	A	T	STY1242	PTS system, glucose-specific IIBC component	4.A.3 Transport Carbohydrates, organic acids and alcohols	1	4.17
1196990	1196990	N	G	A	GGT	AGT	G	S	STY1242	PTS system, glucose-specific IIBC component	4.A.3 Transport Carbohydrates, organic acids and alcohols	1	4.17
1208535	1208535	N	A	T	TGG	AGG	W	R	STY1256	transcription-repair coupling factor (TrcF)	3.A.7 DNA - replication, repair, restriction./modification	1	4.17
1241292	1241292	N	A	G	CTG	CCG	L	P	STY1286	nitrate/nitrite sensor protein NarX	1.B.7.b Anaerobic Respiration	1	4.17
1278803	1278803	N	T	G	AAA	CAA	K	Q	STY1321	conserved hypothetical protein	5.H.a Hypothetical protein	1	4.17
1283955	1283955	N	A	T,G	TGA	CGA/AGA	*/*	R/R	STY1326	indole-3-glycerol phosphate synthase	1.D.4 Aromatic Amino Acid	2	8.33
1289493	1289493	N	T	G	GAT	GAG	D	E	STY1331	putative pseudouridine synthase	3.A.3 Ribosomes - maturation and modification	1	4.17
1296666	1296666	N	G	A	GGC	AGC	G	S	STY1337	cys regulon transcriptional activator	1.D.3 Serine Family	1	4.17
1308498	1308498	N	C	T	AGG	AAG	R	K	STY1349	conserved hypothetical protein	5.I Unknown	1	4.17
1313439	1313439	N	C	T	GGC	AGC	G	S	STY1352	enoyl-[acyl-carrier-protein] reductase (NADH)	1.H Fatty acid biosynthesis	1	4.17
1318260	1318260	N	G	T	CCG	CAG	P	Q	STY1356	peptide transport system ATP-binding protein SapD	4.A.6 Transport Other	1	4.17
1325649	1325649	N	A	G	TTC	CTC	F	L	STY1369	peptide transport periplasmic protein SapA precursor	4.A.6 Transport Other	1	4.17
1399300	1399300	N	C	T	GAG	AAG	E	K	STY1447	putative ribulose-5-phosphate 3-epimerase	5.I Unknown	1	4.17
1399483	1399483	N	T	C	ATT	GTT	I	V	STY1447	putative ribulose-5-phosphate 3-epimerase	5.I Unknown	1	4.17
1425078	1425078	N	G	A	CCT	TCT	P	S	srfA	putative virulence effector protein	4.I Pathogenicity	1	4.17
1445458	1445458	N	G	A	GCG	GTG	A	V	STY1489	nitrite extrusion protein	4.A Transport/binding proteins	1	4.17
1458466	1458466	2	G	A	TGG	TAG	W	*	STY1502	putative secreted protein	3.C.1 Membranes lipoprotein	1	4.17
1485540	1485540	N	T	C	CTG	CCG	L	P	hyaF2	hydrogenase-1 operon protein HyaF2	1.B.7.a Aerobic Respiration	1	4.17
1536085	1536085	N	C	T	GCG	ACG	A	T	STY1589	pyridine nucleotide transhydrogenase subunit-alpha	1.C Central intermediary metabolism	1	4.17
1589539	1589539	N	C	T	GCG	GTG	A	V	STY1658	adenosine deaminase	1.F.4 Salvage of nucleosides and nucleotides	1	4.17

1601143	1601143	N	C	T	CAT	TAT	H	Y	STY1671	glutathione S-transferase	1.G.10 Thioredoxin	1	4.17
1606270	1606270	N	G	A	ATG	ATA	M	I	STY1677	outer membrane lipoprotein SlyB precursor	3.C Cell envelope	1	4.17
1656792	1656792	N	G	T	GAC	TAC	D	Y	ttrC	tetrathionate reductase subunit C (membrane protein)	SPI-2	2	8.33
1701598	1701598	N	A	G	ATA	ATG	I	M	rfc	O-antigen polymerase	3.C.2 Surface polysaccharides & antigens	1	4.17
1721795	1721795	2	G	A	CAG	TAG	Q	*	STY1802	osmotically inducible lipoprotein E precursor	5.I Unknown	1	4.17
1725197	1725197	N	C	T	ATG	ATA	M	I	astE	succinylglutamate desuccinylase	1.A.2 Degradation of amino acids	1	4.17
1756933	1756933	N	G	T	CTC	ATC	L	I	STY1839	putative transcriptional regulator	2 Broad regulatory function	2	8.33
1770310	1770310	N	T	C	GAG	GGG	E	G	STY1861	putative ABC transport ATP-binding subunit	4.A Transport/binding proteins	1	4.17
1772818	1772818	N	C	T	GAT	AAT	D	N	STY1864	putative inner membrane transport protein	4.A Transport/binding proteins	1	4.17
1773651	1773651	N	C	T	GCC	ACC	A	T	STY1865	putative substrate-binding transport protein	4.A Transport/binding proteins	1	4.17
1813911	1813911	N	A	G	ACC	GCC	T	A	STY1921	probable cytochrome oxidase subunit II	1.B.7.c Electron Transport	1	4.17
1861881	1861881	N	T	C	TAC	TGC	Y	C	STY1975	heat shock protein	5.F Adaptions and atypical conditions	1	4.17
1876919	1876919	N	C	T	GGG	GAG	G	E	STY1991	putative acetyltransferase	5.I Unknown	2	8.33
1988113	1988113	N	T	C	GTG	GCG	V	A	STY2136	conserved hypothetical protein	5.H.b Hypothetical protein	1	4.17
1995782	1995782	N	G	A	GTC	ATC	V	I	STY2145	tyrosine-specific transport protein	4.A.1 Transport amino acid and amines	1	4.17
2006311	2006311	N	G	A	GCG	GTG	A	V	STY2161	putative deaminase	1.A Degradation	1	4.17
2051229	2051229	N	C	T	ATG	ATA	M	I	STY2216	putative inner membrane protein	3.C.1 Membranes lipoprotein	1	4.17
2054384	2054384	N	A	C	TTT	GTT	F	V	STY2219	nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyl transferase	1.G.13 Cobalamin	1	4.17
2066384	2066384	N	C	T	CGG	CAG	R	Q	cbiE	precorrin-6Y C5,15-methyltransferase [decarboxylating]	1.A.1 Degradation of carbohydrate	1	4.17
2075855	2075855	N	T	C	TAT	CAT	Y	H	pduC	glycerol dehydratase large subunit	1.A.1 Degradation of carbohydrate	1	4.17
2079993	2079993	N	G	A	GCC	ACC	A	T	pduJ	putative propanediol utilization protein PduJ	1.A.1 Degradation of carbohydrate	1	4.17
2120553	2120553	N	G	T	CCT	CAT	P	H	STY2292	phosphomannomutase	3.C.2 Surface polysaccharides & antigens	1	4.17
2127235	2127235	N	A	G	GTC	GCC	V	A	rfbE	CDP-tyvelose-2-epimerase	3.C.2 Surface polysaccharides & antigens	1	4.17

2145889	2145889	N	T	C	GAC	GGC	D	G	manB	phosphomannomutase	3.C.2 Surface polysaccharides & antigens	1	4.17
2151043	2151043	N	C	G	GAG	GAC	E	D	STY2321	GDP-mannose 4,6-dehydratase	1.C.3 Sugar-nucleotide biosynthesis, conversions	1	4.17
2162476	2162476	N	C	T	CAT	TAT	H	Y	yegH	putative membrane protein	3.C.1 Membranes lipoprotein	1	4.17
2162537	2162537	N	T	C	GTC	GCC	V	A	yegH	putative membrane protein	3.C.1 Membranes lipoprotein	1	4.17
2167085	2167085	N	T	G	CTG	CGG	L	R	STY2336	putative membrane protein	3.C.1 Membranes lipoprotein	2	8.33
2176236	2176236	N	G	A	GAT	AAT	D	N	STY2340	putative RND-family transporter protein	4.A Transport/binding proteins	1	4.17
2185609	2185609	N	G	A	GCC	ACC	A	T	STY2348	hypothetical protein	5.I Unknown	1	4.17
2191798	2191798	N	G	A	CCT	CTT	P	L	STY2357	hypothetical protein	5.I Unknown	1	4.17
2220061	2220061	N	G	A	CCA	CTA	P	L	STY2389	putative two-component system sensor kinase	2 Broad regulatory function	1	4.17
2221672	2221672	N	A	G	GTC	GCC	V	A	STY2389	putative two-component system sensor kinase	2 Broad regulatory function	1	4.17
2239555	2239555	N	C	T	CGT	CAT	R	H	STY2407	FAA-hydrolase-family protein	5.I Unknown	1	4.17
2268919	2268919	N	T	C	ATC	ACC	I	T	STY2437	putative membrane protein	3.C.1 Membranes lipoprotein	1	4.17
2331540	2331540	N	G	A	CTT	TTT	L	F	STY2499	DNA gyrase subunit A	3.A.7 DNA - replication, repair, restriction./modification	1	4.17
2333751	2333751	N	C	T	GAC	AAC	D	N	STY2499	DNA gyrase subunit A	3.A.7 DNA - replication, repair, restriction./modification	1	4.17
2343416	2343416	N	G	A	CGC	TGC	R	C	STY2509	putative transmembrane transport protein	4.A Transport/binding proteins	1	4.17
2350279	2350279	N	A	G	CAG	CGG	Q	R	STY2514	anaerobic glycerol-3-phosphate dehydrogenase subunit B	1.B.7.b Anaerobic Respiration	1	4.17
2366071	2366071	N	G	A	GAG	AAG	E	K	STY2530	conserved hypothetical protein	5.H.b Hypothetical protein	1	4.17
2400798	2400798	N	C	T	GTC	ATC	V	I	STY2563	putative sodium/sulphate transporter	5.I Unknown	1	4.17
2419570	2419570	N	C	T	GAC	AAC	D	N	STY2584	histidine-binding periplasmic protein	4.A.1 Transport amino acid and amines	1	4.17
2463468	2463468	N	T	C	TGG	CGG	W	R	STY2629	putative lipopolysaccharide modification acyltransferase	3.C.2 Surface polysaccharides & antigens	1	4.17
2493604	2493604	N	C	T	GAA	AAA	E	K	STY2655	xanthosine operon transcriptional regulator	2 Broad regulatory function	1	4.17
2493630	2493630	N	G	A	GCG	GTG	A	V	STY2655	xanthosine operon transcriptional regulator	2 Broad regulatory function	1	4.17
2514750	2514750	2	C	T	TGG	TAG	W	*	STY2679	sulphate transport system permease protein CysW	4.A.5 Transport Anions	1	4.17

2515747	2515747	N	C	T	GGA	GAA	G	E	STY2681	thiosulphate-binding protein precursor	4.A.5 Transport Anions	1	4.17
2533579	2533579	N	A	T	TGG	AGG	W	R	STY2699	putative ethanolamine utilization protein EutJ	1.A Degradation	1	4.17
2540803	2540803	N	C	A	GGC	TGC	G	C	maeB	NADP-dependent malate dehydrogenase (decarboxylating)	1.D.4 Aromatic Amino Acid	1	4.17
2567455	2567455	N	G	A	TCA	TTA	S	L	STY2730	conserved hypothetical protein	5.H.b Hypothetical protein	1	4.17
2568142	2568142	N	G	A	CCG	TCG	P	S	STY2731	putative permease	4.A Transport/binding proteins	1	4.17
2572830	2572830	N	C	T	CTC	TTC	L	F	STY2735	putative exported protein	3.C.1 Membranes lipoprotein	1	4.17
2590496	2590496	N	G	A	CCG	TCG	P	S	STY2752	inosine-5'-monophosphate dehydrogenase	1.F.1 Purine ribonucleotide biosynthesis	1	4.17
2609489	2609489	N	G	A	GCT	GTT	A	V	STY2760	putative exported protein	4.I Pathogenicity	1	4.17
2611167	2611167	N	C	T	GCG	ACG	A	T	STY2760	putative exported protein	4.I Pathogenicity	1	4.17
2611377	2611377	N	C	T	GCC	ACC	A	T	STY2760	putative exported protein	4.I Pathogenicity	1	4.17
2626146	2626146	N	A	G	TTC	TCC	F	S	STY2773	putative anaerobic reductase component	1.B.7.b Anaerobic Respiration	1	4.17
2630309	2630309	2	G	A	CAG	TAG	Q	*	pbpC	penicillin-binding protein 1C	3.C.4 Murein sacculus, peptidoglycan	1	4.17
2654483	2654483	N	G	A	CCG	TCG	P	S	STY2799	putative exported protein	3.C.1 Membranes lipoprotein	1	4.17
2655783	2655783	N	T	C	ATG	ACG	M	T	STY2800	stationary phase inducible protein CsiE	5.I Unknown	1	4.17
2677249	2677249	N	G	A	GGG	AGG	G	R	STY2813	putative exported protein	3.C.1 Membranes lipoprotein	1	4.17
2694734	2694734	N	G	A	CCG	CTG	P	L	STY2832	sigma-E factor negative regulatory protein	2 Broad regulatory function	1	4.17
2704993	2704993	N	C	T	GCG	GTG	A	V	STY2843	conserved hypothetical protein	5.H.b Hypothetical protein	1	4.17
2725661	2725661	N	T	C	TAC	TGC	Y	C	STY2855	conserved hypothetical protein	5.H.b Hypothetical protein	1	4.17
2733558	2733558	N	C	A	GAT	TAT	D	Y	STY2864	signal recognition particle protein	3.A.8 Protein translation and modification	2	8.33
2748017	2748017	N	G	A	GGC	GAC	G	D	STY2875	large repetitive protein	SPI-9	1	4.17
2805601	2805601	N	C	T	GCC	GTC	A	V	STY2932	ribonucleoside-diphosphate reductase 2 alpha chain	1.F.5 Miscellaneous	1	4.17
2840369	2840369	N	G	A	GCG	ACG	A	T	STY2963	putative rubredoxin reductase	5.I Unknown	1	4.17
2848073	2848073	N	A	G	TGG	CGG	W	R	STY2971	formate hydrogenlyase subunit 5	1.B.8 Fermentation	1	4.17
2854109	2854109	N	G	A	GGT	AGT	G	S	hypC	hydrogenase isoenzymes formation protein HypC	1.B.7.b Anaerobic Respiration	1	4.17
2870474	2870474	N	A	G	AAG	GAG	K	E	STY2996	AraC-family transcriptional regulator	SPI-1	1	4.17

2904779	2904779	N	G	A	CCG	CTG	P	L	STY3036	possible membrane transport protein	4.A Transport/binding proteins	1	4.17
2915322	2915322	N	T	G	ACG	CCG	T	P	rpoS	RNA polymerase sigma subunit RpoS (sigma-38)	2 Broad regulatory function	1	4.17
2915330	2915330	N	T	A	GAA	GTA	E	V	rpoS	RNA polymerase sigma subunit RpoS (sigma-38)	2 Broad regulatory function	1	4.17
2915627	2915627	2	C	T	TGG	TAG	W	*	rpoS	RNA polymerase sigma subunit RpoS (sigma-38)	2 Broad regulatory function	1	4.17
2918204	2918204	N	C	T	GGT	AGT	G	S	STY3052	stationary-phase survival protein	5.I Unknown	1	4.17
2936316	2936316	N	G	A	ATG	ATA	M	I	sopD	possible secreted protein	4.I Pathogenicity	1	4.17
2955365	2955365	N	G	A	GTT	ATT	V	I	STY3091	possible secreted protein	3.C.1 Membranes lipoprotein	1	4.17
2992175	2992175	N	G	A	GGT	AGT	G	S	STY3124	possible aminotransferase	5.I Unknown	1	4.17
2994859	2994859	N	G	A	ACG	ATG	T	M	STY3127	conserved hypothetical protein	5.H.b Hypothetical protein	1	4.17
3051986	3051986	N	G	A	CGG	TGG	R	W	STY3183	hypothetical protein (assiciated with virulence)	4.I Pathogenicity	1	4.17
3053028	3053028	N	G	A	GCC	ACC	A	T	STY3186	conserved hypothetical protein	5.H.b Hypothetical protein	1	4.17
3109192	3109192	N	G	A	GTG	ATG	V	M	STY3244	galactose-proton symport (galactose transporter)	4.A.3 Transport Carbohydrates, organic acids and alcohols	1	4.17
3117804	3117804	N	G	A	GAC	AAC	D	N	STY3254	putative membrane protein	3.C.1 Membranes lipoprotein	2	8.33
3128447	3128447	N	G	A	GAC	AAC	D	N	STY3268	nucleoside permease	4.A.6 Transport Other	1	4.17
3239753	3239753	N	C	T	GGT	AGT	G	S	cheM	methyl-accepting chemotaxis protein II	4.D Chemotaxis and mobility	1	4.17
3311882	3311882	N	G	A	CCT	CTT	P	L	STY3467	protein chain initiation factor 2	3.A.8 Protein translation and modification	1	4.17
3323883	3323883	N	C	T	CCA	CTA	P	L	STY3479	Penicillin-binding protein (D-alanyl-D-alanine carboxypeptidase)	3.C.4 Murein sacculus, peptidoglycan	1	4.17
3326564	3326564	N	C	A	GGC	TGC	G	C	STY3481	putative membrane protein	3.C.1 Membranes lipoprotein	1	4.17
3329720	3329720	N	C	T	GGC	AGC	G	S	STY3486	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	3.C.4 Murein sacculus, peptidoglycan	1	4.17
3364345	3364345	N	C	T	ATG	ATA	M	I	STY3521	putative GntR-family transcriptional regulator	2 Broad regulatory function	1	4.17
3370871	3370871	N	T	C	GTG	GCG	V	A	degS	serine protease	3.B.3 Degradation of proteins, peptides, glycoproteins	1	4.17
3371437	3371437	N	G	A	GAG	AAG	E	K	degS	serine protease	3.B.3 Degradation of proteins, peptides, glycoproteins	1	4.17
3378518	3378518	N	C	T	GCT	ACT	A	T	STY3536	possible membrane transport protein	4.A Transport/binding proteins	1	4.17
3379980	3379980	N	C	T	GGT	AGT	G	S	STY3538	possible GntR-family transcriptional regulator	2 Broad regulatory function	1	4.17

3437215	3437215	N	T	C	TAC	TGC	Y	C	tatD	putative deoxyribonuclease	3.B.2 Degradation of DNA	1	4.17
3449225	3449225	N	C	A	TGG	TGT	W	C	STY3594	5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase	1.D.2 Aspartate Family	1	4.17
3552039	3552039	N	G	T	GAT	TAT	D	Y	STY3708	possible LysR-family transcriptional regulatory protein	2 Broad regulatory function	1	4.17
3602082	3602082	2	C	T	TGG	TAG	W	*	STY3744	vitamin B12 receptor protein	3.C Cell envelope	1	4.17
3652101	3652101	N	A	G	ACG	GCG	T	A	STY3785	putative glycerol metabolic protein	1.C Central intermediary metabolism	1	4.17
3669615	3669615	N	C	T	CGC	CAC	R	H	STY3806	putative membrane protein	3.C Cell envelope	1	4.17
3680371	3680371	N	C	A	CTC	ATC	L	I	STY3819	possible membrane transport protein	4.A Transport/binding proteins	1	4.17
3688431	3688431	N	G	A	GCC	ACC	A	T	STY3827	L-rhamnose isomerase	1.A.1 Degradation of carbohydrate	1	4.17
3716547	3716547	N	G	T	GGC	GTC	G	V	STY3859	putative glycosyl hydrolase	5.I Unknown	1	4.17
3718028	3718028	N	C	T	CTC	TTC	L	F	STY3860	putative membrane permease	4.A Transport/binding proteins	1	4.17
3754576	3754576	N	G	A	GCG	GTG	A	V	STY3894	D-ribose-binding periplasmic protein	1.A.1 Degradation of carbohydrate	1	4.17
3759518	3759518	N	C	T	GGT	AGT	G	S	STY3898	membrane transport protein	4.A Transport/binding proteins	1	4.17
3774538	3774538	N	A	G	ACT	GCT	T	A	STY3913	ATP synthase beta subunit	1.A Degradation	1	4.17
3795418	3795418	2	C	T	CAG	TAG	Q	*	STY3932	putative membrane transport protein	4.A Transport/binding proteins	1	4.17
3808610	3808610	N	G	A	GAC	AAC	D	N	STY3942	recF protein	3.A.7 DNA - replication, repair, restriction./modification	1	4.17
3820902	3820902	N	G	A	CGC	TGC	R	C	STY3951	Two-component sensor protein histidine protein kinase.	1.B.7.b Anaerobic Respiration	1	4.17
3826057	3826057	N	G	A	ATG	ATA	M	I	STY3956	trimethylamine-N-oxide reductase precursor	1.B.7.b Anaerobic Respiration	1	4.17
3828882	3828882	N	G	A	GGT	AGT	G	S	yhjA	probable cytochrome c peroxidase	5.I Unknown	1	4.17
3903736	3903736	N	G	A	GGG	GAG	G	E	STY4044	putative glycosyl hydrolase	5.I Unknown	1	4.17
3926993	3926993	N	G	A	GGC	AGC	G	S	STY4068	formamidopyrimidine-DNA glycosylase	3.A.7 DNA - replication, repair, restriction./modification	1	4.17
3931488	3931488	N	G	A	GCT	ACT	A	T	waaG	lipopolysaccharide core biosynthesis protein	3.C.2 Surface polysaccharides & antigens	1	4.17
3943300	3943300	N	C	T	CGT	CAT	R	H	rfaD	ADP-L-Glycero-D-mannoheptose-6-epimerase	3.C.2 Surface polysaccharides & antigens	1	4.17
3954677	3954677	N	G	A	GGT	AGT	G	S	cysE	serine acetyltransferase	1.D.3 Serine Family	1	4.17

4001985	4001985	N	C	T	GCG	GTG	A	V	STY4137	D-xylose isomerase	1.A.1 Degradation of carbohydrate	1	4.17
4038663	4038663	2	C	T	TGG	TGA	W	*	STY4178	conserved hypothetical protein	5.H.a Hypothetical protein	1	4.17
4042168	4042168	N	T	A	CTG	CAG	L	Q	yhjO	putative polysaccharide biosynthesis protein catalytic subunit	3.C.2 Surface polysaccharides & antigens	1	4.17
4051472	4051472	N	G	A	ATG	ATA	M	I	STY4188	conserved hypothetical protein	5.H.b Hypothetical protein	1	4.17
4117689	4117689	N	C	T	ACG	ATG	T	M	STY4253	putative exported protein	3.C.1 Membranes lipoprotein	1	4.17
4119751	4119751	N	C	T	CGT	TGT	R	C	STY4255	glycerol-3-phosphate transport system permease protein	4.A.3 Transport Carbohydrates, organic acids and alcohols	1	4.17
4124948	4124948	2	C	T	CAA	TAA	Q	*	ggt	gamma-glutamyltranspeptidase precursor	1.G.10 Thioredoxin	1	4.17
4186093	4186093	N	G	C	GGT	GCT	G	A	yrfA	conserved hypothetical protein	5.H.a Hypothetical protein	1	4.17
4225021	4225021	N	G	A	GGG	GAG	G	E	yheO	conserved hypothetical protein	5.H.b Hypothetical protein	1	4.17
4299207	4299207	N	T	C	TTT	TCT	F	S	STY4428	maltose operon periplasmic protein	3.C.1 Membranes lipoprotein	1	4.17
4304442	4304442	N	G	A	ATG	ATA	M	I	STY4433	LexA repressor	2 Broad regulatory function	1	4.17
4308440	4308440	2	G	A	TGG	TAG	W	*	STY4438	putative exported protein	3.C.1 Membranes lipoprotein	1	4.17
4331402	4331402	N	G	A	GGT	AGT	G	S	STY4458	large repetitive protein	SPI-4	1	4.17
4356323	4356323	N	A	G	GAA	GGA	E	G	STY4470	putative membrane protein	3.C.1 Membranes lipoprotein	1	4.17
4359035	4359035	N	C	T	GAG	AAG	E	K	acs	acetyl-coenzyme A synthetase	1.H Fatty acid biosynthesis	1	4.17
4363103	4363103	N	G	A	GTC	ATC	V	I	STY4476	cytochrome c-type protein NrfB precursor	1.B.7.b Anaerobic Respiration	1	4.17
4369734	4369734	N	C	T	GGT	AGT	G	S	yjcO	putative exported protein	3.C.1 Membranes lipoprotein	1	4.17
4412203	4412203	N	C	T	ACA	ATA	T	I	STY4523	hypothetical protein	SPI-7	1	4.17
4416872	4416872	N	T	G	TTC	GTC	F	V	STY4528	hypothetical protein	SPI-7	1	4.17
4428141	4428141	N	C	T	CCA	CTA	P	L	pilO	putative pilus assembly protein	SPI-7	1	4.17
4442998	4442998	2	G	A	TGG	TAG	W	*	STY4562	hypothetical protein	SPI-7	2	8.33
4516890	4516890	N	C	T	GCC	ACC	A	T	tviE	Vi polysaccharide biosynthesis protein TviE, Glycosyl transferases group 1	ViaB operon	1	4.17
4518117	4518117	N	G	A	CAT	TAT	H	Y	tviE	Vi polysaccharide biosynthesis protein TviE, Glycosyl transferases group 1	ViaB operon	1	4.17
4519552	4519552	N	G	T	CCG	CAG	P	Q	tviD	Vi polysaccharide biosynthesis protein	ViaB operon	1	4.17

4520023	4520023	N	C	T	CGC	CAC	R	H	tviD	Vi polysaccharide biosynthesis protein	ViaB operon	1	4.17
4520030	4520030	N	C	T	GTT	ATT	V	I	tviD	Vi polysaccharide biosynthesis protein	ViaB operon	1	4.17
4520036	4520036	N	C	A	GGT	TGT	G	C	tviD	Vi polysaccharide biosynthesis protein	ViaB operon	1	4.17
4521040	4521040	N	T	A	CAG	CTG	Q	L	tviD	Vi polysaccharide biosynthesis protein	ViaB operon	1	4.17
4521233	4521233	N	A	G	TTC	CTC	F	L	tviD	Vi polysaccharide biosynthesis protein	ViaB operon	1	4.17
4523460	4523460	N	C	T	GAT	AAT	D	N	tviB	Vi polysaccharide biosynthesis protein, UDP-glucose/GDP-mannose dehydrogenase	ViaB operon	1	4.17
4528484	4528484	N	A	G	CAG	CGG	Q	R	STY4665	hypothetical protein	SPI-7	1	4.17
4529343	4529343	N	A	G	GAG	GGG	E	G	STY4666	probable phage integrase	SPI-7	1	4.17
4561764	4561764	N	C	T	GCG	ACG	A	T	frdA	fumarate reductase, flavoprotein subunit	SPI-7	1	4.17
4585807	4585807	N	C	T	CGT	TGT	R	C	purA	adenylosuccinate synthetase	1.F.1 Purine ribonucleotide biosynthesis	1	4.17
4594305	4594305	2	G	A	TGG	TAG	W	*	yjfC	conserved hypothetical protein	5.H.a Hypothetical protein	1	4.17
4607828	4607828	N	C	T	GCG	GTG	A	V	STY4751	putative membrane protein	3.C.1 Membranes lipoprotein	1	4.17
4644476	4644476	N	T	C	GTA	GCA	V	A	STY4785	hypothetical protein	5.I Unknown	1	4.17
4726489	4726489	N	G	A	ACC	ATC	T	I	trpS2	probable tryptophanyl-tRNA synthetase	3.A.5 Amino acyl tRNA synthesis; tRNA modification	1	4.17
4748848	4748848	N	C	T	ACA	ATA	T	I	mrr	mrr restriction system protein	3.A.7 DNA - replication, repair, restriction./modification	1	4.17
4757795	4757795	N	C	T	GGC	GAC	G	D	mdoB	putative phosphoglycerol transferase	5.I Unknown	2	8.33
4773347	4773347	N	C	T	CCC	CTC	P	L	STY4912	hypothetical protein	5.I Unknown	1	4.17

eLIFE elifesciences.org

RESEARCH ARTICLE

A novel ciprofloxacin-resistant subclade of H58 *Salmonella* Typhi is associated with fluoroquinolone treatment failure

Duy Pham Thanh[1], Abhilasha Karkey[2], Sabina Dongol[2], Nhan Ho Thi[1], Corinne N Thompson[1,3,4], Maia A Rabaa[1,3], Amit Arjyal[2], Kathryn E Holt[5], Vanessa Wong[6], Nga Tran Vu Thieu[1], Phat Voong Vinh[1], Tuyen Ha Thanh[1], Ashish Pradhan[7], Saroj Kumar Shrestha[7], Damoder Gajurel[7], Derek Pickard[6], Christopher M Parry[4,8], Gordon Dougan[6], Marcel Wolbers[1,3], Christiane Dolecek[1,3,9], Guy E Thwaites[1,3], Buddha Basnyat[2], Stephen Baker[1,3,4]*

[1]The Hospital for Tropical Diseases, Wellcome Trust Major Overseas Programme, Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam; [2]Oxford University Clinical Research Unit, Patan Academy of Health Sciences, Kathmandu, Nepal; [3]Centre for Tropical Medicine and Global Health, Oxford University, Oxford, United Kingdom; [4]The London School of Hygiene and Tropical Medicine, London, United Kingdom; [5]Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Melbourne, Australia; [6]The Wellcome Trust Sanger Institute, Cambridge, United Kingdom; [7]Civil Services Hospital, Kathmandu, Nepal; [8]School of Tropical Medicine and Global Health, Department of Clinical Research, Nagasaki University, Nagasaki, Japan; [9]Department of Zoology, University of Oxford, Oxford, United Kingdom

***For correspondence:** sbaker@oucru.org

Competing interests: The authors declare that no competing interests exist.

Funding: See page 11

Received: 22 December 2015
Accepted: 19 February 2016
Published: 14 March 2016

Reviewing editor: Prabhat Jha, Saint Michael's Hospital, Canada

Abstract The interplay between bacterial antimicrobial susceptibility, phylogenetics and patient outcome is poorly understood. During a typhoid clinical treatment trial in Nepal, we observed several treatment failures and isolated highly fluoroquinolone-resistant *Salmonella* Typhi (*S.* Typhi). Seventy-eight *S.* Typhi isolates were genome sequenced and clinical observations, treatment failures and fever clearance times (FCTs) were stratified by lineage. Most fluoroquinolone-resistant *S.* Typhi belonged to a specific H58 subclade. Treatment failure with *S.* Typhi-H58 was significantly less frequent with ceftriaxone (3/31; 9.7%) than gatifloxacin (15/34; 44.1%)(Hazard Ratio 0.19, p=0.002). Further, for gatifloxacin-treated patients, those infected with fluoroquinolone-resistant organisms had significantly higher median FCTs (8.2 days) than those infected with susceptible (2.96) or intermediately resistant organisms (4.01)(p<0.001). H58 is the dominant *S.* Typhi clade internationally, but there are no data regarding disease outcome with this organism. We report an emergent new subclade of *S.* Typhi-H58 that is associated with fluoroquinolone treatment failure. Clinical trial registration: ISRCTN63006567.
DOI: 10.7554/eLife.14003.001

Introduction

Enteric (typhoid) fever, a systemic infection caused predominantly by the bacterium *Salmonella enterica* subspecies *enterica* serovar Typhi (*S.* Typhi), remains one of the principal bacterial causes of febrile disease in low-income countries (***Parry et al., 2002***). *S.* Typhi is a distinct, monophyletic lineage of *S. enterica* that is exquisitely adapted to cause disease only in humans (***Roumagnac et al., 2006***), characterised by a non-specific fever with malaise and asymptomatic convalescent carriage

eLife digest People who ingest a type of bacteria called *Salmonella* Typhi can develop the symptoms of typhoid fever. This disease is common in low-income settings in Asia and Africa, and causes a high rate of death in people who are not treated with antimicrobial drugs.

During a study in Nepal, Thanh et al. tried to evaluate which of two antimicrobials was better for treating typhoid fever. One of the drugs – called gatifloxacin – did not work in some of the patients. To understand why this treatment failed, Thanh et al. decoded the entire DNA sequences of all the *Salmonella* Typhi bacteria isolated during the study. Comparing this genetic data to the clinical data of the patients identified a new variant of *Salmonella* Typhi. These bacteria have a specific combination of genetic mutations that render them resistant to the family of drugs that gatifloxacin belongs to – the fluoroquinolones.

Patients infected with the variant bacteria and treated with gatifloxacin were highly likely to completely fail treatment and have longer-lasting fevers. On further investigation Thanh et al. found these organisms were likely recently introduced into Nepal from India.

Fluoroquinolones are amongst the most effective and common antimicrobials used to treat typhoid fever and other bacterial infections. However, the presence of bacteria that are resistant to these compounds in South Asia means that they should no longer be the first choice of drug to treat typhoid fever in this location.

DOI: 10.7554/eLife.14003.002

(***Parry et al., 2002***). There are an estimated 20–30 million new cases of enteric fever per year globally (***Crump and Mintz, 2010***), with the majority occurring in Asia, but there is an increasingly recognised burden of disease across sub-Saharan Africa.

Antimicrobial resistance is a major global health challenge, and resistance against the most commonly used antimicrobials for treating enteric fever has evolved successively over the last 30 years. Ampicillin, chloramphenicol and trimethoprim-sulfamethoxazole were originally standard-of-care for enteric fever. However, multidrug resistance (MDR) against these agents began to emerge in the 1970s and 1980s (***Olarte and Galindo, 1973***; ***Wain et al., 2003***). Consequently, third-generation cephalosporins and fluoroquinolones (FQs) became the most clinically reliable drugs for treating enteric fever (***Kariuki et al., 2015***), and were formally advocated by the World Health Organization (WHO) in 2003 (***World Health Organization, 2003***). *S.* Typhi isolates with acquired resistance against third-generation cephalosporins are rare (***Hendriksen et al., 2015***), but *S.* Typhi exhibiting reduced susceptibility to FQs, induced by sequential mutations in the gene encoding a target protein (*gyrA*), now dominate internationally (***Emary et al., 2012***; ***Kariuki et al., 2010***). The global ascendency of *S.* Typhi strains with reduced susceptibility to FQs has been partly catalysed by the dissemination of a specific MDR lineage (H58) across Asia and Africa (***Wong et al., 2015***). These H58 strains are rapidly displacing other lineages, and strains with *gyrA* mutations may have a fitness advantage, even in the absence of antimicrobial exposure (***Baker et al., 2013***).

We have previously shown that protracted fever clearance times (FCTs) are associated with organisms with higher Minimum Inhibitory Concentrations (MIC) against FQs in enteric fever patients treated with ciprofloxacin and ofloxacin (***Parry et al., 2011***). However, whilst the clinical efficacy of the older FQs in enteric fever is contentious, we have shown that the fourth-generation FQ, gatifloxacin, has remained efficacious for uncomplicated disease, even in patients infected with *S.* Typhi strains with reduced ciprofloxacin susceptibility (MIC $\geq$0.125 µg/mL) (***Pandit et al., 2007***; ***Koirala et al., 2013***; ***Arjyal et al., 2011***).

During a recent randomised controlled trial (RCT) comparing ceftriaxone and gatifloxacin, conducted in Nepal, we observed an increased number of treatment failures associated with FQ-resistant (ciprofloxacin MIC>32 µg/ml) *S.* Typhi, prompting the data safety and monitoring board to stop the trial (***Arjyal et al., 2016***). Aiming to assess the molecular epidemiology of the infecting isolates and investigate how genotype may be related to treatment outcome, we performed whole genome sequencing (WGS) on the *S.* Typhi isolated during this trial, and after stratifying by genotype, we assessed clinical presentation and outcome.

Results

Salmonella Typhi whole genome sequencing

We performed WGS on the 78 available *S.* Typhi isolates from patients in both RCT treatment arms (gatifloxacin and ceftriaxone) (***Supplementary file 1***). The resulting phylogeny, which incorporated reference sequence CT18, indicated that the majority of isolates (65/78; 83.3%) fell within the H58 lineage, while the remaining 13 (16.7%) represented eight different lineages (***Figure 1***). All but four of the H58 strains contained the common DNA gyrase (*gyrA*) mutation in codon 83 (S83F), which confers reduced susceptibility to FQs (ciprofloxacin MIC; 0.125–0 5 µg/ml) (***Parry et al., 2010***). Nested within the S83F H58 group, but separated from the rest of the group by a branch defined by 30 SNPs, was an H58 subclade comprised of 12 isolates containing the S83F *gyrA* mutation, a mutation in *gyrA* at codon 87 (D87N), and an additional mutation in the topoisomerase gene, *parC* (S80I) (H58 triple mutant). Notably, these H58 triple mutants shared high MICs against ciprofloxacin (≥24 µg/ml). Further, an additional two non-H58 RCT isolates with ciprofloxacin MIC≥24 µg/ml had the S83F *gyrA* mutation, an alternative mutation at codon 87 (D87V), the S80I *parC* mutation, and an A364V mutation in *parE* (***Figure 1***, ***Supplementary file 1***). Notably, none of the sequenced isolates harboured plasmid-mediated quinolone resistance genes (PMQR) or contained additional antimicrobial resistance genes within the well-described *S.* Typhi-associated IncH1 family of plasmids.

Clinical presentation of *Salmonella* Typhi infections

We stratified clinical data from the RCT by H58 status of the corresponding *S.* Typhi isolates (H58; N=65, non-H58; N=13) and compared baseline characteristics between these groups. We found no significant differences in demographics and no association between disease severity at presentation

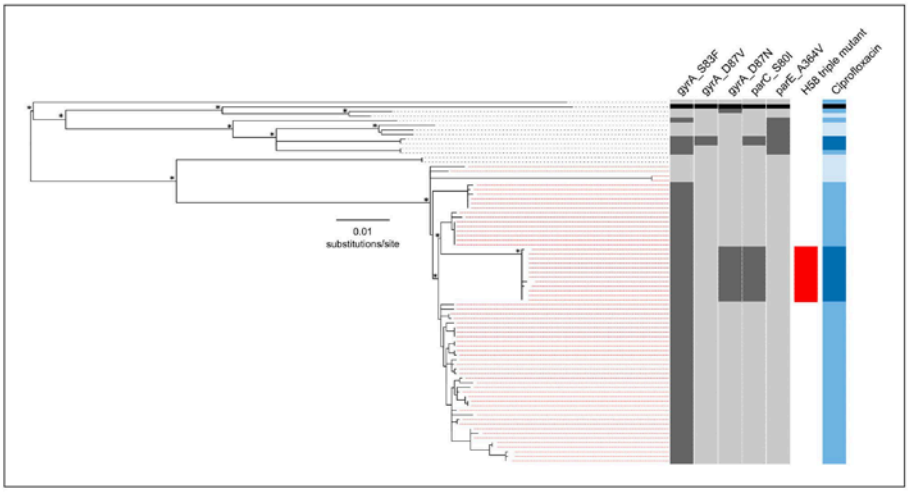

Figure 1. The phylogenetic structure of 78 Nepali *Salmonella* Typhi isolated during a gatifloxacin versus ceftriaxone randomised controlled trial. Maximum likelihood phylogeny based on core-genome SNPs of 78 *Salmonella* Typhi RCT isolates with the corresponding metadata, including the presence of mutations (dark grey) in *gyrA* (S83F, D87V and D87N), *parC* (S80I) and *parC* (A364V) and susceptibility to ciprofloxacin (susceptible, light blue; intermediate, mid-blue and non-susceptible, dark blue) by Minimum Inhibitory Concentration (MIC). The reference strain CT18 was used for context and highlighted by the black boxes. Red lines linking to metadata show isolates belonging to the *Salmonella* Typhi H58 lineage (with H58 triple mutants highlighted), other lineages (non-H58) are shown with black lines. The scale bar indicates the number of substitutions per variable site (see methods). Asterisks indicate ≥85% bootstrap support at nodes of interest.
DOI: 10.7554/eLife.14003.003

between those infected with an H58 *S.* Typhi isolate or a non-H58 isolate (***Supplementary file 2A***). Next, we compared the baseline characteristics of patients stratified by ciprofloxacin susceptibility (susceptible, intermediate and resistant), and found no differences in disease severity or demographics on presentation; the only exception being that FQ-resistant *S.* Typhi were more frequently isolated from adults (***Supplementary file 2B***). A significantly lower proportion of H58 *S.* Typhi (4/65; 6.2%) were susceptible to FQs compared to non-H58 isolates (6/13; 46%) (p=0.001) (***Table 1***) and, overall, H58 isolates had significantly higher (but not resistant) MICs against the majority of tested antimicrobials than non-H58 isolates (***Table 1***).

Treatment failure and fever clearance times

The primary endpoint of the RCT in which these data were generated was a composite for treatment failure (see method and previous publication) (***Arjyal et al., 2016***). Treatment failure with H58 *S.* Typhi was significantly less common in the ceftriaxone group (3/31; 9.7%) than the gatifloxacin group (15/34; 44.1%) (Hazard Ratio (HR) of time to failure 0.19, 95%CI 0.05–0.56, p=0.002) (***Table 2***). Conversely, there was no significant difference in treatment failure between those infected with non-H58 isolates treated with gatifloxacin (0/6; 0%) or ceftriaxone (2/7; 28.6%) (p=0.32). Similarly, time to fever clearance differed significantly between the two treatment groups in H58 infections, with median FCTs of 5.03 days (interquartile range (IQR): 3.18–7.21) in the gatifloxacin group and 3.07 days (IQR: 1.89–4.52) in the ceftriaxone group (p<0.0006). Again, this trend was not mirrored in the non-H58 *S.* Typhi infections, with FCTs of 2.87 (IQR: 2.08–3.7) and 3.12 (IQR: 2.2–4.12) days for gatifloxacin and ceftriaxone, respectively (p=0.61) (***Table 3***). Moreover, in the gatifloxacin arm, H58 *S.* Typhi tended to be associated with a higher risk of treatment failure (p=0.06) and a longer fever clearance time (p=0.013) (***Figure 2***, ***Table 2*** and ***Supplementary file 2C***).

As we identified two non-H58 isolates that were also FQ-resistant (***Figure 1***), we additionally stratified outcome for the gatifloxacin arm (N=40 patients) by FQ susceptibility of the infecting organism. Those infected with FQ-resistant *S.* Typhi failed gatifloxacin treatment more frequently (8/10; 80%) than those infected with an intermediately resistant organism (7/25; 28%) or a susceptible organism (0/5; 0%) (p=0.007) (***Figure 2*** and ***Table 2***). Furthermore, in the gatifloxacin arm, those infected with FQ-resistant organisms had significantly higher median FCTs than those infected with

Table 1. Comparison of antimicrobial susceptibility by *Salmonella* Typhi lineage.

	Non-H58 (N=13)			H58 (N=65)			p value*
E test	**MIC50**	**MIC90**	**GM (range)**	**MIC50**	**MIC90**	**GM (range)**	
Amoxicillin	0.5	1	0.77 (0.38–38)	0.75	>256	1.43 (0.38–>256)	0.0412
Chloramphenicol	3	4	2.7 (1.5–8)	4	12	5.7 (2–>256)	0.0147
Ceftriaxone	0.06	0.06	0.06 (0.05–0.13)	0.09	0.19	0.11 (0.03–0.64)	0.0004
Gatifloxacin	0.13	0.25	0.06 (0.01–2)	0.13	2	0.21 (0.01–3)	0.1197
Nalidixic acid	>256	>256	21.6 (1–>256)	>256	>256	346.8 (1–>256)	0.0004
Ofloxacin	0.25	0.75	0.24 (0.03–>32)	0.5	>3232	1.09 (0.03–>32)	0.0240
Trimethoprim sulphate	0.02	0.05	0.03 (0.02–0.05)	0.05	0.32	0.09 (0.01–>32)	0.0016
Ciprofloxacin	0.13	0.75	0.11 (0.01–>32)	0.38	>32	0.80 (0.02–>32)	0.0051
Ciprofloxacin susceptibility group							0.0008#
- Susceptible	6 (46.2%)			4 (6.2%)			
- Intermediate	4 (30.8%)			48 (73.8%)			
- Resistant	3 (23.1%)			13 (20.0%)			

*Comparisons between *Salmonella* Typhi lineage for MICs and ciprofloxacin susceptibility groups were based on the Wilcoxon rank sum test and Fisher's exact test. respectively.

MIC: minimum inhibitory concentration, measured in µg/ml

#p value for comparison of susceptible vs. intermediate/resistant combined between groups by Fisher's exact test is 0.001.

GM: geometric mean, the upper range of the values was determined by multiplying the MIC by 2 if the result was >X (for example, >256 = 256*2 = 512).

DOI: 10.7554/eLife.14003.004

Table 2. Summary of time to treatment failure by *Salmonella* Typhi lineage and ciprofloxacin susceptibility.

Time to treatment failure	Gatifloxacin (events/N)	Ceftriaxone (events/N)	Hazard ratio of time to failure (95%CI); p value	Heterogeneity test (p value)
H58*				0.020
- H58	15/34	3/31	0.19 (0.05, 0.56); p=0.002	
- Non-H58	0/6	2/7	3.87 (0.31, 534.24); p=0.32	
Ciprofloxacin susceptibility group†				0.08
- Susceptible	0/5	1/5	2.40 (0.13, 350.21); p=0.57	
- Intermediate	7/25	2/27	0.27 (0.05, 0.99); p=0.049	
- Resistant	8/10	2/6	0.27 (0.05, 1.01); p=0.052	

*Likelihood ratio test p=0.06 and 0.40 for comparison of time to treatment failure between H58 vs. non-H58 groups in gatifloxacin arm only and in all patients, respectively

†Likelihood ratio test p=0.007 for comparison of time to treatment failure between MIC groups in gatifloxacin arm only

DOI: 10.7554/eLife.14003.005

S. Typhi with alternative FQ susceptibility profiles (median FCTs (days): susceptible, 2.96 (IQR: 2.13–3.85), intermediate, 4.01 (IQR: 2.76–5.37) and resistant 8.2 (IQR: 5.99–10.5), respectively [p<0.0001]) (***Table 3*** and ***Supplementary file 2D***). Comparatively, the median FCT for those infected with an FQ-resistant organism but randomised to ceftriaxone was 3.83 days (IQR: 2.96–4.7) (p<0.0001 for the between-treatment comparison).

The emergence of fluoroquinolone-resistant *Salmonella* Typhi

To measure the pattern of emergence of FQ-resistant *S.* Typhi in Nepal, we compiled FQ susceptibility data from 837 organisms isolated during enteric fever RCTs conducted at Patan Hospital between 2005 and 2014 (***Figure 3***) (***Pandit et al., 2007***; ***Koirala et al., 2013***; ***Arjyal et al., 2011***). MICs against FQs were generally higher for *S.* Paratyphi A than for *S.* Typhi. There was a significant temporal increase in *S.* Typhi MICs against both ciprofloxacin (p<0.0001) and gatifloxacin (p<0.0001), with a sharp increase from 2009. MICs against gatifloxacin in *S.* Paratyphi A also significantly increased with time (p<0.0001); however, MICs against ciprofloxacin showed only weak evidence of an upward trend over time (p=0.06).

We hypothesised that the H58 triple mutants represented a contemporary importation into Nepal. To explore this, we compared the genomes of the 78 RCT *S.* Typhi isolates with those from 58 supplementary *S.* Typhi isolates from previous studies conducted between 2008 and 2013 in this setting (***Figure 4***, ***Supplementary file 1***) (***Wong et al., 2015***). We found that the majority of the local H58 isolates (84/121; 69.4%) were closely related; these strains represented an 'endemic' Nepali H58 clade containing a single S83F *gyrA* mutation. Additionally, we identified a further five Nepali strains isolated in 2013 that belonged to the H58 triple mutant group, and had an MIC ≥24 µg/ml against ciprofloxacin. Incorporating additional genome sequences from a recent international study of the H58 lineage (***Wong et al., 2015***), we found that all the Nepali H58 triple mutants were very closely related (5 SNPs to nearest neighbour) to H58 triple mutants isolated previously in neighbouring India between 2008 and 2012 (***Figure 4***).

Discussion

Our study shows that a new FQ-resistant subclade of H58 *S.* Typhi has been introduced into Nepal and is associated with a lack of FQ efficacy. This subclade was associated with longer FCTs and treatment failure in patients treated with the FQ, gatifloxacin. For the first time, we can conclusively show how enteric fever patients respond to FQ treatment when infected with a specific subclade of H58, thereby linking organism genotype with a treatment phenotype. Given the international significance of FQs for the treatment of enteric fever and other bacterial infections, our findings have major global health implications for the long-term use and efficacy of this group of antimicrobials.

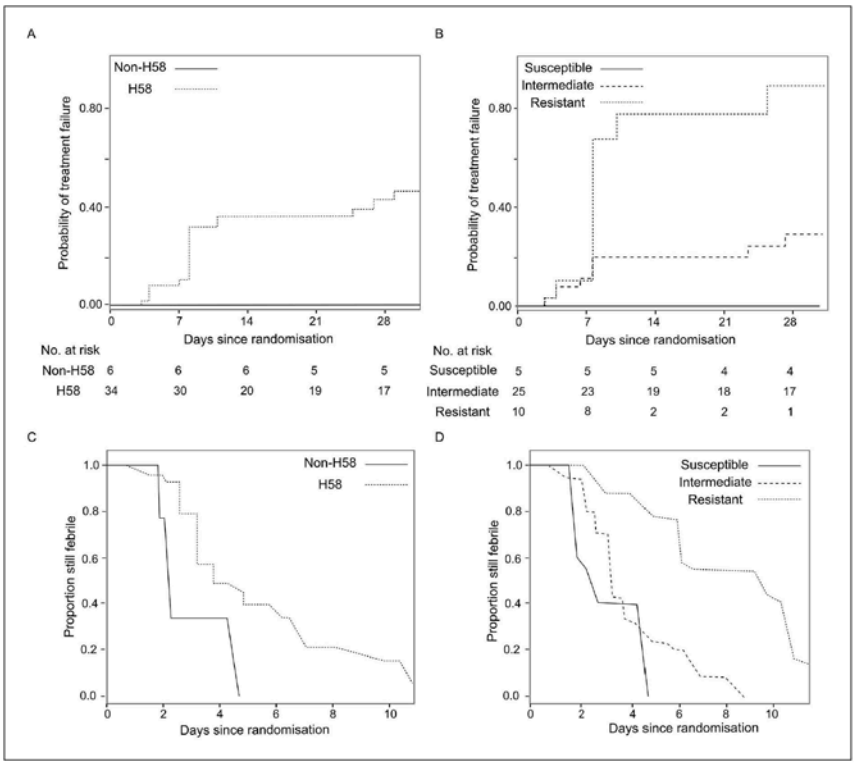

Figure 2. The association of *Salmonella* Typhi lineage and ciprofloxacin susceptibility with treatment failure and fever clearance time in patients randomised to gatifloxacin. (**A**) Kaplan-Meier curve for time to treatment failure by H58 and non-H58 *Salmonella* Typhi. (**B**) Kaplan-Meier curve for time to treatment failure by *Salmonella* Typhi susceptibility group (susceptible, intermediate, resistant to ciprofloxacin). (**C**) Non-parametric maximum likelihood estimators for interval-censored fever clearance time (see methods) by H58 and non-H58 *Salmonella* Typhi. (**D**) Non-parametric maximum likelihood estimators for interval-censored fever clearance time by *Salmonella* Typhi susceptibility group (susceptible, intermediate, resistant to ciprofloxacin).

DOI: 10.7554/eLife.14003.006

Our data suggest these FQ-resistant *S.* Typhi strains circulating in Nepal most likely descended from a single ancestor carrying the triple *gyrA/parC* mutant, such as that isolated in Nepal in 2011 (***Koirala et al., 2012***). This isolate was also associated with treatment failure, although this organism was not genome sequenced and was assumed to be an isolated case. More significantly, several very closely related strains were genome sequenced during an international study of H58 *S.* Typhi (***Wong et al., 2015***). These organisms had the same combination of triple FQ resistance mutations as those described here; our analysis shows they belong to the same subclade of H58. These strains had equivalently high MICs against ciprofloxacin and were isolated in India between 2008 and 2012. However, there were no associated patient outcome data for these strains and other reports from India have been limited. Our data implies that this lineage was introduced into Nepal from India or elsewhere in South Asia within the last 4–5 years and has subsequently entered in an endemic transmission cycle in Kathmandu. Given the large extent of human movement between India and Nepal, we propose this is the most likely route of introduction. However, there is also a small possibility

Table 3. Summary of fever clearance time by Salmonella Typhi lineage and ciprofloxacin susceptibility.

Fever clearance time	Gatifloxacin median (IQR) days	Ceftriaxone median (IQR) days	Acceleration factor (95%CI); p value	Heterogeneity test (p value)
H58[¥]				0.07
- H58	5.03 (3.18, 7.21)	3.07 (1.89 ,4.52)	1.59 (1.22, 2.09); p=0.0006	
- Non-H58	2.87 (2.08, 3.7)	3.12 (2.2, 4.12)	0.90 (0.59, 1.36); p=0.61	
Ciprofloxacin susceptibility group[‡]				0.015
- Susceptible	2.96 (2.13, 3.85)	4.78 (4.01, 5.5)	0.71 (0.49, 1.02); p=0.07	
- Intermediate	4.01 (2.76, 5.37)	2.63 (1.52, 4.05)	1.31 (0.97, 1.76); p=0.07	
- Resistant	8.2 (5.99, 10.5)	3.83 (2.96, 4.7)	2.23 (1.57, 3.17); p<0.0001	

[¥]p=0.013 and p=0.029 for comparison of interval censored time to fever clearance between H58 vs. non-H58 groups in gatifloxacin arm only and in all patients, respectively

[‡]p<0.0001 for comparison of interval censored time to fever clearance between MIC groups in gatifloxacin arm only

DOI: 10.7554/eLife.14003.007

that multiple strains independently gained resistance against FQs through the same selective pressure.

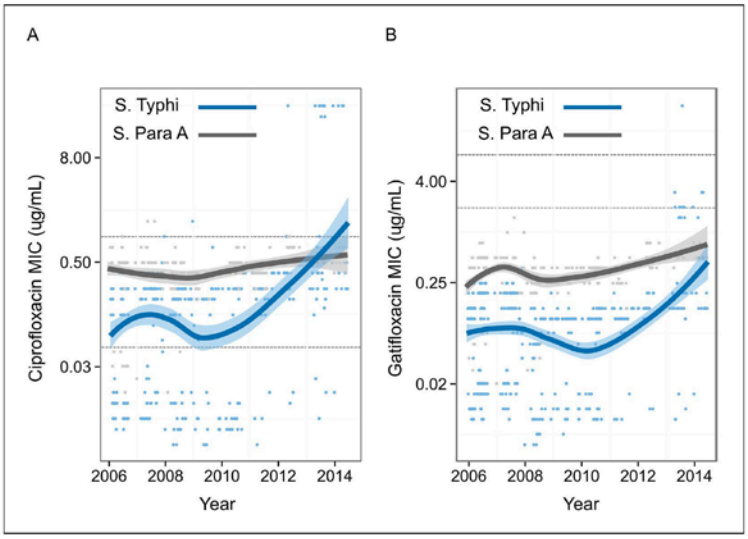

Figure 3. Minimum Inhibitory Concentrations of Nepali *Salmonella* Typhi and *Salmonella* Paratyphi against ciprofloxacin and gatifloxacin over ten years. Minimum Inhibitory Concentrations (µg/ml) for 568 Nepali *Salmonella* Typhi (blue) and 269 Nepali *Salmonella* Paratyphi A (grey) against (**A**) ciprofloxacin and (**B**) gatifloxacin collected from four randomised controlled trials conducted between 2005–2014 at Patan Hospital in Kathmandu, Nepal (***Pandit et al., 2007***; ***Koirala et al., 2013***; ***Arjyal et al., 2011***). The smoothed line derived from the generalized additive model showing a non-linear increase in Minimum Inhibitory Concentrations over time, with shading representing the 95% confidence interval. Lower and upper horizontal lines represent the current CLSI cut-offs for susceptible/intermediate and intermediate/resistant, respectively (***CLSI, 2012***).

DOI: 10.7554/eLife.14003.008

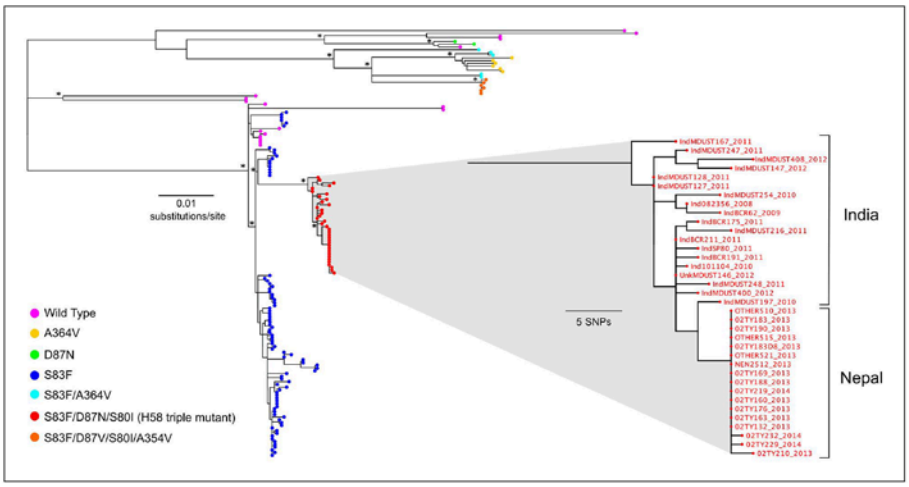

Figure 4. The phylogenetic structure of fluoroquinolone resistant *Salmonella* Typhi in a regional context. Maximum likelihood phylogeny based on core-genome SNPs of 136 (78 from the RCT) *Salmonella* Typhi isolates from Nepal and neighbouring India (***Supplementary file 1***). Main tree shows the overall phylogenetic structure and the presence of specific combinations of mutations in *gyrA* (S83F, D87V and D87N), *parC* (S80I) and *parE* (A364V). The inset shows a magnified view of the fluoroquinolone-resistant *Salmonella* Typhi H58 triple mutants from Nepal and their close association with similarly fluoroquinolone-resistant *Salmonella* Typhi H58 triple mutants from India (***Wong et al., 2015***). The scale bar on the primary tree indicates the number of substitutions per variable site, while that in the inset indicates genetic distance in number of SNPs (see methods). Asterisks indicate ≥85% bootstrap support at nodes of interest.

DOI: 10.7554/eLife.14003.009

Appropriate antimicrobial therapy is critical in the treatment of enteric fever, as effective drugs curtail symptoms and prevent life threatening complications. Our data has substantial repercussions for enteric fever treatment, and we advocate that FQs should no longer be used for empirical enteric fever therapy on the Indian subcontinent, as we predict these strains are likely to be widespread and are associated with poor outcomes with FQ therapy. Notably, in the RCT from which these data were derived we used the newer generation FQ, gatifloxacin, which binds to a different location on the DNA gyrase than the older FQs and is not as susceptible to the common resistance mutations (***Lu et al., 1999***). The isolates in this study were not generally resistant to gatifloxacin according to the current CLSI guidelines for Enterobacteriaceae (***CLSI, 2012***); we suggest that these guidelines be modified specifically for *S.* Typhi to reflect these new clinical data. We additionally propose that *S.* Typhi genotyping, mapping and susceptibility testing is performed routinely and rapidly in reference laboratories inside and outside of South Asia to monitor the international spread of these strains and ensure the provision of alternative efficacious therapies to returning travellers (***Lee et al., 2013***; ***García-Fernández et al., 2015***). In cases of infection with these FQ-resistant isolates, we suggest that ceftriaxone and azithromycin be used as alternatives, and do not currently recommend a return to the use of first-line drugs without contemporary data on treatment outcome. Whilst none of the isolates in this study were MDR, we predict a rapid return of MDR strains if there is a hasty return to older first-line alternatives.

This study has limitations. First, the clinical data was collected from one study in a single location, thus limiting utility outside this setting. Second, the overall sample size (and the gatifloxacin group subsampling) of those with culture-positive *S.* Typhi-associated enteric fever was relatively small, and the analysis presented here was performed in a *post hoc* manner. Notwithstanding these limitations, we were able to show a highly significant association between disease outcome and susceptibility

profile of the infecting organism. Further, by using WGS, we were able to pinpoint causative mutations, identify the subclade responsible for treatment failure and relate these strains to other isolates circulating outside Nepal in other parts of South Asia. The methodologies presented here, in which clinical outcome data are combined with genome sequences and antimicrobial susceptibility data, should become the gold standard for informing empiric treatment for all invasive bacterial infections and understanding the role of bacterial genotype and resistance profile on disease outcome for other bacterial infections. No other combination of methodologies would provide the granularity of data required to understand the epidemiology and clinical impact of this emergent strain in detail.

In conclusion, our data, for the first time, show a significant association between *S.* Typhi genotype, antimicrobial susceptibility and disease outcome for those treated with gatifloxacin in a cohort of Nepali enteric fever patients. A FQ-resistant variant of Typhi H58 has emerged in Nepal and is associated with the clinical failure of FQs. Our data suggest these isolates are likely widespread in the subcontinent and FQs should not be recommended for empirical enteric fever therapy in this setting.

Materials and methods

Study design and setting

The RCT from which the organisms and corresponding clinical data originated for these analyses was conducted at Patan Hospital and the Civil Hospital in the Lalitpur area of Kathmandu, Nepal, between 2011 and 2014, as described previously (***Arjyal et al., 2016***). The trial was registered at www.clinicaltrials.gov (ISRCTN63006567). Briefly, patients were randomly assigned to seven days of treatment with either oral gatifloxacin (400 mg tablets, Square Pharmaceuticals Limited, Bangladesh) at a dose of 10 mg/kg once daily or intravenous ceftriaxone (Powercef, 1000mg injection vial, Wockhardt Ltd, India), injected over 10 min at a dose of 60 mg/kg up to a maximum of two grams (aged 2 to 13 years) or two grams (≥14 years) once daily.

A detailed description of the RCT from which these data were generated has been previously published (***Arjyal et al., 2016***). The primary endpoint was a composite of treatment failure, defined as the occurrence of at least one of the following events: fever clearance time (FCT) (time from the first dose of a study drug until the temperature dropped to ≤37·5°C and remained there for at least two days) more than seven days post-treatment initiation; requirement for rescue treatment as judged by the treating physician; blood culture positivity for *S.* Typhi or *S.* Paratyphi on day eight of treatment (microbiological failure); culture-confirmed or syndromic enteric fever relapse within 28 days of initiation of treatment; and the development of any enteric fever-related complication (e.g. clinically significant bleeding, fall in the Glasgow Coma Score, perforation of the gastrointestinal tract and hospital admission) within 28 days after the initiation of treatment. Time to treatment failure was defined as the time from the first dose of treatment until the date of the earliest failure event. FCTs were calculated electronically using twice-daily recorded temperatures and treated as interval-censored outcomes. Patients without fever clearance or relapse, respectively, were censored at the time of their last follow-up visit (additional details regarding study procedures can be found in Arjyal *at al.* 2016 (***Arjyal et al., 2016***)).

Blood (3 ml if aged <14 years; 8 ml if aged ≥14 years) was taken from all patients for bacterial culture on enrolment. Adult blood samples were inoculated into media containing tryptone soya broth and sodium polyanethol sulphonate, up to a total volume of 50 mL. Bactec Peds Plus culture bottles (Becton Dickinson, New Jersey, USA) were used for paediatric blood samples. Culture results were reported for up to seven days, positive bottles were subcultured onto blood, chocolate and MacConkey agar and presumptive *Salmonella* colonies were identified using standard biochemical tests and serotype-specific antisera (Murex Biotech, Dartford, England). Antimicrobial susceptibility testing was performed by the modified Bauer-Kirby disc diffusion method with zone size interpretation based on CLSI guidelines (***CLSI, 2012***). Etests were used to determine MICs, following the manufacturer's recommendations (bioMérieux, France). Ciprofloxacin MICs were used to categorise *S.* Typhi isolates as susceptible (≤0.06 µg/mL), intermediate (0.12–0.5 µg/mL) and resistant (≥1 µg/mL) following CLSI guidelines (***CLSI, 2012***).

Whole genome sequencing and analysis

Genomic DNA from Nepali *S.* Typhi organisms originating from this RCT (78 isolates) was extracted using the Wizard Genomic DNA Extraction Kit (Promega, Wisconsin, USA) (***Supplementary file 1***) (***Karkey et al., 2013***). Two µg of genomic DNA was subjected to WGS on an Illumina Miseq platform, following the manufacturer's recommendations to generate 250bp/100bp paired-end reads. All reads were mapped to the reference sequence of *S.* Typhi CT18 (accession no: AL515582) using SMALT (version 0.7.4). Candidate single nucleotide polymorphisms (SNPs) were called against the reference sequence using SAMtools (***Li et al., 2009***) and filtered with a minimal phred quality of 30 and a quality cut-off of 0.75. The allele at each locus in each isolate was determined by reference to the consensus base in that genome, using *samtools mpileup* and removing low confidence alleles with consensus base quality $\leq$20, read depth $\leq$5 or a heterozygous base call. SNPs called in phage regions, repetitive sequences or recombinant regions were excluded, (***Wong et al., 2015***) resulting in a final set of 1,607 chromosomal SNPs. Strains belonging to haplotype H58 were defined by the SNP *glpA*-C1047T (position 2348902 in *S.* Typhi CT18, BiP33) (***Emary et al., 2012***; ***Holt et al., 2008***; ***Parkhill et al., 2001***).

A maximum likelihood (ML) phylogeny was estimated using a 1440 SNP alignment of the 78 RCT isolates in RAxML (version 7.8.6) with the generalized time-reversible substitution model (GTR) and a gamma distribution, with support for the phylogeny assessed via 1000 bootstrap replicates. The alignment was then compared to a global *S.* Typhi sequence database, with a particular focus on identifying sequences with a mutational profile suggestive of shared ancestry with a divergent H58 clade identified in the previous phylogeny. A secondary ML phylogenetic tree was then inferred from the SNP alignment of the 136 Nepali Typhi along with 19 recently described Typhi H58 with the aforementioned mutational profile, using the same parameters as above (1642 SNPs; ***Supplementary file 1***) (***Wong et al., 2015***). Raw sequence data are available in the European Nucleotide Archive (ENA) (***Supplementary file 1***).

Statistical analysis

Comparison of baseline characteristics within patient groups, stratified by the H58 status or susceptibility category of their corresponding *S.* Typhi isolates was performed using the Kruskal Wallis test for continuous variables and Fisher's exact test for categorical variables. Time to treatment failure was analysed using Firth's penalized maximum likelihood bias reduction method for Cox regression as a solution for the non-convergence of likelihood function in the case of zero event counts in subgroups (***Firth, 1993***). For comparisons between treatment arms, H58 status, or ciprofloxacin susceptibility group, the model included treatment arm, H58 status, or susceptibility group as a single covariate. Confidence intervals (CI) and *p*-values were calculated by profile-penalized likelihood. FCT was analysed as an interval-censored outcome, i.e. as the time interval from the last febrile temperature assessment until the first afebrile assessment, using parametric Weibull accelerated failure time models (***Kalbfleisch and Prentice, 2002***). Median and inter-quartile range (IQR) FCT calculations for subgroups were based on models for each subgroup separately. Acceleration factors were based on models that included treatment arm as the only covariate. The non-parametric maximum likelihood estimator (NPMLE) was used to visualize the distribution of FCT between groups. Heterogeneity between subgroups was tested with models that included an interaction between treatment arm and the sub-grouping variable. To study the emergence of FQ resistance, data from previous enteric fever trials from 2005–2014 (***Pandit et al., 2007***; ***Koirala et al., 2013***; ***Arjyal et al., 2011***) was pooled and generalized additive models (GAM) were used to examine potential non-linear trends of ciprofloxacin and gatifloxacin MICs over time. All analyses were performed using R software version 3.2.2 (***Team, 2012***).

Acknowledgements

This project was funded by the Wellcome Trust of Great Britain (106158/Z/14/Z). SB is a Sir Henry Dale Fellow, jointly funded by the Wellcome Trust and the Royal Society (100087/Z/12/Z). KEH is supported by fellowship #1061409 from the NHMRC of Australia. CD was funded by the Li Ka Shing Foundation Global Health Programme at the University of Oxford. DTP and AK are funded as leadership fellows through the Oak Foundation. The funders had no role in study design, data collection

and analysis, decision to publish, or preparation of the manuscript. We wish to acknowledge Mangal Rawal, Sumi Munankarmi, Bibek Karki, Radheshyam KC, Sudeep Dhoj Thapa, Rabin Gautam, Priyanka Tiwari, Manisha Risal, Surendra Shrestha, Balmukunda Neupane, Nabraj Regmi, Krishna Prajapati, Bimal Thapa and the trial monitors Nguyen Thi Phuong Dung and Nguyen Thi Thanh Thuy for their assistance in conducting this trial.

Additional information

Funding

Funder	Grant reference number	Author
Wellcome Trust	100087/Z/12/Z	Stephen Baker
Royal Society	100087/Z/12/Z	Stephen Baker
National Health and Medical Research Council	1061409	Kathryn E Holt
Li Ka Shing Foundation		Christiane Dolecek
Wellcome Trust	106158/Z/14/Z	Kathryn E Holt Gordon Dougan Christiane Dolecek Buddha Basnyat Stephen Baker
The Oak Foundation	OCAY-15-547	Duy Pham Thanh Abhilasha Karkey Stephen Baker

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

DPT, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article; AK, SD, VW, NTVT, PVV, THT, DP, Acquisition of data, Contributed unpublished essential data or reagents; NHT, CNT, MAR, KEH, Analysis and interpretation of data, Drafting or revising the article; AA, BB, Conception and design, Acquisition of data; AP, SKS, DG, Recruited patients into study and collected clinical data, Acquisition of data; CMP, GD, GET, Conception and design, Drafting or revising the article; MW, Analysis and interpretation of data, Contributed unpublished essential data or reagents; CD, Conception and design, Acquisition of data, Contributed unpublished essential data or reagents; SB, Conception and design, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article

Author ORCIDs

Maia A Rabaa, http://orcid.org/0000-0003-0529-2228

Ethics

Clinical trial Registry: ISRCTN. Registration ID: ISRCTN63006567.

Human subjects: This study was performed following the principles of the declaration of Helsinki. Written informed consent to participate in all studies from Nepal contributing data for this analysis was required from all patients. For those aged <18 years, written informed consent was obtained from a parent or an adult guardian. The protocol was reviewed and approved by the Ethics Committee of the Nepal Health Research Council (NHRC) and the Oxford Tropical Research Ethics Committee (OxTREC) UK.

Additional files

Supplementary files

• Supplementary file 1. Table of *Salmonella* Typhi isolates and their corresponding sequencing metadata used in this study.

DOI: 10.7554/eLife.14003.010

• Supplementary file 2. (A) Table of baseline characteristics by *Salmonella* Typhi lineage. (B) Table of baseline characteristics grouped by *Salmonella* Typhi ciprofloxacin susceptibility. (C) Table of treatment failure in detail by *Salmonella* Typhi lineage in the gatifloxacin treatment group. (D) Table of treatment failure in detail by ciprofloxacin susceptibility in the gatifloxacin treatment group.
DOI: 10.7554/eLife.14003.011

Major datasets

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database, license, and accessibility information
Duy Pham Thanh, Stephen Baker, Kathryn E Holt, Gordon Dougan, Vanessa Wong	2015	Sequence data	http://www.ebi.ac.uk/ena/data/view/PRJEB10959	PRJEB10959

References

Arjyal A, Basnyat B, Koirala S, Karkey A, Dongol S, Agrawaal KK, Shakya N, Shrestha K, Sharma M, Lama S, Shrestha K, Khatri NS, Shrestha U, Campbell JI, Baker S, Farrar J, Wolbers M, Dolecek C. 2011. Gatifloxacin versus chloramphenicol for uncomplicated enteric fever: an open-label, randomised, controlled trial. *The Lancet Infectious Diseases* **11**:445–454. doi: 10.1016/S1473-3099(11)70089-5

Arjyal A, Basnyat B, Nhan HT, Koirala S, Giri A, Joshi N, Shakya M, Pathak KR, Mahat SP, Prajapati SP, Adhikari N, Thapa R, Merson L, Gajurel D, Lamsal K, Lamsal D, Yadav BK, Shah G, Shrestha P, Dongol S, Karkey A, Thompson CN, Thieu NTV, Thanh DP, Baker S, Thwaites GE, Wolbers M, Dolecek C. 2016. Gatifloxacin versus ceftriaxone for uncomplicated enteric fever in nepal: an open-label, two-centre, randomised controlled trial. *The Lancet Infectious Diseases*. doi: 10.1016/S1473-3099(15)00530-7

Baker S, Duy PT, Nga TVT, Dung TTN, Phat VV, Chau TT, Turner AK, Farrar J, Boni MF. 2013. Fitness benefits in fluoroquinolone-resistant *salmonella* typhi in the absence of antimicrobial pressure. *eLife* **2**:e01229–e01229. doi: 10.7554/eLife.01229

CLSI. 2012. Clinical and laboratory standards institute: performance standards for antimicrobial susceptibility testing. *Twentieth Informational Supplement*:M100–S22.

Crump JA, Mintz ED. 2010. Global trends in typhoid and paratyphoid fever. *Clinical Infectious Diseases* **50**:241–246. doi: 10.1086/649541

Emary K, Moore CE, Chanpheaktra N, An KP, Chheng K, Sona S, Duy PT, Nga TVT, Wuthiekanun V, Amornchai P, Kumar V, Wijedoru L, Stoesser NE, Carter MJ, Baker S, Day NPJ, Parry CM. 2012. Enteric fever in cambodian children is dominated by multidrug-resistant H58 salmonella enterica serovar typhi with intermediate susceptibility to ciprofloxacin. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **106**:718–724. doi: 10.1016/j.trstmh.2012.08.007

Firth D. 1993. Bias reduction of maximum likelihood estimates. *Biometrika* **80**:27–38 Available: http://biomet.oxfordjournals.org/content/80/1/27.abstract . Accessed 24 June 2015. doi: 10.1093/biomet/80.1.27

García-Fernández A, Gallina S, Owczarek S, Dionisi AM, Benedetti I, Decastelli L, Luzzi I. 2015. Emergence of ciprofloxacin-resistant salmonella enterica serovar typhi in italy. *PLOS ONE* **10**:e0132065. doi: 10.1371/journal.pone.0132065

Hendriksen RS, Leekitcharoenphon P, Mikoleit M, Jensen JD, Kaas RS, Roer L, Joshi HB, Pornruangmong S, Pulsrikarn C, Gonzalez-Aviles GD, Reuland EA, Al Naiemi N, Wester AL, Aarestrup FM, Hasman H. 2015. 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? *Journal of Clinical Microbiology* **53**:677–680. doi: 10.1128/JCM.03104-14

Holt KE, Parkhill J, Mazzoni CJ, Roumagnac P, Weill F-X, Goodhead I, Rance R, Baker S, Maskell DJ, Wain J, Dolecek C, Achtman M, Dougan G. 2008. High-throughput sequencing provides insights into genome variation and evolution in salmonella typhi. *Nature Genetics* **40**:987–993. doi: 10.1038/ng.195

Kalbfleisch JD, Prentice RL. 2002. *The Statistical Analysis of Failure Time DataJohn Wiley & Sons*. 2nd ed Hoboken, NJ, USA: John Wiley & Sons, Inc. doi: 10.1002/9781118032985

Kariuki S, Gordon MA, Feasey N, Parry CM. 2015. Antimicrobial resistance and management of invasive salmonella disease. *Vaccine* **33**:C21–C29. doi: 10.1016/j.vaccine.2015.03.102

Kariuki S, Revathi G, Kiiru J, Mengo DM, Mwituria J, Muyodi J, Munyalo A, Teo YY, Holt KE, Kingsley RA, Dougan G. 2010. Typhoid in kenya is associated with a dominant multidrug-resistant salmonella enterica serovar typhi haplotype that is also widespread in southeast asia. *Journal of Clinical Microbiology* **48**:2171–2176. doi: 10.1128/JCM.01983-09

Karkey A, Thompson CN, Tran Vu Thieu N, Dongol S, Le Thi Phuong T, Voong Vinh P, Arjyal A, Martin LB, Rondini S, Farrar JJ, Dolecek C, Basnyat B, Baker S. 2013. Differential epidemiology of salmonella typhi and

paratyphi a in kathmandu, nepal: a matched case control investigation in a highly endemic enteric fever setting. *PLoS Neglected Tropical Diseases* **7**:e2391. doi: 10.1371/journal.pntd.0002391

Koirala KD, Thanh DP, Thapa SD, Arjyal A, Karkey A, Dongol S, Shrestha UM, Farrar JJ, Basnyat B, Baker S. 2012. Highly resistant salmonella enterica serovar typhi with a novel gyrA mutation raises questions about the long-term efficacy of older fluoroquinolones for treating typhoid fever. *Antimicrobial Agents and Chemotherapy* **56**:2761–2762. doi: 10.1128/AAC.06414-11

Koirala S, Basnyat B, Arjyal A, Shilpakar O, Shrestha K, Shrestha R, Shrestha UM, Agrawal K, Koirala KD, Thapa SD, Karkey A, Dongol S, Giri A, Shakya M, Pathak KR, Campbell J, Baker S, Farrar J, Wolbers M, Dolecek C. 2013. Gatifloxacin versus ofloxacin for the treatment of uncomplicated enteric fever in nepal: an open-label, randomized, controlled trial. *PLoS Neglected Tropical Diseases* **7**:e2523. doi: 10.1371/journal.pntd.0002523

Lee C-J, Su L-H, Huang Y-C, Chiu C-H. 2013. First isolation of ciprofloxacin-resistant salmonella enterica serovar typhi in taiwan. *Journal of Microbiology, Immunology and Infection* **46**:469–473. doi: 10.1016/j.jmii.2013.01.002

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R.1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079. doi: 10.1093/bioinformatics/btp352

Lu T, Zhao X, Drlica K. 1999. Gatifloxacin activity against quinolone-resistant gyrase: allele-specific enhancement of bacteriostatic and bactericidal activities by the C-8-methoxy group. *Antimicrobial Agents and Chemotherapy* **43**:2969–2974.

Olarte J, Galindo E. 1973. Salmonella typhi resistant to chloramphenicol, ampicillin, and other antimicrobial agents: strains isolated during an extensive typhoid fever epidemic in mexico. *Antimicrobial Agents and Chemotherapy* **4**:597–601. doi: 10.1128/AAC.4.6.597

Pandit A, Arjyal A, Day JN, Paudyal B, Dangol S, Zimmerman MD, Yadav B, Stepniewska K, Campbell JI, Dolecek C, Farrar JJ, Basnyat B. 2007. An open randomized comparison of gatifloxacin versus cefixime for the treatment of uncomplicated enteric fever. *PLoS ONE* **2**:e542. doi: 10.1371/journal.pone.0000542

Parkhill J, Dougan G, James KD, Thomson NR, Pickard D, Wain J, Churcher C, Mungall KL, Bentley SD, Holden MTG, Sebaihia M, Baker S, Basham D, Brooks K, Chillingworth T, Connerton P, Cronin A, Davis P, Davies RM, Dowd L, White N, Farrar J, Feltwell T, Hamlin N, Haque A, Hien TT, Holroyd S, Jagels K, Krogh A, Larsen TS, Leather S, Moule S, Ó'Gaora P, Parry C, Quail M, Rutherford K, Simmonds M, Skelton J, Stevens K, Whitehead S, Barrell BG. 2001. Complete genome sequence of a multiple drug resistant salmonella enterica serovar typhi CT18. *Nature* **413**:848–852. doi: 10.1038/35101607

Parry CM, Hien TT, Dougan G, White NJ, Farrar JJ. 2002. Typhoid fever. *New England Journal of Medicine* **347**: 1770–1782. doi: 10.1056/NEJMra020201

Parry CM, Thuy CT, Dongol S, Karkey A, Vinh H, Chinh NT, Duy PT, Thieu Nga TV, Campbell JI, Van Minh Hoang N, Arjyal A, Bhutta ZA, Bhattacharya SK, Agtini MD, Dong B, Canh DG, Naheed A, Wain J, Tinh Hien T, Basnyat B, Ochiai L, Clemens J, Farrar JJ, Dolecek C, Baker S. 2010. Suitable disk antimicrobial susceptibility breakpoints defining salmonella enterica serovar typhi isolates with reduced susceptibility to fluoroquinolones. *Antimicrobial Agents and Chemotherapy* **54**:5201–5208. doi: 10.1128/AAC.00963-10

Parry CM, Vinh H, Chinh NT, Wain J, Campbell JI, Hien TT, Farrar JJ, Baker S. 2011. The influence of reduced susceptibility to fluoroquinolones in salmonella enterica serovar typhi on the clinical response to ofloxacin therapy. *PLoS Neglected Tropical Diseases* **5**:e1163. doi: 10.1371/journal.pntd.0001163

Roumagnac P, Weill F-X, Dolecek C, Baker S, Brisse S, Chinh NT, Le TAH, Acosta CJ, Farrar J, Dougan G, Achtman M. 2006. Evolutionary history of salmonella typhi. *Science* **314**:1301–1304. doi: 10.1126/science.1134933

Team R. 2012. R: a language and environment for statistical computing. *R Foundation for Statistical Computing*. Vienna, Austria: Open access available http//cran r-project org.

The World Health Organization. 2003. Background document: the diagnosis, treatment and management of typohoid fever. http://www.who.int/rpc/TFGuideWHO.pdf.

Wain J, Diem Nga LT, Kidgell C, James K, Fortune S, Song Diep T, Ali T, O Gaora P, Parry C, Parkhill J, Farrar J, White NJ, Dougan G. 2003. Molecular analysis of incHI1 antimicrobial resistance plasmids from salmonella serovar typhi strains associated with typhoid fever. *Antimicrobial Agents and Chemotherapy* **47**:2732–2739. doi: 10.1128/AAC.47.9.2732-2739.2003

Wong VK, Baker S, Pickard DJ, Parkhill J, Page AJ, Feasey NA, Kingsley RA, Thomson NR, Keane JA, Weill F-X, Edwards DJ, Hawkey J, Harris SR, Mather AE, Cain AK, Hadfield J, Hart PJ, Thieu NTV, Klemm EJ, Glinos DA, Breiman RF, Watson CH, Kariuki S, Gordon MA, Heyderman RS, Okoro C, Jacobs J, Lunguya O, Edmunds WJ, Msefula C, Chabalgoity JA, Kama M, Jenkins K, Dutta S, Marks F, Campos J, Thompson C, Obaro S, MacLennan CA, Dolecek C, Keddy KH, Smith AM, Parry CM, Karkey A, Mulholland EK, Campbell JI, Dongol S, Basnyat B, Dufour M, Bandaranayake D, Naseri TT, Singh SP, Hatta M, Newton P, Onsare RS, Isaia L, Dance D, Davong V, Thwaites G, Wijedoru L, Crump JA, De Pinna E, Nair S, Nilles EJ, Thanh DP, Turner P, Soeng S, Valcanis M, Powling J, Dimovski K, Hogg G, Farrar J, Holt KE, Dougan G. 2015. Phylogeographical analysis of the dominant multidrug-resistant H58 clade of salmonella typhi identifies inter- and intracontinental transmission events. *Nature Genetics* **47**:632–639. doi: 10.1038/ng.3281

PLOS | NEGLECTED TROPICAL DISEASES

RESEARCH ARTICLE

The Molecular and Spatial Epidemiology of Typhoid Fever in Rural Cambodia

Duy Pham Thanh[1☯], Corinne N. Thompson[1,2,3☯], Maia A Rabaa[1,2], Soeng Sona[4], Sun Sopheary[4], Varun Kumar[4], Catrin Moore[4], Nga Tran Vu Thieu[1], Lalith Wijedoru[4], Kathryn E. Holt[5], Vanessa Wong[6,7], Derek Pickard[6], Guy E. Thwaites[1,2], Nicholas Day[2,4], Gordon Dougan[6,7], Paul Turner[2,4‡], Christopher M. Parry[3,4,8‡], Stephen Baker[1,2,3‡*]

1 The Hospital for Tropical Diseases, Wellcome Trust Major Overseas Programme, Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam, **2** Centre for Tropical Medicine, Oxford University, Oxford, United Kingdom, **3** The London School of Hygiene and Tropical Medicine, London, United Kingdom, **4** Cambodia-Oxford Medical Research Unit, Angkor Hospital for Children, Siem Reap, Cambodia, **5** Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria, Australia, **6** The Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United Kingdom, **7** The Department of Medicine, University of Cambridge, Cambridge, United Kingdom, **8** School of Tropical Medicine and Global Health, Nagasaki University, Nagasaki, Japan

☯ These authors contributed equally to this work.
‡ PT, CMP and SB are joint senior authors.
* sbaker@oucru.org

OPEN ACCESS

Citation: Pham Thanh D, Thompson CN, Rabaa MA, Sona S, Sopheary S, Kumar V, et al. (2016) The Molecular and Spatial Epidemiology of Typhoid Fever in Rural Cambodia. PLoS Negl Trop Dis 10(6): e0004785. doi:10.1371/journal.pntd.0004785

Editor: Edward T. Ryan, Massachusetts General Hospital, UNITED STATES

Received: March 18, 2016

Accepted: May 27, 2016

Published: June 22, 2016

Data Availability Statement: The raw sequence data for this study are available in the European Nucleotide Archive (ENA) under the accession numbers described in S1 Table in the manuscript.

Funding: This project was funded by the Wellcome Trust of Great Britain (106158/Z/14/Z). SB is a Sir Henry Dale Fellow, jointly funded by the Wellcome Trust and the Royal Society (100087/Z/12/Z). KEH is supported by fellowship #1061409 from the NHMRC of Australia. DPT is funded as a leadership fellow through the Oak Foundation. The bacterial surveillance at Angkor Hopsital for Children was funded by the Wellcome Trust and the Li Ka Shing-

Abstract

Typhoid fever, caused by the bacterium *Salmonella* Typhi, is an endemic cause of febrile disease in Cambodia. The aim of this study was to better understand the epidemiology of pediatric typhoid fever in Cambodia. We accessed routine blood culture data from Angkor Hospital for Children (AHC) in Siem Reap province between 2007 and 2014, and performed whole genome sequencing (WGS) on the isolated bacteria to characterize the *S*. Typhi population. The resulting phylogenetic information was combined with conventional epidemiological approaches to investigate the spatiotemporal distribution of *S*. Typhi and population-level risk factors for reported disease. During the study period, there were 262 cases of typhoid within a 100 km radius of AHC, with a median patient age of 8.2 years (IQR: 5.1–11.5 years). The majority of infections occurred during the rainy season, and commune incidences as high as 11.36/1,000 in children aged <15 years were observed over the study period. A population-based risk factor analysis found that access to water within households and increasing distance from Tonle Sap Lake were protective. Spatial mapping and WGS provided additional resolution for these findings, and confirmed that proximity to the lake was associated with discrete spatiotemporal disease clusters. We confirmed the dominance of MDR H58 *S*. Typhi in this population, and found substantial evidence of diversification (at least seven sublineages) within this single lineage. We conclude that there is a substantial burden of pediatric typhoid fever in rural communes in Cambodia. Our data provide a platform for additional population-based typhoid fever studies in this location, and suggest that this would be a suitable setting in which to introduce a school-based vaccination programme with Vi conjugate vaccines.

University of Oxford Global Health Programme. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Author Summary

Typhoid fever is an infectious disease caused by the bacterium *Salmonella* Typhi. The disease is generally restricted to those living in low-income settings with poor sanitation. Typhoid fever is a common cause of fever requiring hospital treatment in Cambodia, but limited data is available on the epidemiology of the disease. To better understand typhoid fever in Cambodia, we accessed routine hospital data for typhoid fever from a single healthcare facility treating sick children in Siem Reap in the central Cambodia between 2007 and 2014. We mapped the location of these cases and examined population-based risk factors for reported disease. Additionally, we decoded the genomes of the *S.* Typhi isolated from children attending the hospital to understand how the organism has evolved and spread throughout the population. We found a large burden of typhoid fever in children in this largely rural setting in central Cambodia. We also found that disease was associated with the rainy season and that living close to Tonle Sap Lake increased the risk of disease. The genomes of the sequenced bacteria showed that a diverse range of strains were circulating during the study, and allowed us to identify signatures of location- and time-specific outbreaks. Our work provides baseline data for additional typhoid fever studies in the population living in this location and findings suggest that rural Cambodia would be a suitable setting in which to introduce a school-based vaccination program with new typhoid vaccines.

Introduction

The bacterium *Salmonella enterica* serovar Typhi (*S.* Typhi) is the cause of the human infection typhoid fever, a systemic disease predominantly diagnosed in children and young adults in low-income settings [1]. *S.* Typhi is primarily contracted via ingestion of food or water contaminated with human feces from patients excreting the organism, and typhoid fever remains a major public health issue in areas with poor sanitation and limited access to safe water [2]. The control of typhoid fever is largely dependent on improving the availability of clean water, hygienic food preparation and access to adequate sanitation, but such interventions are substantial challenges in many locations where typhoid remains endemic [3]. As a result, active case detection and appropriate antimicrobial therapy are currently the principal methods for controlling this disease in endemic locations. The lack of rapid and reliable diagnostics and the emergence of antimicrobial resistance (AMR) reduce the effectiveness of these strategies [4,5].

The World Health Organization (WHO) currently recommends the use of licensed typhoid vaccines in areas where the burden of typhoid fever is high and AMR organisms are prevalent, though limited programmatic use of vaccines has occurred in endemic countries [6]. Identifying areas that have a burden of typhoid fever warranting immunization programmes can be challenging. Much of the current focus for vaccine implementation is on highly populated urban slums with poor infrastructure [7–9]. However, it is unclear if the epidemiology of typhoid is similar between urban and rural regions in the developing world. Understanding the dominant modes of transmission and epidemiological risk factors for typhoid fever in both urban and rural endemic regions is vital for controlling and preventing the disease [3].

New molecular tools now permit an unprecedented insight into how *S.* Typhi may be circulating locally and internationally [10,11,7]. Comparing the composition of phylogenetically informative Single Nucleotide Polymorphisms (SNPs) across the *S.* Typhi genome allows the subtyping of populations of *S.* Typhi and inference of evolutionary relationships between isolates [12,13]. SNP-based typing methods, and now whole genome sequencing (WGS), have

been successfully used to study the molecular epidemiology of typhoid in different settings, revealing the importance of environmental transmission and the diversity of commonly co-circulating haplotypes of *S.* Typhi within localized human populations [7,14–16]. Such molecular approaches play an important role in identifying dominant transmission pathways and can also capture both the emergence of AMR and the dynamics of the bacterial population. We now know that the current population of *S.* Typhi has been driven by a clonal expansion and international dispersal of a specific haplotype (H58) in Asia and Africa. This H58 haplotype now dominates internationally and is associated with a multidrug resistant (MDR) phenotype (non-susceptibility to ampicillin, chloramphenicol and trimethoprim-sulphamethoxazole) and reduced susceptibility to fluoroquinolones [11].

Typhoid fever is endemic in Cambodia, although only limited data regarding the morbidity, mortality and risk factors for the disease are available in published literature. A community-based study conducted near Phnom Penh (the capital city) between 2006 and 2009 reported routine isolation of MDR *S.* Typhi from the blood of febrile patients [17] Additionally, a hospital-based cross-sectional study of pediatric bloodstream infections in a children's hospital in Siem Reap (in northwest Cambodia, near the World Heritage Site of Angkor Wat) between 2007 and 2011 found that *S.* Typhi was the most commonly isolated pathogen in this setting and confirmed the presence and dominance of H58 *S.* Typhi (98/102; 96% of *S.* Typhi isolates) exhibiting reduced susceptibility to ciprofloxacin [18,19]. In the present study, we aimed to utilize the precision of WGS to characterize the H58 *S.* Typhi population in Siem Reap, Cambodia. Further, we combined the resulting phylogenetic information with additional epidemiological approaches to investigate the spatiotemporal distribution of *S.* Typhi and population-level risk factors for typhoid fever infection in this location.

Methods

Ethics statement

The study involved characterization of stored bacterial isolates cultured from specimens taken for routine clinical care. Therefore, it was not possible to obtain consent from the patient or their parent/guardian for participation in this retrospective study, but all patient data was anonymized. The study protocol was reviewed and approved by both the Angkor Hospital for Children Institutional Review Board (AHC IRB; reference 423/13) and the Oxford Tropical Research Ethics Committee (OxTREC; reference 512–13).

Study site and setting

This study was conducted at Angkor Hospital for Children (AHC) in Siem Reap City in Cambodia between January 2007 and December 2014. AHC is one of two pediatric hospitals in Siem Reap City and has approximately 125,000 attendees and 4,000 admissions per year. The patients attending AHC are <16 years of age and come from a wide geographical radius and attend the hospital for various conditions. The majority of patients reside in the province of Siem Reap, which is located in northwest Cambodia and is bordered in the south by the Tonle Sap Lake, the largest freshwater lake in Southeast Asia. According to available census data, the province had a population of 896,443 people living in an area of 10,299 km^2 in 2008; the province is subdivided administratively into 12 districts, 100 communes (which are within districts) and 907 villages. [20]. Cambodia has a tropical climate with a dry and wet season each year. During the wet season (April–October) the area of the Tonle Sap Lake can expand dramatically, increasing from 3,500 km^2 up to approximately 14,500 km^2, with the depth increasing from 0.5 m up to 6–9 m [21].

Definition of the case population and the control population

Case and control populations were identified from the electronic hospital and laboratory information system of AHC. For the purposes of this study, the case population was defined as the population of hospital inpatients from whom *S.* Typhi was isolated from a blood culture. The control population was defined as the patient population admitted to AHC who did not have typhoid fever based on the recorded discharge diagnosis (International Classification of Disease (ICD)-10 code). Patients with a discharge diagnosis of typhoid fever but without blood culture confirmation (n = 410) were not included in the risk factor analysis. Additionally, for the mapping and population risk factor analyses, cases that lived outside of a 100km radius from AHC were excluded. Data on age, sex, home location (commune level), admission and discharge dates for cases and controls were extracted from the electronic hospital information system. If a case or control was readmitted to the hospital with the same discharge diagnosis within a seven-day period, only the initial admission was included in the analysis.

Data sources

Commune-level census data were obtained from the Cambodian National Report on General Population Census of 2008 [20]. The extracted information included details regarding demographic indicators, age structure, literacy and education, housing and household characteristics, and access to toilet facilities and drinking water. Based on this report, a commune was classified as urban if the population density exceeded 200/km^2, less than half of men were employed in agriculture and the total population exceeded 2,000. Monthly average precipitation was collected from Siem Reap Weather Station and MRCS (Mekong River Commission Secretariat) [22]. Shuttle Radar Topography Mission (SRTM) elevation data were obtained from the CGIAR Consortium for Spatial Information (CGIAR-CSI) [23]. Shapefile layers containing 2008 commune-level population census data were accessed from Open Development Cambodia, an open-access data website providing data on Cambodia and its economic and social development (http://www.opendevelopmentcambodia.net).

Typhoid diagnosis and bacterial identification

Routine diagnosis of typhoid fever was performed by blood culture. Blood (1–4 ml) was taken for bacterial culture from all patients with fever including those with a clinical suspicion of typhoid fever. Blood was inoculated into media containing tryptone soya broth and sodium polyanethol sulphonate, up to a total volume of 25mL. Blood culture bottles were incubated for up to seven days, with blind sub-cultures at 24 hours, 48 hours and 7 days or if the broth was cloudy. Positive bottles were subcultured onto sheep blood, chocolate and MacConkey agar and presumptive *Salmonella* colonies were identified using standard biochemical tests and serotype-specific antisera (Murex Biotech, Dartford, England). Antimicrobial susceptibility testing was performed by the modified Bauer-Kirby disc diffusion method with zone size interpretation based on CLSI guidelines [24]. Etests were used to determine MICs, following the manufacturer's recommendations (bioMérieux, France). Ciprofloxacin MICs were used to categorize *S.* Typhi isolates as susceptible ($\leq$0.06 μg/mL), intermediate (0.12–0.5 μg/mL) and resistant ($\geq$1 μg/mL) following CLSI guidelines [24].

Statistical analysis

Rates of hospitalized typhoid fever were calculated at the commune level using the population under the age of 15 years from 2008. Multivariable negative binomial regression was used to identify commune-level risk factors associated with the rate of cases per 1,000 population

under the age of 15 years. Interaction between commune level factors was evaluated using the likelihood ratio test. Variables included in the evaluation of the final model included those with significant associations ($p<0.10$) in the univariate analysis and *a priori* sanitation and water source variables. Variables that did not add significantly to the fit of the final model (determined by the likelihood ratio test) were not included. All analyses were performed in STATA (v13, College Station, TX, USA) and plots were created in R v3.1.1 (R Foundation for Statistical Computing, Vienna, Austria, https://cran.r-project.org/) using ggplot2 [25].

Spatiotemporal clustering detection

Spatiotemporal clustering analysis was performed using Moran's I and SaTScan methodologies. First, Moran's I test was used to evaluate global autocorrelation amongst communes that reported at least one case (n = 78) of typhoid fever in GeoDa software (v1.6.7, https://geodacenter.asu.edu/). This test statistic provides an evaluation of whether the rates across the area of interest are spatially random (Moran's I = 0), over-dispersed (Moran's I<0) or clustered (Moran's I>0) [26]. Next, Kulldorff's scan statistic in SaTScan (v9.1.1, http://www.satscan.org/) was used to identify the location of clusters of communes with high rates of typhoid fever over space and time [27,28]. A cylindrical window was used to scan the area for clusters, with the size of the circle corresponding to the spatial scan and the height of the cylinder corresponding to time. The significance of the detected clusters was assessed by a likelihood ratio test, with a *p*-value obtained by 999 Monte Carlo simulations generated under the null hypothesis of random spatiotemporal distribution. In this analysis, scan windows were used to fit discrete Poisson models. For the sublineage-specific analyses, all case communes were included and those without cases of a specific sublineage were classified as having 0 cases. The upper limit for cluster detection was specified as 25% of the study population over each year. All maps were created in ArcGIS 10.2 (ESRI, Redlands, CA, USA).

Whole-genome sequencing and phylogenetic analysis

Of the 284 *S.* Typhi isolates collected between 2007 and 2014, a total of 209 (74%) collected between 2007 and 2012, were subjected to genomic DNA extraction using the Wizard Genomic DNA Extraction Kit (Promega, Wisconsin, USA) (S1 Table). Two micrograms of genomic DNA was subjected to WGS on an Illumina HiSeq2000 platform following the manufacturer's recommendations to generate 100bp paired-end reads. All reads were mapped to the reference sequence of *S.* Typhi strain CT18 (Accession no: AL513382), plasmid pHCM1 (AL513383) and pHCM2 (AL513384) using SMALT (version 0.7.4). Candidate SNPs were called against the reference sequence using SAMtools and filtered with a minimal mapping quality of 30 and a quality ratio cut-off of 0.75. The allele at each locus in each isolate was determined by reference to the consensus base in that genome, using *samtools mpileup* and removing low confidence alleles with consensus base quality $\leq$20, read depth $\leq$5 or a heterozygous base call. SNPs in phage regions, repetitive sequences or recombinant regions were excluded, resulting in a final set of 750 chromosomal SNPs. Strains belonging to haplotype H58 were defined by the SNP *glpA*-C1047T (position 2,348,902 in *S.* Typhi CT18, BiP33, as previously described [12,13]).

A maximum likelihood phylogenetic tree was constructed from a 188 chromosomal SNP alignment of H58 isolates with RAxML (version 7.8.6) using the generalized time-reversible model (GTR) and a gamma distribution to model site-specific rate variation (GTR+Γ nucleotide substitution model in RAxML). Support for the ML phylogeny was assessed via 1,000 bootstrap pseudo-analyses of the alignment data. Phylogenetic subgrouping was defined based on monophyletic groups (lineages) with well-supported bootstrap value ($\geq$85%).

To investigate the short-term divergence within the bacterial population and the transmission within the local population, a minimum spanning tree was reconstructed from the SNP alignment of lineage III and lineage IV identified in the ML tree (accounting for 95% of isolates) using the goeBURST algorithm in Phyloviz software (version 1.1) [29]. This algorithm identified seven sublineages based on similarity among allelic profiles and frequency of isolation within the population. Sequences with identical SNP profiles and isolated at the highest frequency within each sublineage were assigned as founder genotypes (viewed as the central nodes within each of the sublineages), with descendant genotypes (represented by terminal nodes surrounding the founder genotype) assigned based on similarity to founder SNP profiles. These descendent genotypes can differ from the parental genotype by a single or multiple SNPs. The raw sequence data for this study are available in the European Nucleotide Archive (ENA) under the accession numbers described in S1 Table.

Results

Baseline characteristics

Between 2007 and 2014, there were 284 microbiologically confirmed cases of typhoid fever caused by *S.* Typhi at AHC in Siem Reap. *S.* Paratyphi A was uncommon, with only three cases in 2008 followed by an isolated outbreak in 2013–2014 (38 cases). A total of 262/284 (93%) of the confirmed *S.* Typhi cases lived within a 100 km radius of AHC and spanned 78 communes; these 78 communes were selected for the spatial comparison and the typhoid fever population level risk factor analyses. During this same period there were 19,877 admissions with an ICD-10 discharge diagnosis other than typhoid fever originating from the same geographic area. The baseline characteristics of all communes and those with at least one case of typhoid fever are shown in Table 1.

Of the 262 cases of typhoid fever living within a 100 km radius of AHC, the median age was 8.2 years (interquartile range (IQR): 5.1–11.5 years). Additionally, 62/262 (24%) of the cases were less than five years of age and 142/262 (54%) were female. As shown in Fig 1a, the absolute number of confirmed cases of typhoid fever increased dramatically (from 12 cases per year to 71 cases per year) between 2009 and 2012, but then declined in 2013 and 2014 (28 and 45 cases in 2013 and 2014; respectively); data from our non-confirmed typhoid cases also reflected this trend. Over this same time period (2009 to 2014) the number of patients attending AHC for other conditions (control population) mirrored the distribution of the cases (Fig 1b). There was seasonal variation in the number of typhoid cases, with the majority of the cases (178/262; 68%) occurring during the early monsoon months (April, May, June and July) (Fig 1c & 1d). In late monsoon months (August to October), the number of cases declined to less than two cases per month and generally remained below this threshold in the dry season (November to March) (Fig 1c & 1d).

Spatiotemporal clustering of typhoid fever cases

The majority of *S.* Typhi cases (241/284; 85%) originated from communes located within Siem Reap province (Fig 2). The median population density in communes with at least once case of typhoid fever was 119 people/km^2 (IQR: 60–212), and 70/78 (90%) of communes with a typhoid fever case were classified as rural. Compared to typhoid cases, the non-typhoid fever population controls came from a larger area (243 communes), the median population density of which was lower at 106 people/km^2 (IQR: 53–210); however, a similar proportion of these communes (220/243; 91%) was also classified as rural (Fig 2).

The estimated median commune level minimum incidence of reported cases of typhoid fever over the study period was 0.62/1,000 children aged <15 years (IQR: 0.37–1.02; range:

Table 1. Baseline characteristics of all communes and those with at least one case of typhoid fever.

Characteristic	All communes		Typhoid communes	
	median	IQR	median	IQR
	n = 243		n = 78	
Population density/km^2	105.7	53–210	119.4	60–214
Elevation, m	17	12–28	18	11–35
Distance to lake, km	45.3	24–63	30.0	14–49
Average household size	4.8	4.6–5.0	5.0	4.8–5.1
Percent of population <15 yr	36.4%	34–39%	36.9%	35–39%
Median age of population, yr	19.5	18–21	19.4	18–20
Adult literacy	72.8%	59–82%	69.3%	57–77%
Female adult literacy	65.3%	50–75%	62.5%	49–71%
Total attending school	28.6%	26–31%	28.0%	24–31%
Female attending school	26.8%	24–29%	26.0%	23–28%
Female education >25 years /1,000 population				
Primary not completed	85.6	63–101	78.5	59–101
Primary/Lower secondary	27.8	16–55	22.0	16–39
Secondary or above	0.51	0.1–1.6	0.4	0.1–1.4
Toilet, % of households				
None	83.1%	63–92%	85.2%	68–92%
Sewage	5.3%	2–14%	4.6%	2–15%
Septic tank	3.9%	1–16%	2.8%	1–11%
Pit latrine	2.0%	1–5%	1.2%	1–5%
Drinking water, % of households				
Piped	1.5%	1–4%	2.1%	1–4%
Tube/pipe well	10.2%	3–28%	23.6%	8–63%
Dug well	26.9%	11–56%	23.6%	9–63%
Spring/river	23.9%	4–54%	4.7%	1–27%
Drinking water location, % of households				
Within premises	19.3%	10–35%	27.8%	18–56%
Near premises	31.1%	22–40%	28.0%	21–34%
Away premises	41.8%	23–56%	35.0%	11–52%

IQR: interquartile range

doi:10.1371/journal.pntd.0004785.t001

0.5–11.36). The reported incidence varied significantly across the 78 communes. Kampong Kleang commune (Soutr Nikom district, Siem Reap) showed the highest incidence of typhoid fever over the study period with 11.36 cases of typhoid fever /1,000 population of children aged <15 years (Fig 2c). This area is renowned for its floating villages and is situated on the edge of Tonle Sap Lake, approximately 35 km southeast of Siem Reap City. The second highest incidence was identified in Kaoh Chiveang commune (Aek Phnum district, Battambang, 33 km southwest of Siem Reap City) with 4.1 cases/1000 people aged <15 years over the study period (Fig 2c). Both of these areas experience heavy flooding when the Tonle Sap Lake expands during the rainy season.

Overall, there was some evidence of positive spatial autocorrelation (case clustering) across the 78 communes that had at least one case of typhoid fever between 2007 and 2014 (Moran's I = 0.11, $p<0.056$). The magnitude of this autocorrelation varied over time, and was the most significant in 2013 (Moran's I = 0.19, $p<0.019$) but was non-significant in other years. We

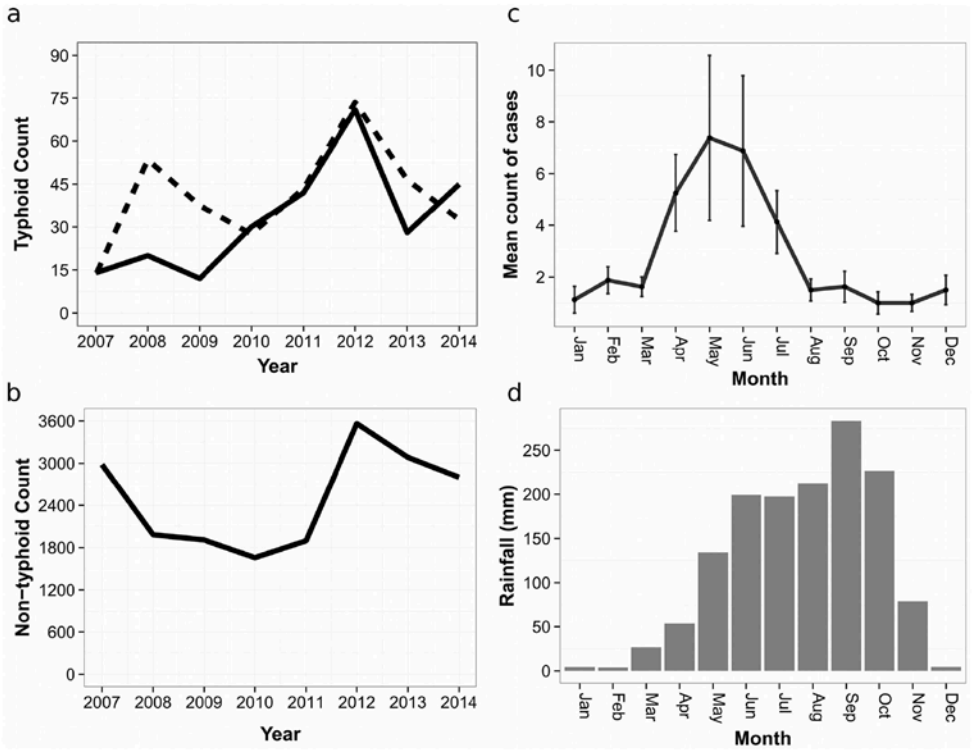

Fig 1. The annual and seasonal distribution of typhoid fever cases at Angkor Hospital for Children in Cambodia. a) The annual number of culture confirmed (solid line) and non-confirmed (broken line) typhoid cases at AHC from 2007 to 2014. b) The annual number of total admissions at AHC from 2007 to 2014. c) The mean monthly count of typhoid cases aggregated from 2007 to 2014. d) The average monthly rainfall (mm) per month over the study period.

doi:10.1371/journal.pntd.0004785.g001

were able to identify three significant spatiotemporal clusters associated with high rates of typhoid fever. The first occurred in 2008 toward the west of the study area and had a radius of 23.8 km; this cluster had 1.27 predicted cases and 10 observed cases (relative risk [RR] = 8.17, $p = 0.002$). The second cluster occurred in 2012 in the central northern area and had a radius of 10.8 km, with 1.67 predicted cases and 12 observed cases (RR = 7.47, $p<0.001$). The final cluster occurred in 2013 in the southeastern area and had a radius of 15.5 km, with 0.88 predicted cases and 14 observed cases (RR = 16.8, $p<0.0001$) (Fig 2d).

The population structure of *Salmonella* Typhi in Siem Reap province, Cambodia

The resulting WGS data demonstrated that 97% (203/209) of the sequenced Cambodian isolates could be attributed to haplotype H58. The majority (199/203, 98%) of the H58 isolates

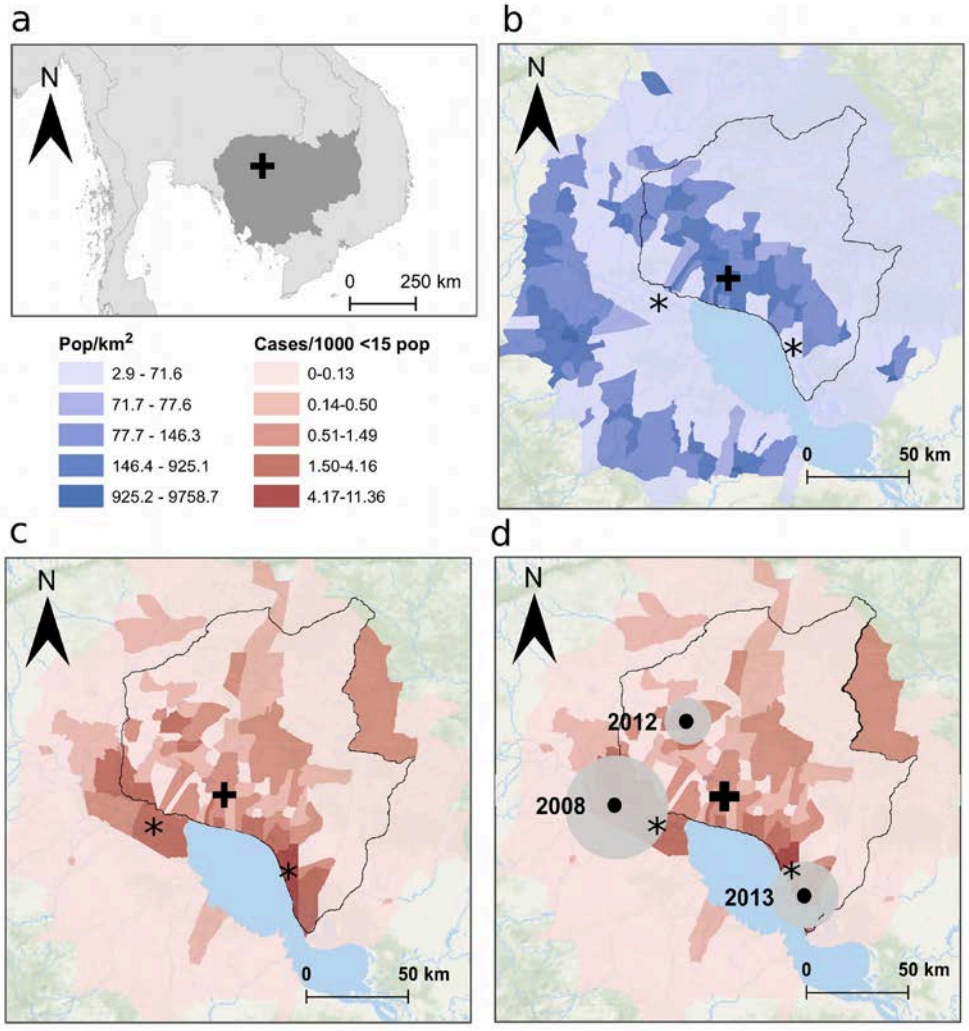

Fig 2. The spatial distribution of typhoid fever cases in Siem Reap province, Cambodia. a) North oriented map of Cambodia, the black cross shows the location of AHC. b) Map showing the population density (people/km^2, color-coding in key) of the 78 communes within the typhoid study area. AHC is shown by the black cross, the black border denotes Siem Reap province and the left and right asterisks are mark the locations of the communes with highest incidence of typhoid fever; Kaoh Chiveang and Kampong Kleang, respectively. c) Map of the study area showing the rate of reported typhoid cases per 1,000 population under the age of 15 years (color-coding in key). d) Map of the study area showing significant spatiotemporal clusters of typhoid during the study period, the size of the grey circles corresponds to the radius of the cluster and the years of the clusters are denoted.

doi:10.1371/journal.pntd.0004785.g002

exhibited intermediate susceptibility against fluoroquinolones (0.12–0.5 μg/mL) via the common amino acid substitution of serine to phenylalanine at codon 83 (S83F) in the DNA gyrase protein encoded by *gyrA*. There was a strong association between haplotype H58 and an IncHI1 plasmid, which confers an MDR phenotype, with 89% (180/203) of the H58 isolates harboring the common IncHI1 plasmid and the corresponding antimicrobial resistance phenotype. For the six non-H58 isolates, no mutations were observed in the *gyrA* gene, while two (33%) carried the same IncHI1 plasmid as found in the H58 isolates. We identified 188 SNPs across the H58 population and, from a SNP-based phylogeny, identified the circulation of at least four lineages of H58 circulating in the selected area of Cambodia between 2007 and 2012 (Fig 3a). These lineages, designated here as I-IV, differed from each other by as little as three to

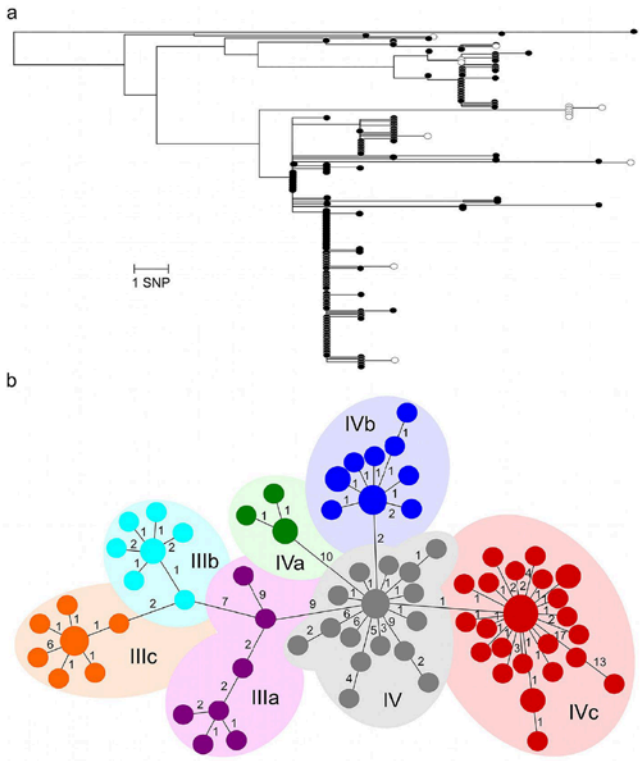

Fig 3. The phylogenetic structure of the H58 lineage of Cambodian *Salmonella* Typhi. a) Maximum likelihood phylogenetic tree of the 203 H58 isolates identified during this project (scale bar denotes SNP differences). The sub-lineages are shown on the major branches. Isolates exhibiting a multi-drug resistance (MDR) phenotype are indicated by black nodes. The tree is midpoint-rooted for the purpose of clarity. Bootstrap values >85% are indicated by an asterisk. b) Minimum spanning tree subdividing H58 lineage III and IV into the various sublineages (IIIa, IIIb, IIIc, IV, IVa, IVb, IVc). The various sublineages are color-coded for reference and the number of each variant is indicated by the cluster size. The number on each of the branches signifies the number of SNPs between each cluster.

doi:10.1371/journal.pntd.0004785.g003

five SNPs and were phylogenetically well-supported (bootstrap values ≥ 87%). The majority of the H58 isolates fell into lineage IV (152/203, 75%) and lineage III (41/203, 20%).

The spatiotemporal distribution of *Salmonella* Typhi genotypes

To investigate short-term evolutionary traits within the identified lineages, we constructed a SNP-based minimum spanning tree (Fig 3b). Using these data, we were able to investigate the local population dynamics and detected several clonal clusters emerging from lineage III (IIIa-IIIc) and lineage IV (IVa-IVc); SNPs defining these sublineages are shown in S2 Table. Our data show a complex temporal distribution of *S*. Typhi H58 sublineages circulating in this location between 2007 and 2012 (Fig 4a). The distribution of these various strains was highly dynamic, with strain replacements, potential extinctions and the specific microevolution and expansion of H58-IVc (Fig 4a). In 2011 and 2012, H58-IVc became the dominant genotype, accounting for 44% (18/42) and 85% (61/72) of all *S*. Typhi isolates in these years, respectively.

We next aimed to identify spatiotemporal clustering of the various *S*. Typhi H58 sublineages, and found that IIIc, IV, IVb and IVc all displayed significant evidence of clustering over space and time. Notably, the locations of these clusters were generally different between sublineages, signifying some degree of geographical variation of the circulating *S*. Typhi strains. For example, we identified significant clustering of H58-IIIc in the western part of the study area in 2011 ($p<0.001$, RR: 26.7, radius: 36km) (Fig 4b) and clustering of the emergent

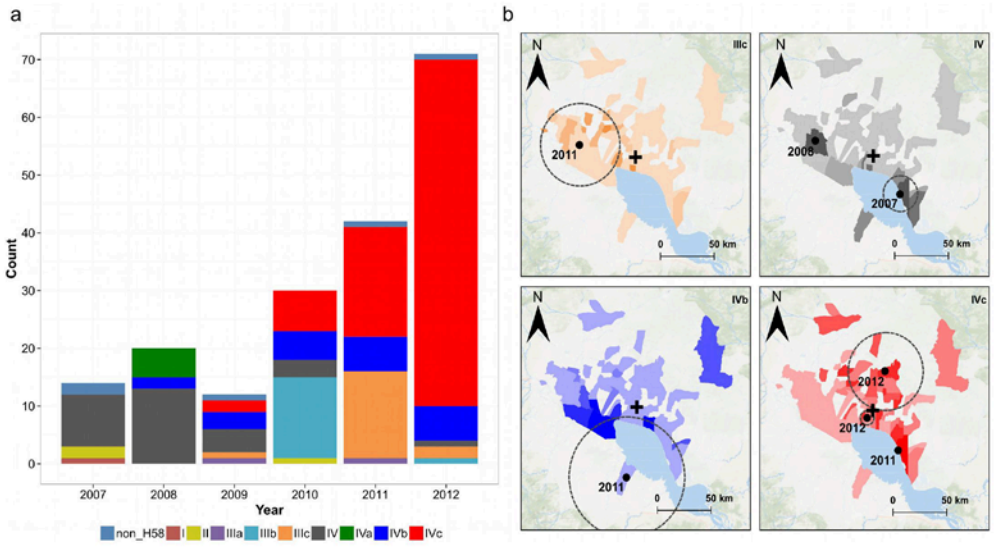

Fig 4. The spatiotemporal distribution of the various *Salmonella* Typhi lineages/sublineages in Siem Reap province, Cambodia. a) Bar chart shows the annual distribution of the various *S*. Typhi lineages/sublineages from 2007 to 2012; sublineages are color-coded as in Fig 3b. b) Maps showing significant spatiotemporal clusters identified for sublineages IIIc, IV, IVb and IVc. The timing of each cluster is shown by the year in black text and the dotted circle represents the radius of the detected cluster. Background colors represent the rate of each sublineage per 1,000 population aged under 15 years. The incidence rates vary between sublineages, ranging from 0 to a maximum of 0.8 (IIIc), 3.12 (IV), 2.56 (IVb) and 5.84 (IVc) 5.84 cases/1,000 population aged under 15 years.

doi:10.1371/journal.pntd.0004785.g004

H58-IVc strain in both 2011 (Kampong Khleang commune, $p<0.001$, RR: 39.4, radius: <1km) and in two locations in 2012 (smaller cluster, $p = 0.017$, RR: 5.17, radius: 6.2km; larger cluster, $p<0.001$, RR: 5.87, radius: 33.9km).

Population risk factors for typhoid fever

Finally, we investigated associations between rates of typhoid in children and demographic and sanitation variables at the commune level. We found a number of significant risk factors (e.g. low female education level and collection of drinking water near the household premises) and protective factors (e.g. higher population density, elevation, distance from lake and attendance at school) associated with the rate of typhoid hospitalizations in the univariate analysis (Table 2). However, after controlling for confounders, we found that the distance of the centroid of the commune to the perimeter of the lake was strongly and significantly associated with rate of typhoid cases (10km increase in distance from the lake, incidence rate ratio (IRR): 0.38, 95%CI 0.26–0.55, $p<0.001$) (Table 2). Furthermore, the relative numbers of households within the commune connected to public sewage services and households using a sunken well were also strongly protective, however these associations were reversed through interaction with increasing number of households using wells and distance from the lake, respectively

Table 2. Regression results highlighting factors associated with typhoid cases.

Commune characteristic	Univariable		Multivariable	
	IRR (95%CI)	p	IRR (95%CI)	P
Population density^	0.81 (0.70–0.95)	0.008		
Elevation, 10m	0.89 (0.81–0.99)	0.026		
Distance to lake, 10km	0.81 (0.74–0.89)	<0.001	0.38 (0.26–0.55)	<0.001
Average household size	1.54 (0.66–3.57)	0.317		
Total attending school/1,000^	0.11 (0.04–0.33)	<0.001		
Female education >25 years /1000 population^				
Primary not completed	2.59 (1.53–4.38)	<0.001		
Primary/Lower secondary	0.94 (0.70–1.25)	0.654		
Secondary or above	0.96 (0.80–1.17)	0.714		
Toilets per 1000 people^				
None	1.24 (0.84–1.79)	0.246		
Sewage	0.96 (0.81–1.15)	0.676	0.44 (0.25–0.80)	0.007
x households with wells			1.19 (1.07–1.32)	0.001
Septic tank	0.89 (0.78–1.02)	0.090		
Pit latrine	0.95 (0.79–1.15)	0.621		
Drinking water, hh/1000 hh^				
Piped	0.87 (0.74–1.03)	0.104		
Tube/pipe well	0.82 (0.73–0.92)	0.001		
Dug well	0.83 (0.64–0.83)	<0.001	0.31 (0.19–0.50)	<0.001
x distance to lake			1.16 (1.09–1.25)	<0.001
Spring/river	1.15 (1.05–1.25)	0.003		
Drinking water location, hh/1000 hh^				
Within premises	0.71 (0.55–0.93)	0.013	0.65 (0.49–0.86)	0.003
Near premises	3.38 (2.24–5.10)	<0.001		
Away premises	0.88 (0.73–1.06)	0.177		

^log of the variable was included; hh: household

doi:10.1371/journal.pntd.0004785.t002

(Table 2). Finally, a high number of households reporting drinking water retrieval from 'within the household premises' was also associated with a significant protective effect (log households/1,000 households, IRR: 0.65, 95%CI: 0.49–0.86, $p = 0.003$).

Discussion

In this study we combined conventional epidemiological methods, current genome sequencing tools and geospatial mapping to add insight into the epidemiology of typhoid fever in pediatric patients attending a single healthcare facility in central Cambodia. The majority of recent typhoid fever studies originate from urban locations in low-income countries. This study provides a new perspective into this important community-acquired infection from a predominantly rural setting. The primary finding of this study is that there is a considerable and widespread burden of pediatric typhoid fever in rural Cambodia, thus questioning the dogma that typhoid fever is predominantly geographically restricted to urban populations with poor sanitation systems [9,7]. Our data are consistent with findings from a recent study conducted across sub-Saharan Africa [30]. The Typhoid Surveillance in Africa Programme (TSAP) found a large burden of typhoid fever in younger children and almost equivalent population incidences between urban and rural settings. This distribution was most apparent in West Africa (Burkina Faso and Ghana) and was similarly restricted to children aged less than 15 years [30]. Therefore, we infer that the epidemiology of typhoid fever in Cambodia may be more similar to contemporary observations from sub-Saharan Africa, as opposed to the urban distribution that has commonly been observed across much of Asia [7,31].

The impending availability of Vi-conjugate vaccine raises the question of who should be given this vaccine and when it should be given to obtain maximum benefit in the control of typhoid fever [32,33]. This issue is complicated by a lack of population-based incidence data and a poor understanding of the burden of disease in school and preschool aged children, for whom the conjugated form of the Vi polysaccharide vaccine would be particularly beneficial [34]. Our data indicate a substantial burden of typhoid fever in school and preschool aged children in this area, with a hospital-based incidence (i.e. a minimum population incidence) of 11.36 cases of typhoid fever /1,000 population in children aged <15 years over the study period. The overall burden of typhoid fever in this population is likely to be greater than we have estimated due to poor sensitivity of blood culture and restriction of the study to a single healthcare center. Siem Reap province could be a suitable location in which to trial, or even introduce, the next generation typhoid vaccines in Cambodia that have been tested elsewhere [32]. Further, we suggest that immunizing school-aged children in the period prior to the wet season may provide the most economic and prudent approach for vaccine introduction.

Between 2007 and 2012, we observed a sharp increase in the number of typhoid cases concurrent with an increasing geographic expansion. We also observed that typhoid fever in this population followed a seasonal pattern, suggesting an association with rainfall and potentially with localized flooding and the contamination of water sources. The population-based risk factors support these hypotheses, as living further away from Tonle Sap Lake and access to water within the household were highly protective. Additionally, we found that two communes located next to the lake (Kaoh Chiveang and Kampong Kleang) had the highest incidence of typhoid fever and had large clusters of cases in 2008 and 2013. This case clustering in specific locations warrants further investigation at the household level to understand specific sanitation-associated risk factors and likely exposures to *S.* Typhi in this setting [35]. It appears that access to lake water in some of these communes, such as Kaoh Chiveang, is vital for the household water supply and we hypothesize that the lake water is more prone to localised fecal contamination at specific times throughout the year.

Using targeted SNP-specific PCR, we have previously shown that MDR H58 *S*. Typhi strains dominate in this population [18]. Our WGS investigation confirmed these findings and identified additional diversification in this population. We were able to separate these H58 strains into seven (IIIa, IIIb, IIIc, IV, IVa, IVb, IVc) major sublineages. These discrete groups varied in size and were segregated by only limited numbers of SNPs. We did observe some evidence of expansion of sublineage IVc between 2009 and 2012; this correlated with several spatiotemporal clusters suggesting small disease outbreaks. We currently cannot explain the expansion of this group and our strain selection for sequencing was limited by the availability of strains isolated only up to 2012. Despite some clustering of closely related strains, the overall temporal and spatial distribution of strains was random, with a range of *S*. Typhi H58 sublineages circulating throughout the study period, which is similar to patterns described in urban settings in Asia [14,36].

This study has some limitations. First the data originated from patients attending a single healthcare facility, without the added support of healthcare utilization data. This approach, while cost-effective, induces bias in the spatial and risk factor analyses. Furthermore, while the associations identified in the regression analysis are plausible and provide direction for future investigations, they should be viewed with caution. The population level census data does not allow examination of exposures at an individual or household-level and provides only broad epidemiological evidence. However, the association with distance to the lake and water and sanitation variables suggests these factors should be examined more rigorously in the future with respect to the dynamics of typhoid fever outbreaks. Similarly, the identification and location of the spatiotemporal clusters should be interpreted with some degree of caution. Communes without cases were not included in the cluster analyses due to a lack of data as to whether these regions truly lacked typhoid cases. A dataset with more complete spatial information on presence and absence of typhoid would permit a more reliable analysis.

In conclusion, we find a large burden of typhoid fever in children in rural Cambodia. Our conventional population-based risk factor analysis identified access to water in the household and increasing distance from Tonle Sap Lake as protective against typhoid fever in communes. Spatial mapping and WGS provided additional resolution to investigate these findings and confirmed that proximity to that lake was associated with discrete disease clusters. We confirmed the dominance of MDR H58 *S*. Typhi in this location and found a substantial amount of diversification within this lineage. Our data provide a platform for additional studies in the Cambodian population and suggest that this is a suitable location in which to introduce Vi conjugate vaccines for school children.

Supporting Information

S1 Table. Strain list and accession numbers for organisms used in this study.
(XLS)

S2 Table. SNPs defining H58 sublineages.
(XLS)

S1 Checklist. STROBE checklist.
(DOC)

Acknowledgments

We wish to acknowledge the staff of Ankor Hospital for Children for their assistance with this study and their on-going commitment in providing healthcare to the local community.

Author Contributions

Conceived and designed the experiments: DPT CNT MAR PT CMP SB. Performed the experiments: DPT CNT MAR SSon SSop VK CM NTVT LW KEH VW DP PT CMP. Analyzed the data: DPT CNT MAR NTVT KEH VW DP GET PT CMP. Contributed reagents/materials/analysis tools: DPT CNT MAR LW KEH VW DP GET ND GD PT CMP SB. Wrote the paper: DPT CNT MAR VK CM LW KEH GET ND GD PT CMP SB.

References

1. Crump JA, Luby SP, Mintz ED, et al. (2004) The global burden of typhoid fever. Bull World Health Organ 82: 346–353. PMID: 15298225
2. Schwartz E (2010) Typhoid and Paratyphoid Fever. Trop Dis Travel 366: 144–153.
3. Crump JA, Mintz ED (2010) Global Trends in Typhoid and Paratyphoid Fever. Clin Infect Dis 50: 241–246. doi: 10.1086/649541 PMID: 20014951
4. World Health Organization (2003) Background document: The diagnosis, treatment and prevention of typhoid fever. World Health Organization: 38.
5. Parry CM, Wijedoru L, Arjyal A, Baker S (2011) The utility of diagnostic tests for enteric fever in endemic locations. Expert Rev Anti Infect Ther 9: 711–725. A doi: 10.1586/eri.11.47 PMID: 21692675
6. World Health Organization (2008) Typhoid vaccines: WHO position paper. Wkly Epidemiol Rec 83: 49–59. PMID: 18260212
7. Baker S, Holt KE, Clements ACA, Karkey A, Arjyal A, et al. (2011) Combined high-resolution genotyping and geospatial analysis reveals modes of endemic urban typhoid fever transmission. Open Biol 1: 110008. doi: 10.1098/rsob.110008 PMID: 22645647
8. Brooks WA, Hossain A, Goswami D, Nahar K, Alam K, et al. (2005) Bacteremic typhoid fever in children in an urban slum, Bangladesh. Emerg Infect Dis 11: 326–329. PMID: 15752457
9. Khan MI, Ochiai RL, Soofi SB, Von-Seidlein L, Khan MJ, et al. (2012) Risk factors associated with typhoid fever in children aged 2–16 years in Karachi, Pakistan. Epidemiol Infect 140: 665–672. doi: 10.1017/S0950268811000938 PMID: 21676350
10. Feasey NA, Gaskell K, Wong V, Msefula C, Selemani G, et al. (2015) Rapid emergence of multidrug resistant, H58-lineage Salmonella typhi in Blantyre, Malawi. PLoS Negl Trop Dis 9: e0003748. doi: 10.1371/journal.pntd.0003748 PMID: 25909750
11. Wong VK, Baker S, Pickard DJ, Parkhill J, Page AJ, et al. (2015) Phylogeographical analysis of the dominant multidrug-resistant H58 clade of Salmonella Typhi identifies inter- and intracontinental transmission events. Nat Genet 47: 632–639. doi: 10.1038/ng.3281 PMID: 25961941
12. Holt KE, Parkhill J, Mazzoni CJ, Roumagnac P, Goodhead I, et al. (2008) High-throughput sequencing provides insights into genome variation and evolution in Salmonella Typhi. Nat Genet 40: 987–993. doi: 10.1038/ng.195 PMID: 18660809
13. Roumagnac P, Weill F-X, Dolecek C, Baker S, Brisse S, et al. (2006) Evolutionary history of Salmonella typhi. Science 314: 1301–1304. PMID: 17124322
14. Holt KE, Dolecek C, Chau TT, Duy PT, La TTP, et al. (2011) Temporal fluctuation of multidrug resistant salmonella typhi haplotypes in the mekong river delta region of Vietnam. PLoS Negl Trop Dis 5: e929. doi: 10.1371/journal.pntd.0000929 PMID: 21245916
15. Kariuki S, Revathi G, Kiiru J, Mengo DM, Mwituria J, et al. (2010) 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 48: 2171–2176. doi: 10.1128/JCM.01983-09 PMID: 20392916
16. Baker S, Holt K, van de Vosse E, Roumagnac P, Whitehead S, et al. (2008) High-throughput genotyping of Salmonella enterica serovar Typhi allowing geographical assignment of haplotypes and pathotypes within an urban District of Jakarta, Indonesia. J Clin Microbiol 46: 1741–1746. doi: 10.1128/JCM.02249-07 PMID: 18322069
17. Kasper MR, Sokhal B, Blair PJ, Wierzba TF, Putnam SD (2010) Emergence of multidrug-resistant Salmonella enterica serovar Typhi with reduced susceptibility to fluoroquinolones in Cambodia. Diagn Microbiol Infect Dis 66: 207–209. doi: 10.1016/j.diagmicrobio.2009.09.002 PMID: 19800753
18. Emary K, Moore CE, Chanpheaktra N, An KP, Chheng K, et al. (2012) Enteric fever in Cambodian children is dominated by multidrug-resistant H58 Salmonella enterica serovar Typhi with intermediate susceptibility to ciprofloxacin. Trans R Soc Trop Med Hyg 106: 718–724. doi: 10.1016/j.trstmh.2012.08.007 PMID: 23122884

19. Wijedoru LPM, Kumar V, Chanpheaktra N, Chheng K, Smits HL, et al. (2012) Typhoid Fever among Hospitalized Febrile Children in Siem Reap, Cambodia. J Trop Pediatr 58: 68–70. doi: 10.1093/tropej/fmr032 PMID: 21508082

20. National Institute of Statistics (2008) General Population Census of Cambodia 2008. Ministerial Plan: 1–37.

21. Kirsch H (2008) Watershed Inventory Siem Reap, Cambodia: 9–14.

22. Kummu M (2009) Water management in Angkor: Human impacts on hydrology and sediment transportation. J Environ Manage 90: 1413–1421. doi: 10.1016/j.jenvman.2008.08.007 PMID: 18926615

23. Jarvis A, Reuter HI, Nelson A, Guevara E (2008) Hole-filled seamless SRTM data V4. Int Cent Trop Agric: available from http://srtm.csi.cgiar.org.

24. CLSI (2012) Clinical and Laboratory Standards Institute: Performance standards for antimicrobial susceptibility testing; twentieth informational supplement. M100–S22.

25. Wickham H (2009) ggplot2: elegant graphics for data analysis. New York: Springer-Verlag.

26. Pfeiffer D, Robinson T, Stevenson M, Stevens K, Rogers D, et al. (2008) Spatial Analysis in Epidemiology. New York City: Oxford University Press.

27. Kulldorff M (1997) A spatial scan statistic. Commun Stat—Theory Methods 26: 1481–1496.

28. Kulldorff M, Heffernan R, Hartman J, Assunção R, Mostashari F (2005) A Space–Time Permutation Scan Statistic for Disease Outbreak Detection. PLoS Med 2: e59. PMID: 15719066

29. Francisco AP, Vaz C, Monteiro PT, Melo-Cristino J, Ramirez M, et al. (2012) PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. BMC Bioinformatics 13: 87. doi: 10.1186/1471-2105-13-87 PMID: 22568821

30. Nichols C, Cruz Espinoza LM, Von Kalckreuth V, Aaby P, Ahmed El Tayeb M, et al. (2015) Bloodstream infections and frequency of pretreatment associated with age and hospitalization status in Sub-Saharan Africa. Clin Infect Dis 61: S372–S379. doi: 10.1093/cid/civ730 PMID: 26449954

31. Cruz Espinoza LM, Nichols C, Adu-Sarkodie Y, Al-Emran HM, Baker S, et al. (2016) Variations of Invasive Salmonella Infections by Population Size in Asante Akim North Municipal, Ghana. Clin Infect Dis 62 Suppl 1: S17–S22. doi: 10.1093/cid/civ787 PMID: 26933015

32. Date KA, Bentsi-Enchill A, Marks F, Fox K (2015) Typhoid fever vaccination strategies. Vaccine 33: C55–C61. doi: 10.1016/j.vaccine.2015.04.028 PMID: 25902360

33. Crump JA (2015) Building the case for wider use of typhoid vaccines. Vaccine 33: C1–C2. doi: 10.1016/j.vaccine.2015.04.033 PMID: 25910921

34. Mai NL, Phan VB, Vo AH, Tran CT, Lin FYC, et al. (2003) Persistent efficacy of Vi conjugate vaccine against typhoid fever in young children. N Engl J Med 349: 1390–1391. PMID: 14523155

35. Karkey A, Jombart T, Walker AW, Thompson CN, Torres A, et al. (2016) The Ecological Dynamics of Fecal Contamination and Salmonella Typhi and Salmonella Paratyphi A in Municipal Kathmandu Drinking Water. PLoS Negl Trop Dis 10: e0004346. doi: 10.1371/journal.pntd.0004346 PMID: 26735696

36. Pham Thanh D, Karkey A, Dongol S, Ho Thi N, Thompson CN, et al. (2016) A novel ciprofloxacin-resistant subclade of H58 Salmonella Typhi is associated with fluoroquinolone treatment failure. Elife. 2016 Mar 11; 5. pii: e14003. doi: 10.7554/eLife.14003 PMID: 26974227

nature genetics

Phylogeographical analysis of the dominant multidrug-resistant H58 clade of *Salmonella* Typhi identifies inter- and intracontinental transmission events

The emergence of multidrug-resistant (MDR) typhoid is a major global health threat affecting many countries where the disease is endemic. Here whole-genome sequence analysis of 1,832 *Salmonella enterica* serovar Typhi (*S.* Typhi) identifies a single dominant MDR lineage, H58, that has emerged and spread throughout Asia and Africa over the last 30 years. Our analysis identifies numerous transmissions of H58, including multiple transfers from Asia to Africa and an ongoing, unrecognized MDR epidemic within Africa itself. Notably, our analysis indicates that H58 lineages are displacing antibiotic-sensitive isolates, transforming the global population structure of this pathogen. H58 isolates can harbor a complex MDR element residing either on transmissible IncHI1 plasmids or within multiple chromosomal integration sites. We also identify new mutations that define the H58 lineage. This phylogeographical analysis provides a framework to facilitate global management of MDR typhoid and is applicable to similar MDR lineages emerging in other bacterial species.

S. Typhi, the primary global cause of human typhoid (enteric fever), is a monophyletic serovar of *S. enterica*. Unlike many *Salmonella*, *S.* Typhi are highly restricted to infection of humans and are associated with systemic infection, prolonged fever and an asymptomatic carrier state[1]. Typhoid is still a common disease in many regions of the world with poor infrastructure and limited economic development and is also a risk for travelers who visit such regions[2]. It is estimated that 20–30 million cases of typhoid occur annually, although deaths are less frequently reported than before the availability of effective antimicrobials[3,4].

In addition to improvements in access to clean water and sanitation, typhoid can potentially be controlled by other interventions such as vaccination[5–7] and antimicrobial therapy[8]. Chloramphenicol, ampicillin and trimethoprim-sulfamethoxazole were traditional first-line drugs commonly used to treat acute typhoid, and these agents continue to be used in areas of the world where *S.* Typhi are deemed susceptible. However, since the 1970s, *S.* Typhi have emerged that display multidrug resistance, defined as resistance to the above antimicrobials, compromising treatment[9–11]. Since the 1990s, alternative treatment options have included fluoroquinolones, third-generation cephalosporins (such as ceftriaxone) and the azalide azithromycin[1]. The early emergence of MDR *S.* Typhi was driven in large part by the acquisition of IncHI1 plasmids carrying antibiotic resistance genes[12] and, more recently, by chromosomal mutations associated with resistance to fluoroquinolones, and MDR strains have been reported across Asia and Africa[13–16].

Phylogenetic analysis, initially based on subgenomic DNA sequences but later on whole-genome DNA sequences, showed that the global *S.* Typhi population is highly clonal and likely originated from a common ancestor that moved into the human population several thousand years ago[17–19]. It also indicated that the population is relatively small and that recombination between *S.* Typhi and other Salmonellae is rare[12,19,20]. Simple SNP-based typing schemes have been developed that stratify the *S.* Typhi population into haplotypes, and these schemes are now used to unequivocally map new isolates to the phylogeny[17,19,21,22]. Notably, this approach identified a single emerging, highly clonal MDR haplotype of *S.* Typhi, H58, which is being reported with increasing frequency from many countries in Africa and Asia[12,17,19,23]. Within the H58 lineage, IncHI1 MDR plasmids of the restricted subtype PST6 (ref. 23) and chromosomal point mutations conferring quinolone resistance are common[14,24–26]. However, relatively little is known about the emergence and evolutionary history of the H58 lineage or how it is moving across endemic regions. Here we have used phylogenetic analysis based on the whole-genome sequences of a global collection of *S.* Typhi from 63 countries to investigate the genomic architecture of this highly successful *S.* Typhi lineage.

RESULTS

Phylogeography of H58

Of a global collection of 1,832 sequenced *S.* Typhi (listed in **Supplementary Tables 1** and **2**), 853 (47%) belonged to haplotype H58, initially defined by the SNP *glpA*-C1047T (position 2,348,902 in *S.* Typhi CT18, BiP33 in ref. 17). The earliest H58 isolates in our collection were from 1992 (Fiji) and 1993 (Fiji and Vietnam), and H58 isolates were represented every year from 1992 to 2013, at a mean rate of 40% per year (**Fig. 1a**). H58 isolates formed a tight cluster within the whole-genome maximum-likelihood phylogeny (**Fig. 1b**), forming a unique lineage separated by 151 SNPs from the nearest neighboring non-H58 cluster, which consisted exclusively of isolates from Fiji

A full list of authors and affiliations appears at the end of the paper.

Received 21 December 2014; accepted 23 March 2015; published online 11 May 2015; doi:10.1038/ng.3281

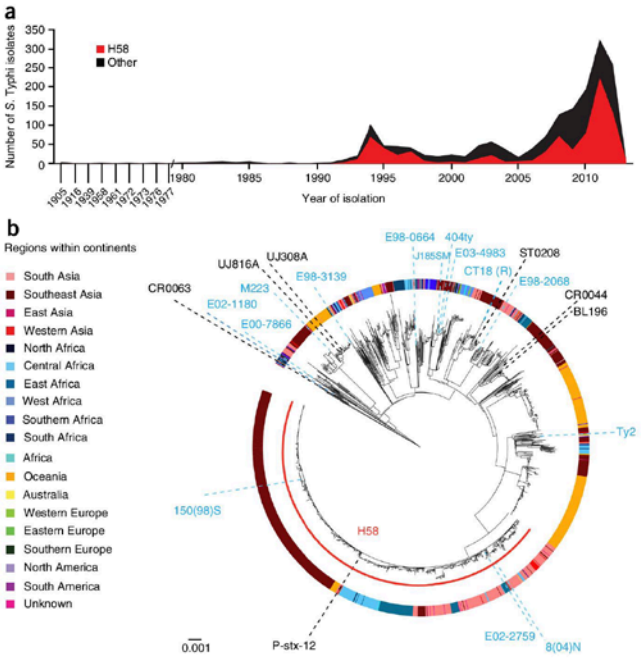

Figure 1 Population structure of the 1,832 *S.* Typhi isolates analyzed in this study. (**a**) Temporal distribution of the *S.* Typhi isolates included in the study. (**b**) Rooted maximum-likelihood tree of *S.* Typhi inferred from 22,145 SNPs, rooted using an outgroup (*S. enterica* serovar Paratyphi A, isolate 9953_5_4_Outgroup_ParatyphiA_IndoA270_2010). The colored ring indicates the geographical origin of the isolates. Red arc, H58 lineage; labeled blue dashed lines, public reference genomes reported in Holt *et al.*[19], including the CT18 (R) reference genome (AL513382); black dashed lines, other publicly available genomes. Branch lengths are indicative of the estimated substitution rate per variable site.

($n = 140$) and Tonga ($n = 2$). Individual H58 isolates differed from the most recent common ancestor (MRCA) of the H58 lineage by a median of just two SNPs, and the median distance between pairs of H58 isolates was six SNPs, strongly indicative of recent clonal expansion. Nearly all of the H58 isolates (93%; 797/853) had ≤5 isolate-specific SNPs, consistent with frequent transmission relative to substitution mutations. This finding was in contrast to that for the rest of the *S.* Typhi tree, which included a wide diversity of isolates (**Fig. 1b**), among which only 66% (642/979) had ≤5 isolate-specific SNPs ($P < 0.0001$, Fisher's exact test) (**Supplementary Fig. 1**).

The population structure within H58 was consistent with our previous work defining two major sublineages of H58 (I and II)[22,25,27] but provided much greater resolution of substructure with a strong phylogeographical signal (**Fig. 2**). Our H58 *S.* Typhi isolates were collected from 21 countries across Asia, Africa and Oceania. We observed strong phylogeographical clustering within 13 countries (**Fig. 2**), indicating transmission of H58 within these locations. Although our sample spans distinct time periods in different locations, in most cases, the localized subclades were isolated from the same country over ≥4 years, indicating the establishment of long-term local reservoirs (**Fig. 3a,b**).

These data demonstrate that H58 is now widely disseminated across distinct geographical areas, and the phylogeny provides several insights into the spatial patterns of its spread. There were numerous instances of very closely related isolates from different countries (**Fig. 2**), which indicate likely transfer events or regional outbreaks and identify routes for geographical dissemination (**Fig. 3a,b**). Maximum-likelihood analysis of inter-region transfers based on the maximum-likelihood phylogeny and locations

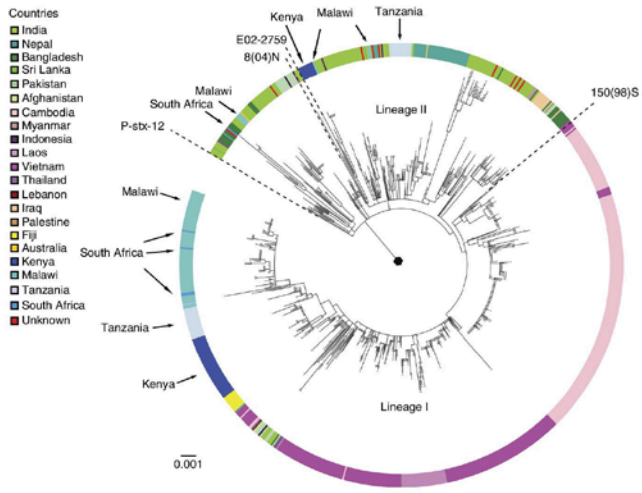

Figure 2 Population structure of the *S.* Typhi H58 lineage. Rooted maximum-likelihood phylogeny inferred from 1,534 SNPs identified in the 853 H58 isolates, rooted using an *S.* Typhi isolate from the nearest neighboring cluster of non-H58 isolates as an outgroup (black filled circle; isolate 10060_5_62_Fij107364_2012). The colored ring indicates the countries of isolation; countries discussed in the text are labeled around the tree. Branch lengths are indicative of the estimated substitution rate per variable site.

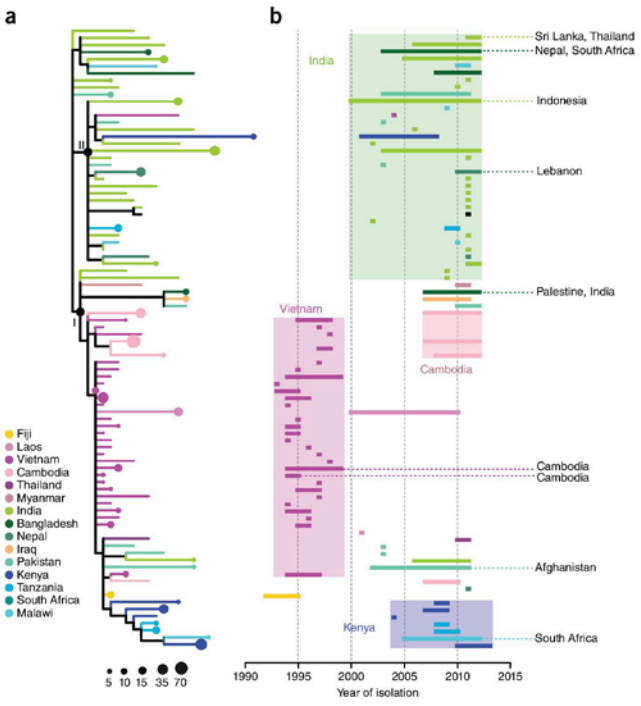

Figure 3 Geographical persistence and routes for dissemination of *S.* Typhi H58. (**a**) Maximum-likelihood tree for the H58 lineage (I and II), with clades containing isolates from a single country collapsed into nodes (circles), sized to indicate the number of isolates in the clade and colored by country of isolation. Branches are colored to indicate the country of origin of descendant nodes. (**b**) Years of isolation for each phylogeographical cluster in the tree, indicated by lines spanning the earliest and latest years of isolation for each cluster and colored to indicate the country. Four regions with extensive local clonal expansion are highlighted by shaded boxes, spanning the phylogenetic (*y* axis) and temporal (*x* axis) extent of the expansion. Locations from which singleton isolates were clustered within the phylogeographical clusters are shown to the right, indicative of further onward transmission.

where isolates were collected highlighted several candidate intercontinental transfers (**Fig. 4**). These data suggest that South Asia was an early hub for H58, from which it was propagated to many locations around the world, including countries in Southeast Asia, western Asia and East Africa, as well as Fiji (**Figs. 2–4**). Most of the diversity in lineage II was present among Indian isolates, with unique local subclusters detected in neighboring countries (Nepal and Pakistan) and in Africa, indicative of occasional transfers out of Asia. In contrast, lineage I was associated mainly with Southeast Asia (Vietnam, Cambodia and Laos), with evidence of transmission to Thailand, Pakistan, Fiji and Africa.

There have been sporadic reports of the emergence of *S.* Typhi in Africa[14,28–30]. Indeed, H58 isolates were predominant among the eastern and southern African *S.* Typhi isolates (63%) (**Supplementary Fig. 2**); in contrast, the H58 lineage was relatively rare in northern, western and central Africa. H58 lineages I and II were detected in Kenya, Tanzania, Malawi and South Africa, providing compelling evidence for multiple introductions of H58 *S.* Typhi from South Asia into the continent (**Figs. 2–4**). In addition, we uncovered evidence of an unreported recent wave of transmission of H58, based on 138 isolates, from Kenya to Tanzania and on to Malawi and South Africa (**Figs. 2** and **3**). These isolates differed from one another by an average of 10 (range of 0–30) SNPs, consistent with a recent clonal expansion. Therefore, this analysis demonstrates an ongoing epidemic of H58 typhoid across countries in eastern and southern Africa.

Dating the emergence of H58

Estimating mutation rates and divergence dates within the *S.* Typhi population has been challenging, as *S.* Typhi is known to establish persistent asymptomatic carriage, during which time it likely evolves at a different rate than during acute infection, disrupting the molecular clock[17]. Here a temporal signal was barely detectable across the full *S.* Typhi maximum-likelihood tree, assessed via linear regression of root-to-tip branch lengths on the basis of year of isolation (correlation coefficient (R) = 0.09 (95% confidence interval (CI) = 0.04–0.13); P = 0.0002, Fisher's exact test). However, a moderate signal was evident within the H58 subtree (R = 0.60 (95% CI = 0.56–0.64); $P < 1 \times 10^{-6}$, Fisher's exact test; **Supplementary Fig. 2**). This temporal signal was entirely destroyed by randomization of isolation dates (mean R = 0.01), indicating that uneven sampling of the H58 lineage across space and time was not solely responsible for the observed association. We propose that a temporal signal was detectable within the H58 tree because these data capture epidemic spread over a relatively short time span (2–3 decades), whereas the wider *S.* Typhi tree represents much more variable population dynamics over thousands of years of evolution, including recent periods of endemic transmission, that differ greatly from the clonal expansion of H58.

We therefore proceeded to estimate the divergence date of the H58 lineage via Bayesian phylodemographical modeling of the H58 population, implemented in Bayesian Evolutionary Analysis Sampling Trees (BEAST)[31]. To limit potential bias due to highly variable sampling intensities in different geographical locations, we performed BEAST analyses on several cross-sections of 114 H58 isolates (13% of the total), each sampled from 21 countries and spanning the years 1992 to 2013. The combined estimate for the median substitution rate within the H58 population was 1.42×10^{-7} substitutions per site per year (95% highest posterior density (HPD) = 1.0×10^{-7} to 1.8×10^{-7}), equivalent to the accumulation of 0.63 SNPs per genome per year (95% HPD = 0.59 to 0.67). The analyses predicted that the MRCA of all extant H58 strains existed ~25 years ago (median calendar year for divergence, 1989; 95% HPD = 1985–1992) and that the effective population size of H58 increased dramatically after 1993 (**Supplementary Fig. 2**). Although the BEAST analysis was limited

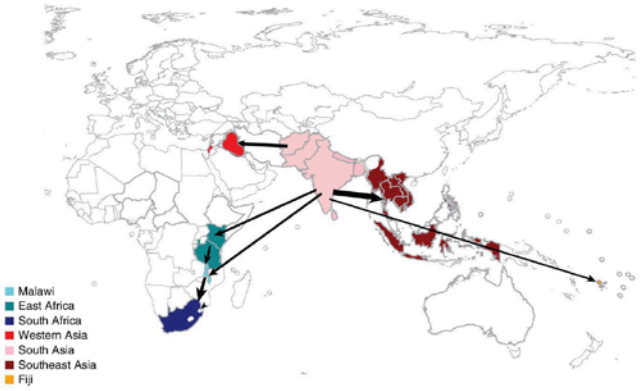

Figure 4 Major geographical transfers within the H58 lineage, inferred from the phylogenetic tree. The size of each arrow indicates the relative number of likely transfers between regions or countries.

because of the moderate strength of the temporal signal, these results are consistent with the very low numbers of SNPs within the H58 lineage, the clear evidence of clonal expansion from the maximum-likelihood tree (**Figs. 1b** and **3a**,**b**) and epidemiological data reporting increasing rates of multidrug resistance in Asia in the early 1990s (refs. 13,32).

Multidrug resistance in H58

The H58 lineage is associated with high levels of multidrug resistance and reduced susceptibility to fluoroquinolones[12,14,25]. Acquired resistance genes were identified in 671 of the 1,832 *S.* Typhi isolates, including 68% of the H58 isolates in comparison to just 9% of non-H58 isolates ($P < 1 \times 10^{-16}$, Fisher's exact test). In our collection, which included 15 countries with ≥2 MDR isolates, H58 was significantly associated ($P < 0.01$) with multidrug resistance in nearly all these locations, with the exception of central and western Africa, where multidrug resistance was detected but H58 was not (**Supplementary Fig. 3**). The most common resistance genes detected were *blaTEM-1* (ampicillin resistance), *dfrA7*, *sul1* and *sul2* (resistance to trimethoprim and sulfonamides, respectively, and to trimethoprim-sulfamethoxazole collectively), *catA1* (chloramphenicol resistance) and *strAB* (streptomycin resistance). These genes were each found in >540 H58 isolates, including in 525 isolates that carried all 7 genes. These genes are encoded within a Tn*2670*-like complex transposable element comprising transposon Tn*6029*, which carries *blaTEM-1*, *strAB* and *sul2*, inserted into transposon Tn*2670*, which itself comprises Tn*21* carrying a class I integron (including *sul1*, with *dfrA7* in the gene cassette) inserted into Tn*9* carrying *catA1* (refs. 12,33) (**Fig. 5**). In addition, 405 H58 isolates harbored the *tetB* gene located in Tn*10*. Other acquired resistance genes were rare, identified in <1% of the H58 isolates.

Previous reports linked multidrug resistance in H58 *S.* Typhi to the Tn*2670*-like element described above encoded on IncHI1 plasmids of the PST6 genotype[12,23]. Here we identified IncHI1-PST6 plasmids in 74% of the H58 isolates harboring the MDR element, including isolates from Southeast Asia, East Africa and South Asia (**Supplementary Table 3**), indicating intercontinental transmission of the IncHI1-PST6 MDR plasmid with its H58 *S.* Typhi host (**Fig. 6**). However, the remaining MDR H58 isolates lacked IncHI1 plasmid sequences, indicating that the resistance-conferring

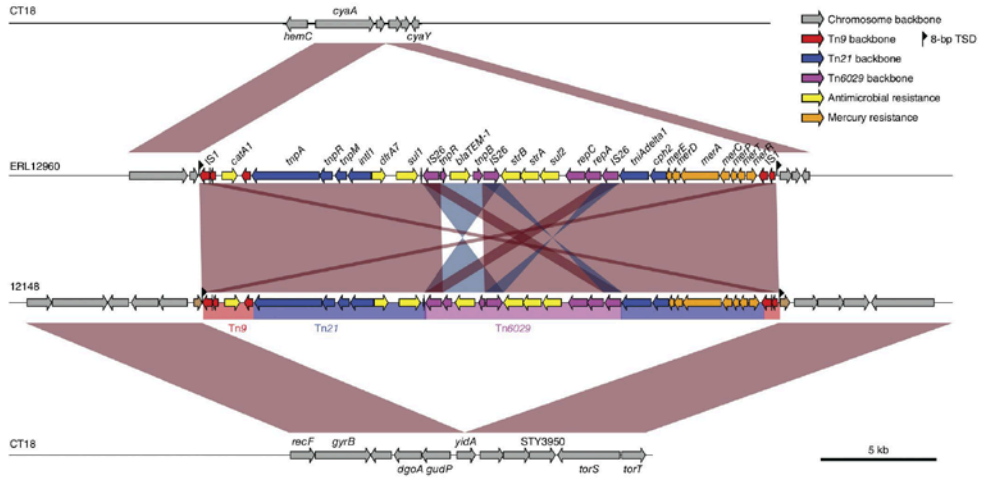

Figure 5 Insertion site of the 24-kb composite transposon in CT18. A comparative analysis using genoPlotr[48] of CT18 and H58 isolates ERL12960 (ERR343327) and 12148 (ERR343322) showed two integration sites of the transposon in the chromosome. The nucleotide sequence of the composite transposon was identical to that in the IncHI1-PST6 plasmid of H58 isolate 10425_1_48_Viety3-193_1997, which was sequenced for comparison. Tn, transposon; TSD, target site duplication.

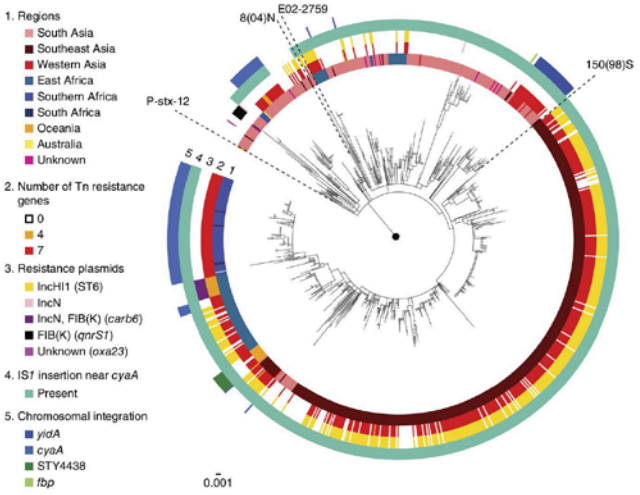

Figure 6 Acquired multidrug resistance in the *S.* Typhi H58 lineage. Maximum-likelihood phylogeny from 1,534 SNPs of 853 H58 isolates rooted using an *S.* Typhi isolate from the nearest neighboring cluster of non-H58 isolates as an outgroup (isolate 10060_5_62_Fij107364_2012) and surrounded by five colored rings representing (1) geographical origin in terms of region, (2) number of transposon-encoded resistance genes, (3) presence of resistance plasmids, (4) presence of IS*1* insertion near the *cyaA* gene and (5) the site of chromosomal integration. Black radial dashed lines show the positions of public reference strains. Branch lengths are indicative of the estimated substitution rate per variable site.

genes must be located elsewhere in the genomes of these 139 isolates. We screened all H58 isolates for known plasmid replicons and identified non-IncHI1 MDR plasmids in 23 isolates but not among isolates carrying the Tn*2670*-like MDR element (**Supplementary Table 3**).

Nearly all H58 isolates carried a copy of insertion sequence IS*1* between the chromosomal genes STY3618 and STY3619, near *cyaA* (**Fig. 6**). We propose that this insertion was acquired early in the H58 lineage, originating from the IncHI1-PST6 plasmid. Consistent with this hypothesis is the present analysis showing that IS*1* was absent from only a few isolates that were basal in the tree (**Fig. 6**). Long-read sequencing of 2012 Indian isolate ERL12960 (**Supplementary Table 4**) confirmed that the entire MDR locus, flanked by copies of IS*1*, was integrated at this location near *cyaA* (**Fig. 5**), in agreement with other recent reports[15,16]. The nucleotide sequence was identical to that in the IncHI1-PST6 plasmid of H58 isolate 10425_1_48_Viety3-193_1997, which we also sequenced for comparison. Further *in silico* analysis of our data identified new insertion sequences in MDR isolates that lacked the IncHI1 plasmid, including (i) 25 phylogenetically related isolates (mostly from Bangladesh, Pakistan and Iraq) with IS*1* in the *yidA* gene, (ii) 9 related isolates from Fiji with IS*1* in STY4438 and (iii) 1 isolate from India with IS*1* in the *fbp* gene (**Fig. 6**). Long-read sequencing of the 2012 isolate 12148 from India confirmed integration of the MDR locus in the *yidA* gene (**Fig. 5**); we confirmed integration at the other two sites by PCR. The distribution of isolates in the H58 tree indicates single integration events at each of the *yidA*, STY4438 and *fbp* loci but numerous independent integrations at the *cyaA* site (**Fig. 6**). The latter suggests that the Tn*2670*-like element, which is flanked by IS*1* sequences, may target existing IS*1* sequences during its mobilization.

In addition, we found that four isolates harbored genes associated with azithromycin resistance. These isolates were 10593_2_14_Alg05-8683_2005 from Algeria carrying *ereA* and three isolates from Indonesia carrying either *msrA* (10349_1_90_Indo404ty_1983) or *msrD* (9953_5_22_IndoA340_2010 and 9953_5_48_IndoA377_2010).

Quinolone resistance in H58

The primary targets of the fluoroquinolones are the DNA gyrase subunits (*gyrA* and *gyrB*) and the topoisomerase IV components (*parC* and *parE*)[34]. Nonsynonymous mutations in the quinolone resistance–determining regions (QRDR) of each gene can decrease susceptibility to fluoroquinolones such as ciprofloxacin, which is commonly used in the treatment of typhoid[35]. Here we found that nonsynonymous changes in the QRDR of these four genes were far more common in the H58 isolates (59%) than in other *S.* Typhi (13%; $P < 1 \times 10^{-6}$, Fisher's exact test; **Supplementary Table 5**). The most frequent QRDR mutations were changes in codon 83 of *gyrA* encoding p.Ser83Phe (45% of H58 isolates) and p.Ser83Tyr (9% of H58 isolates) substitutions (**Supplementary Table 5**). The distribution of *gyrA* substitutions within the H58 phylogeny indicates that these mutations have arisen independently on multiple occasions, consistent with our previous observations of convergent evolution[17,19] and confirming that this region of *gyrA* is under strong positive selection (**Supplementary Fig. 4**). The accumulation of multiple mutations within the *gyr* and *par* genes can result in a higher minimum inhibitory concentration (MIC) for fluoroquinolones[35]. Additionally, we detected multiple mutations in these genes in 199 isolates: 190 H58 isolates (predominantly from Cambodia and India) and 9 non-H58 isolates (mainly from India) (**Supplementary Fig. 4** and **Supplementary Table 5**).

Transmissible fluoroquinolone resistance can occur in *Salmonella* via plasmid-mediated acquisition of *qnr* genes[36]. Here we identified such genes in seven H58 isolates, which carried both *gyrA* mutations and the *qnrS1* gene that confer high-level resistance. The *qnrS1* gene was present within a mobile element also containing *blaTEM-1*, *sul2* and *catB4*, in association with and possibly mobilized by IS*26*, on an IncFIB(K) plasmid (**Fig. 6** and **Supplementary Table 3**).

Trends in antimicrobial resistance in H58

The data show some geographical differences in patterns of antimicrobial resistance within the H58 lineage (**Supplementary Fig. 5**). Multidrug resistance was common among H58 isolates from Southeast Asia in the 1990s, and in recent years *gyrA* mutations have arisen on this background, resulting in high rates of MDR H58 with reduced susceptibility to fluoroquinolones. This observation likely reflects the therapeutic use of fluoroquinolones to treat typhoid over this period. Although we have few examples of H58 isolates from South Asia before 2000, the pattern is clearly different in this region, with the majority of isolates from recent years almost all harboring *gyrA* mutations but with low rates of multidrug resistance[35].

The situation appears to be different yet again in Africa, where the majority of recent isolates (mainly from Malawi, Kenya and South Africa) were identified as MDR but without *gyrA* mutations, potentially reflecting the continued use of traditional antimicrobial agents.

Genomic signatures of the H58 lineage

Because the *S.* Typhi H58 lineage emerged rapidly over the past 30 years, we searched for any distinctive genetic signatures, other than those for multidrug resistance, that might be facilitating its dissemination. Prophage-like elements are known hotspots for variation within *S.* Typhi and other *S. enterica*[37]; however, H58 genomes shared five of the seven prophage-like elements previously identified in the *S.* Typhi reference isolate CT18 (haplotype H1), and only rare acquisitions of new phages were found (five phages, affecting 10% of the H58 isolates; **Supplementary Fig. 6** and **Supplementary Table 6**). The SNPs that define the H58 lineage include several nonsynonymous changes in genes associated with pathogenicity, adaptation and chaperones (**Supplementary Table 7**). The affected genes consist of *ssaP*, encoding a *Salmonella* pathogenicity island–2 (SPI-2)-associated protein involved in intracellular survival and persistence[38], and the regulatory genes *sirA* and *csrB*, which have been implicated in *Salmonella* virulence[39,40]. Additionally, H58-associated SNPs also included changes in genes involved in central or intermediary metabolism, including *trpE*, *rlpB*, *betC*, *rtn*, *metH*, *yfbT*, *pub*, *nuoG* and *iaaA*, and genes involved in membranes or structures, for example, *lip1*, *yhdA*, *kefA*, *kcsA*, *yegT*, *lsrC*, *yajI* and SBOV18161 (*hyaE*)[41]. All *S.* Typhi display substantial genome degradation, via the accumulation of deletions and inactivating mutations within coding sequences, to form pseudogenes; each *S.* Typhi genome has >200 pseudogenes, reflecting a loss of ~4% of protein-coding capacity[19,20]. Here we found that all H58 isolates additionally harbored a point mutation in the *sptP* gene (STY3001) resulting in a premature stop codon at position 185, effectively rendering H58 strains deficient for SptP protein. SptP is an SPI-1 effector protein known to have a role in modulating the host cell actin cytoskeleton via its GAP domain that targets CDC42 and RAC-1 (ref. 42). We previously identified this nonsense mutation in seven sequenced H58 isolates and a second distinct nonsense mutation in the same gene in the H50 isolate E98-3139. Convergent loss-of-function mutations such as this are quite rare in *S.* Typhi and may reflect a selective advantage for inactivation of this gene[20].

DISCUSSION

Here, we provide the first comprehensive global phylogeographical analysis of the emerging MDR-associated *S.* Typhi clade known as H58, covering many of the key geographical regions where typhoid remains endemic. This analysis indicates a major ongoing clonal replacement of resident non-H58 *S.* Typhi haplotypes by this clade and identifies previously unappreciated inter- and intracontinental transmission events. Smaller regional studies performed in different Asian countries and in Kenya have described the emergence of the H58 haplotype at local and country levels[12,14,15,17,19,25]. However, here we show the true global impact of H58, which is transforming the *S.* Typhi population structure across the world. Indeed, we show definitively that H58 has expanded dramatically since the early 1990s and that the MDR phenotype of H58 strains is likely influenced by different regional antibiotic usage[11]. Further, our analyses show an ongoing epidemic of MDR typhoid moving across Africa, potentially driven by this antimicrobial usage. The existence of this epidemic is supported by the available epidemiological data, which include increasing numbers of reports of MDR typhoid in Africa, and by observations from members of this consortium[14,28–30].

The H58 lineage has previously been associated with multidrug resistance, which may be a key factor driving its current expansion[14,25]. Here we show that this association holds across numerous countries in Asia and Africa, such that the majority of the global burden of MDR typhoid can be attributed to the H58 lineage. Intriguingly, our data indicate that multidrug resistance in H58 is tightly linked to the presence of a single Tn*2670*-like element that was probably first introduced via the IncHI1-PST6 plasmid but has since transferred to the *S.* Typhi chromosome in numerous distinct integration events, each affecting different sublineages of H58. Such integrations have recently been noted in isolates from Bangladesh (*cyaA* and *yidA* sites)[15,43] and Zambia (*cyaA* site)[16]; however, our data provide important context for these observations, showing that integrations are relatively frequent and have been occurring since the emergence of H58. Integration of the MDR locus into the chromosome may facilitate loss of the large IncHI1 plasmid, thus moderating any potential fitness burden while maintaining the MDR phenotype. A similar phenomenon has been observed in *Shigella sonnei*, where chromosomal integration of an MDR transposon in the late 1970s appears to be associated with global dissemination of a single successful sublineage[44], and in *Salmonella* that have acquired the *Salmonella* Genomic Island[45]. H58 strains also harbor a higher frequency of quinolone resistance–associated mutations than other *S.* Typhi, potentially owing to enhanced exposure of the large, diversifying and frequently MDR H58 population to fluoroquinolones, coupled with the lack of a fitness cost associated with these mutations[46].

These results provide, to our knowledge, the first global, large-scale, genome-based study of an MDR clade of *S.* Typhi. The global dissemination of H58 requires urgent international attention. Indeed, the arrival of S. Typhi H58 in Africa appears to be transforming the epidemiology of the disease, with MDR outbreaks of typhoid being reported where the disease was previously unappreciated or absent. It will be particularly important to control antimicrobial prescribing practices, including the use of prophylactic antibiotics such as trimethoprim-sulfamethoxazole[47] in this region, as such use likely promotes multidrug resistance. This study highlights the need for longstanding routine surveillance to capture epidemics and monitor changes in bacterial populations as a means to facilitate public health measures, such as the use of effective antimicrobials and the introduction of vaccine programs, to reduce the vast and neglected morbidity and mortality caused by typhoid.

URLs. Sprai, http://zombie.cb.k.u-tokyo.ac.jp/sprai/; SMALT, http://www.sanger.ac.uk/resources/software/smalt/; Path-O-Gen, http://tree.bio.ed.ac.uk/software/pathogen/; Velvet Optimizer, http://www.ebi.ac.uk/~zerbino/velvet/; ISmapper, https://github.com/jhawkey/IS_mapper.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Raw sequence data have been submitted to the European Nucleotide Archive (ENA) under accession ERP001718.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

This work was supported by the Wellcome Trust. We would like to thank the members of the Pathogen Informatics Team and the core sequencing teams at the Wellcome Trust Sanger Institute (Cambridge, UK). We are grateful to D. Harris

for his superb work in managing the sequence data. We also thank L. Fabre for her excellent technical assistance.

This work was supported by a number of organizations. The authors affiliated with the Wellcome Trust Sanger Institute were funded by Wellcome Trust award 098051; N.A.F. was supported by Wellcome Trust research fellowship WT092152MA. N.A.F., R.S.H. and this work were supported by a strategic award from the Wellcome Trust for the Malawi-Liverpool Wellcome Trust Clinical Research Programme (101113/Z/13/Z). C.M.P. was funded by the Wellcome Trust Mahidol University Oxford Tropical Medicine Research Programme, supported by the Wellcome Trust (Major Overseas Programmes–Thailand Unit Core Grant), the European Society for Paediatric Infectious Diseases and the University of Oxford–Li Ka Shing Global Health Foundation. D.D., P.N. and V.D. were supported by the Wellcome Trust (core grant 089275/H/09/Z). K.E.H. was supported by the National Health and Medical Research Council of Australia (fellowship 1061409) and the Victorian Life Sciences Computation Initiative (VLSCI; grant VR0082). C.A.M. was supported by a Clinical Research Fellowship from GlaxoSmithKline, and P.J.H. was supported by a UK Medical Research Council PhD studentship. This work forms part of a European Union Framework Programme 7 Marie Curie Actions Industry Academia Partnerships and Pathways (IAPP) Consortium Programme, entitled GENDRIVAX (Genome-Driven Vaccine Development for Bacterial Infections), involving the Wellcome Trust Sanger Institute, KEMRI Nairobi and the Novartis Vaccines Institute for Global Health. The authors affiliated with the Institut Pasteur were funded by the Institut Pasteur, the Institut de Veille Sanitaire and the French government 'Investissement d'Avenir' program (Integrative Biology of Emerging Infectious Diseases Laboratory of Excellence, grant ANR-10-LABX-62-IBEID). C.H.W. was supported by the UK Medical Research Council (MR/J003999/1). C.O. was supported by Society in Science and the Branco Weiss Fellowship, administered by ETH Zurich. A.K.C. was supported by the UK Medical Research Council (G1100100/1). J.J. was supported by the antibiotic resistance surveillance project in the Democratic Republic of the Congo, funded by project 2.01 of the Third Framework Agreement between the Belgian Directorate General of Development Cooperation and the Institute of Tropical Medicine (Antwerp, Belgium). F.M. was supported by a research grant from the Bill and Melinda Gates Foundation. The findings and conclusions contained within this publication are those of the authors and do not necessarily reflect positions or policies of the Bill and Melinda Gates Foundation. J.A. Crump was supported by the joint US National Institutes of Health–National Science Foundation Ecology and Evolution of Infectious Disease program (R01 TW009237), the UK Biotechnology and Biological Sciences Research Council (BBSRC; BB/J010367/1) and UK BBSRC Zoonoses in Emerging Livestock Systems awards BB/L017679, BB/L018926 and BB/L018845. S.K. was supported by US National Institutes of Health grant R01 AI099525-02. S.B. is a Sir Henry Dale Fellow, jointly funded by the Wellcome Trust and the Royal Society (100087/Z/12/Z). S.O. was supported by the National Institute of Allergy and Infectious Diseases of the US National Institutes of Health (R01 AI097493). The content is solely the responsibility of the authors and does not necessarily represent the official views of the US National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. C.D. was supported by the University of Oxford–Li Ka Shing Global Health Foundation.

AUTHOR CONTRIBUTIONS

Study design: V.K.W., S.B., J. Parkhill , N.R.T., K.E.H. and G.D. Sequencing data generation: A.J.P., J.A.K. and E.J.K. Data analysis: V.K.W., K.E.H., J. Parkhill, N.R.T., A.J.P., J.A.K., D.J.E., J. Hawkey, S.R.H., A.E.M., A.K.C., J. Hadfield, C.O., R.A.K., E.J.K., D.A.G. and D.J.P. Isolate acquisition and processing and clinical data collection: D.J.P., S.B., N.A.F., N.R.T., F.-X.W., P.J.H., N.T.V.T., R.F.B., C.H.W., S.K., M.A.G., R.S.H., J.J., O.L., W.J.E., C.M., J.A. Chabalgoity, M.K., K.J., S. Dutta, F.M., J.C., C.T., S.O., C.A.M., C.D., K.H.K., A.M.S., C.M.P., A.K., E.K.M., J.I.C., S. Dongol, B.B., M.D., D.B., T.T.N., S.P.S., M.H., P.N., R.S.O., L.I., D.D., V.D., G.T., L.W., J.A. Crump, E.D.P., S.N., E.J.N., D.P.T., P.T., S.S., M.V., J. Powling, K.D., G.H., J.F. and K.E.H. Manuscript writing: V.K.W., S.B., K.E.H. and G.D. All authors contributed to manuscript editing. Project oversight: S.B., K.E.H. and G.D.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Parry, C.M., Hien, T.T., Dougan, G., White, N.J. & Farrar, J.J. Typhoid fever. *N. Engl. J. Med.* **347**, 1770–1782 (2002).
2. Connor, B.A. & Schwartz, E. Typhoid and paratyphoid fever in travellers. *Lancet Infect. Dis.* **5**, 623–628 (2005).
3. Crump, J.A. & Mintz, E.D. Global trends in typhoid and paratyphoid fever. *Clin. Infect. Dis.* **50**, 241–246 (2010).
4. Mogasale, V. *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* **2**, e570–e580 (2014).
5. Bodhidatta, L., Taylor, D.N., Thisyakorn, U. & Echeverria, P. Control of typhoid fever in Bangkok, Thailand, by annual immunization of schoolchildren with parenteral typhoid vaccine. *Rev. Infect. Dis.* **9**, 841–845 (1987).
6. Fraser, A., Paul, M., Goldberg, E., Acosta, C.J. & Leibovici, L. Typhoid fever vaccines: systematic review and meta-analysis of randomised controlled trials. *Vaccine* **25**, 7848–7857 (2007).
7. Khan, M.I., Ochiai, R.L. & Clemens, J.D. Population impact of Vi capsular polysaccharide vaccine. *Expert Rev. Vaccines* **9**, 485–496 (2010).
8. Bhutta, Z.A. Current concepts in the diagnosis and treatment of typhoid fever. *Br. Med. J.* **333**, 78–82 (2006).
9. Olarte, J. & Galindo, E. *Salmonella* Typhi resistant to chloramphenicol, ampicillin, and other antimicrobial agents: strains isolated during an extensive typhoid fever epidemic in Mexico. *Antimicrob. Agents Chemother.* **4**, 597–601 (1973).
10. Anderson, E.S. The problem and implications of chloramphenicol resistance in the typhoid bacillus. *J. Hyg. (Lond.)* **74**, 289–299 (1975).
11. Mirza, S.H., Beeching, N.J. & Hart, C.A. Multi-drug resistant typhoid: a global problem. *J. Med. Microbiol.* **44**, 317–319 (1996).
12. Holt, K.E. *et al.* Emergence of a globally dominant IncHI1 plasmid type associated with multiple drug resistant typhoid. *PLoS Negl. Trop. Dis.* **5**, e1245 (2011).
13. Chau, T.T. *et al.* Antimicrobial drug resistance of *Salmonella enterica* serovar Typhi in Asia and molecular mechanism of reduced susceptibility to the fluoroquinolones. *Antimicrob. Agents Chemother.* **51**, 4315–4323 (2007).
14. Kariuki, S. *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.* **48**, 2171–2176 (2010).
15. Chiou, C.S. *et al.* Antimicrobial resistance in *Salmonella enterica* serovar Typhi from Bangladesh, Indonesia, Taiwan and Vietnam. *Antimicrob. Agents Chemother.* **58**, 6501–6507 (2014).
16. Hendriksen, R.S. *et al.* Genomic signature of multi-drug resistant *Salmonella* Typhi related to a massive outbreak in Zambia during 2010 and 2012. *J. Clin. Microbiol.* **53**, 262–272 (2015).
17. Roumagnac, P. *et al.* Evolutionary history of *Salmonella* Typhi. *Science* **314**, 1301–1304 (2006).
18. Kidgell, C. *et al.* *Salmonella* Typhi, the causative agent of typhoid fever, is approximately 50,000 years old. *Infect. Genet. Evol.* **2**, 39–45 (2002).
19. Holt, K.E. *et al.* High-throughput sequencing provides insights into genome variation and evolution in *Salmonella* Typhi. *Nat. Genet.* **40**, 987–993 (2008).
20. Holt, K.E. *et al.* Pseudogene accumulation in the evolutionary histories of *Salmonella enterica* serovars Paratyphi A and Typhi. *BMC Genomics* **10**, 36 (2009).
21. Baker, S. *et al.* High-throughput genotyping of *Salmonella enterica* serovar Typhi allowing geographical assignment of haplotypes and pathotypes within an urban District of Jakarta, Indonesia. *J. Clin. Microbiol.* **46**, 1741–1746 (2008).
22. Holt, K.E. *et al.* High-throughput bacterial SNP typing identifies distinct clusters of *Salmonella* Typhi causing typhoid in Nepalese children. *BMC Infect. Dis.* **10**, 144 (2010).
23. Phan, M.D. *et al.* Variation in *Salmonella enterica* serovar Typhi IncHI1 plasmids during the global spread of resistant typhoid fever. *Antimicrob. Agents Chemother.* **53**, 716–727 (2009).
24. Le, T.A. *et al.* Clonal expansion and microevolution of quinolone-resistant *Salmonella enterica* serotype Typhi in Vietnam from 1996 to 2004. *J. Clin. Microbiol.* **45**, 3485–3492 (2007).
25. Holt, K.E. *et al.* Temporal fluctuation of multidrug resistant *Salmonella* Typhi haplotypes in the Mekong River delta region of Vietnam. *PLoS Negl. Trop. Dis.* **5**, e929 (2011).
26. Baker, S. *et al.* Combined high-resolution genotyping and geospatial analysis reveals modes of endemic urban typhoid fever transmission. *Open Biol.* **1**, 110008 (2011).
27. Holt, K.E. *et al.* High-resolution genotyping of the endemic *Salmonella* Typhi population during a Vi (typhoid) vaccination trial in Kolkata. *PLoS Negl. Trop. Dis.* **6**, e1490 (2012).
28. Lutterloh, E. *et al.* Multidrug-resistant typhoid fever with neurologic findings on the Malawi-Mozambique border. *Clin. Infect. Dis.* **54**, 1100–1106 (2012).
29. Phoba, M.F. *et al.* Multidrug-resistant *Salmonella enterica*, Democratic Republic of the Congo. *Emerg. Infect. Dis.* **18**, 1692–1694 (2012).
30. Breiman, R.F. *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* **7**, e29119 (2012).
31. Drummond, A.J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
32. Wain, J. *et al.* Quinolone-resistant *Salmonella* Typhi in Viet Nam: molecular basis of resistance and clinical response to treatment. *Clin. Infect. Dis.* **25**, 1404–1410 (1997).
33. Holt, K.E. *et al.* Multidrug-resistant *Salmonella enterica* serovar Paratyphi A harbors IncHI1 plasmids similar to those found in serovar Typhi. *J. Bacteriol.* **189**, 4257–4264 (2007).
34. Hooper, D.C. Quinolone mode of action—new aspects. *Drugs* **45** (suppl. 3), 8–14 (1993).
35. Menezes, G.A., Harish, B.N., Khan, M.A., Goessens, W.H. & Hays, J.P. Antimicrobial resistance trends in blood culture positive *Salmonella* Typhi isolates from Pondicherry, India, 2005–2009. *Clin. Microbiol. Infect.* **18**, 239–245 (2012).

36. Keddy, K.H., Smith, A.M., Sooka, A., Ismail, H. & Oliver, S. Fluoroquinolone-resistant typhoid, South Africa. *Emerg. Infect. Dis.* **16**, 879–880 (2010).
37. Thomson, N. *et al.* The role of prophage-like elements in the diversity of *Salmonella enterica* serovars. *J. Mol. Biol.* **339**, 279–300 (2004).
38. Hensel, M. *et al.* Functional analysis of *ssaJ* and the *ssaK/U* operon, 13 genes encoding components of the type III secretion apparatus of *Salmonella* Pathogenicity Island 2. *Mol. Microbiol.* **24**, 155–167 (1997).
39. Teplitski, M., Goodier, R.I. & Ahmer, B.M. Pathways leading from BarA/SirA to motility and virulence gene expression in *Salmonella*. *J. Bacteriol.* **185**, 7257–7265 (2003).
40. Lawhon, S.D. *et al.* Global regulation by CsrA in *Salmonella typhimurium*. *Mol. Microbiol.* **48**, 1633–1645 (2003).
41. Parkhill, J. *et al.* Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* **413**, 848–852 (2001).
42. Fu, Y. & Galan, J.E. A *Salmonella* protein antagonizes Rac-1 and Cdc42 to mediate host-cell recovery after bacterial invasion. *Nature* **401**, 293–297 (1999).
43. Ashton, P.M. *et al.* MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. *Nat. Biotechnol.* **33**, 296–300 (2015).
44. Holt, K.E. *et al. Shigella sonnei* genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe. *Nat. Genet.* **44**, 1056–1059 (2012).
45. Le Hello, S. *et al.* The global establishment of a highly-fluoroquinolone resistant *Salmonella enterica* serotype Kentucky ST198 strain. *Front. Microbiol.* **4**, 395 (2013).
46. Baker, S. *et al.* Fitness benefits in fluoroquinolone-resistant *Salmonella* Typhi in the absence of antimicrobial pressure. *eLife* **2**, e01229 (2013).
47. Bwakura-Dangarembizi, M. *et al.* A randomized trial of prolonged co-trimoxazole in HIV-infected children in Africa. *N. Engl. J. Med.* **370**, 41–53 (2014).
48. Guy, L., Kultima, J.R. & Andersson, S.G. genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* **26**, 2334–2335 (2010).

Vanessa K Wong[1,2], Stephen Baker[3–5], Derek J Pickard[1], Julian Parkhill[1], Andrew J Page[1], Nicholas A Feasey[6], Robert A Kingsley[1,7], Nicholas R Thomson[1,5], Jacqueline A Keane[1], François-Xavier Weill[8], David J Edwards[9], Jane Hawkey[9,10], Simon R Harris[1], Alison E Mather[1], Amy K Cain[1], James Hadfield[1], Peter J Hart[11], Nga Tran Vu Thieu[3], Elizabeth J Klemm[1], Dafni A Glinos[1], Robert F Breiman[12–14], Conall H Watson[15], Samuel Kariuki[1,12], Melita A Gordon[16], Robert S Heyderman[17], Chinyere Okoro[1], Jan Jacobs[18,19], Octavie Lunguya[20,21], W John Edmunds[15], Chisomo Msefula[17,22], Jose A Chabalgoity[23], Mike Kama[24], Kylie Jenkins[25], Shanta Dutta[26], Florian Marks[27], Josefina Campos[28], Corinne Thompson[3,4], Stephen Obaro[29–31], Calman A MacLennan[1,11,32], Christiane Dolecek[4], Karen H Keddy[33], Anthony M Smith[33], Christopher M Parry[34,35], Abhilasha Karkey[36], E Kim Mulholland[5,37], James I Campbell[3,4], Sabina Dongol[36], Buddha Basnyat[36], Muriel Dufour[38], Don Bandaranayake[39], Take Toleafoa Naseri[40], Shalini Pravin Singh[41], Mochammad Hatta[42], Paul Newton[4,43], Robert S Onsare[12], Lupeoletalalei Isaia[44], David Dance[4,43], Viengmon Davong[43], Guy Thwaites[3,4], Lalith Wijedoru[45,46], John A Crump[47], Elizabeth De Pinna[48], Satheesh Nair[48], Eric J Nilles[49], Duy Pham Thanh[3], Paul Turner[4,45,50], Sona Soeng[50], Mary Valcanis[51], Joan Powling[51], Karolina Dimovski[51], Geoff Hogg[51], Jeremy Farrar[3,4], Kathryn E Holt[9,52] & Gordon Dougan[1,52]

[1]Wellcome Trust Sanger Institute, Hinxton, UK. [2]Department of Microbiology, Addenbrooke's Hospital, Cambridge University Hospitals National Health Service (NHS) Foundation Trust, Cambridge, UK. [3]Hospital for Tropical Diseases, Wellcome Trust Major Overseas Programme, Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam. [4]Centre for Tropical Medicine and Global Health, Nuffield Department of Clinical Medicine, Oxford University, Oxford, UK. [5]Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK. [6]Liverpool School of Tropical Medicine, Liverpool, UK. [7]Institute of Food Research, Norwich Research Park, Norwich, UK. [8]Institut Pasteur, Unité des Bactéries Pathogènes Entériques, Paris, France. [9]Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Melbourne, Victoria, Australia. [10]Faculty of Veterinary and Agricultural Sciences, University of Melbourne, Melbourne, Victoria, Australia. [11]Institute of Biomedical Research, School of Immunity and Infection, College of Medicine and Dental Sciences, University of Birmingham, Birmingham, UK. [12]Kenya Medical Research Institute (KEMRI), Nairobi, Kenya. [13]Centers for Disease Control and Prevention, Atlanta, Georgia, USA. [14]Emory Global Health Institute, Atlanta, Georgia, USA. [15]Centre for the Mathematical Modelling of Infectious Diseases, Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK. [16]Institute of Infection and Global Health, University of Liverpool, Liverpool, UK. [17]Malawi-Liverpool Wellcome Trust Clinical Research Programme, College of Medicine, University of Malawi, Blantyre, Malawi. [18]Department of Clinical Sciences, Institute of Tropical Medicine, Antwerp, Belgium. [19]Department of Microbiology and Immunology, Katholieke Universiteit (KU) Leuven, University of Leuven, Leuven, Belgium. [20]National Institute for Biomedical Research, Kinshasa, Democratic Republic of the Congo. [21]University Hospital of Kinshasa, Kinshasa, Democratic Republic of the Congo. [22]Microbiology Department, College of Medicine, University of Malawi, Blantyre, Malawi. [23]Departamento de Desarrollo Biotecnológico, Instituto de Higiene, Facultad de Medicina, Montevideo, Uruguay. [24]Ministry of Health, Suva, Fiji. [25]Fiji Health Sector Support Program, Suva, Fiji. [26]National Institute of Cholera and Enteric Diseases, Kolkata, India. [27]International Vaccine Institute, Department of Epidemiology, Seoul, Republic of Korea. [28]Enteropathogen Division, Administración Nacional de Laboratorios e Institutos de Salud (ANLIS) Carlos G. Malbran Institute, Buenos Aires, Argentina. [29]Division of Pediatric Infectious Diseases, University of Nebraska Medical Center, Omaha, Nebraska, USA. [30]University of Abuja Teaching Hospital, Abuja, Nigeria. [31]Bingham University, Karu, Nigeria. [32]Novartis Vaccines Institute for Global Health, Siena, Italy. [33]Centre for Enteric Diseases, National Institute for Communicable Diseases, Division in the National Health Laboratory Service, University of the Witwatersrand, Johannesburg, South Africa. [34]Department of Clinical Research, London School of Hygiene and Tropical Medicine, London, UK. [35]Graduate School of Tropical Medicine and Global Health, Nagasaki University, Nagasaki, Japan. [36]Patan Academy of Health Sciences, Wellcome Trust Major Overseas Programme, Oxford University Clinical Research Unit, Kathmandu, Nepal. [37]Murdoch Childrens Research Institute, Melbourne, Victoria, Australia. [38]Enteric and Leptospira Reference Laboratory, Institute of Environmental Science and Research, Ltd. (ESR), Porirua, New Zealand. [39]National Centre for Biosecurity and Infectious Disease, Institute of Environmental Science and Research, Porirua, New Zealand. [40]Samoa Ministry of Health, Apia, Samoa. [41]National Influenza Center, World Health Organization, Center for Communicable Disease Control, Suva, Fiji. [42]Department of Microbiology, Hasanuddin University, Makassar, Indonesia. [43]Lao Oxford Mahosot Wellcome Trust Research Unit, Microbiology Laboratory, Mahosot Hospital, Vientiane, Laos. [44]National Health Services, Tupua Tamasese Meaole Hospital, Apia, Samoa. [45]Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand. [46]Paediatric Emergency Medicine, Chelsea and Westminster Hospital, London, UK. [47]Centre for International Health and Otago International Health Research Network, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand. [48]*Salmonella* Reference Service, Public Health England, Colindale, London, UK. [49]Emerging Disease Surveillance and Response, Division of Pacific Technical Support, World Health Organization, Suva, Fiji. [50]Cambodia-Oxford Medical Research Unit, Angkor Hospital for Children, Siem Reap, Cambodia. [51]Microbiological Diagnostic Unit–Public Health Laboratory, Department of Microbiology and Immunology at the Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, Victoria, Australia. [52]These authors contributed equally to this work. Correspondence should be addressed to V.K.W. (vw1@sanger.ac.uk).

ONLINE METHODS

Bacterial isolates and sequencing. A total of 1,832 *S.* Typhi isolates were included in the study. These were isolated between 1905 and 2013 and originated from 63 countries spanning 6 continents (Asia, Africa, North and South America, Europe, and Australia and Oceania) (**Supplementary Table 1**). This collection included 14 of the 19 *S.* Typhi isolates previously sequenced by Holt *et al.*[19], including the 2 isolates with finished reference genomes, CT18 (AL513382) and Ty2 (AE014613) (**Supplementary Table 2**). Seven other public *S.* Typhi references sequences were downloaded from public databases and included in the analysis (**Supplementary Table 2**). Information on the source was available for 1,531 of our *S.* Typhi isolates. The isolates included in our study were supplied by numerous contributing laboratories and were cultured from a wide range of clinical specimens, including blood (90.9%; 1,391/1,531), stool or rectal swab (8.3%; 127/1,531), urine (0.4%; 6/1,531), pus (0.3%; 5/1,531), gallbladder fluid (0.1%; 2/1,531), pleural fluid (0.1%; 1/1,531) and cerebrospinal fluid (0.1%; 1/1,531). There were data on the age of the patient for 826 *S.* Typhi isolates, and these comprised 330 (40%) isolates originating from children (0 to ≤16 years old) and 496 (60%) isolates cultured from adults (>16 years old).

Each of the collaborating laboratories used their own individual methodologies for isolation of whole-genomic DNA. For 231 isolates, DNA was prepared using the Wizard Genomic DNA kit (Promega) according to the manufacturer's instructions. Index-tagged paired-end Illumina sequencing libraries were prepared as previously described[49]. These were combined into pools each containing 96 uniquely tagged libraries and sequenced on the Illumina HiSeq 2000 or HiSeq 2500 platform according to the manufacturer's protocols to generate tagged 100-bp paired-end reads.

In addition, the genomes of five H58 isolates were sequenced on the PacBio RS II platform (Pacific Biosciences) for better resolution of the integration of the large composite transposable element into the chromosome (**Supplementary Table 4**). Genomic DNA (3 µg) was sheared using the HydroShear Plus (Digilab), and a library was prepared using DNA Template Prep Kit 2.0 (Pacific Biosciences), according to the manufacturer's instructions. Sequencing was performed on SMRT cells with XL polymerase and DNA Sequencing Kit C2 (Pacific Biosciences). *De novo* assembly was performed with Sprai v0.9.5 (see URLs) and HGAP v2.1.0 (ref. 50) with default parameters. The contigs from Sprai were circularized with a script in the Sprai package when the script detected a significant overlap between the beginning and end of contigs.

Read alignment and SNP detection. For analysis of SNPs, paired-end Illumina reads were mapped to the CT18 reference genome of *S.* Typhi, including the chromosome and pHCM1 and pHCM2 plasmids[41], using SMALT (version 0.7.4) (see URLs) as previously described[51,52]. Candidate SNPs were identified as previously described[19,53], using SAMtools command mpileup -d 1000 -DSugBf ref bam > results.bcf; bcftools view -cg results.bcf[54]. SNP calls with quality scores above 30 were collated into a single list of variant sites, and the allele at each SNP site in each isolate was determined by reference to the consensus base call for that genome (using SAMtools and removing low-confidence alleles with consensus base quality <50, SNPs contained in <75% of reads and those with mapping quality <30, read depth <4, <2 reads per strand, strand bias $P < 0.001$, mapping bias $P < 0.001$ or tail bias $P < 0.001$). SNPs called in phage regions, repetitive sequences (354 kb; ~7.4% of bases in the CT18 reference chromosome, as defined previously[19]) or recombinant regions (~180 kb; <4% of the CT18 reference chromosome, identified using an approach described previously[49]) were excluded, resulting in a final set of 22,145 chromosomal SNPs identified in an alignment of length 4,275,037 bp.

Phylogenetic analysis. The maximum-likelihood phylogenetic tree shown in **Figure 1b** was built from the 22,145-SNP alignment of all 1,832 isolates, plus a *S.* Paratyphi A strain included as an outgroup for tree rooting, using RAxML (version 7.8.6)[55] with the generalized time-reversible model and a Gamma distribution to model site-specific rate variation (the GTR+Γ substitution model; GTRGAMMA in RAxML). Support for the maximum-likelihood phylogeny was assessed via 100 bootstrap pseudoanalyses of the alignment data. A maximum-likelihood phylogenetic tree was also inferred separately from the SNP alignment of 853 H58 *S.* Typhi isolates using the same parameters as above (**Fig. 2**). The H58 phylogenetic tree was rooted using an *S.* Typhi isolate from the nearest neighboring cluster of non-H58 isolates (isolate Fij107364). All maximum-likelihood trees were displayed and annotated using iTOL[56,57]. To simplify visualization of the H58 phylogeny (**Fig. 3a**), clades containing only isolates from a single country were collapsed manually in R. Geographical transitions were inferred from this collapsed H58 tree using discrete trait transition modeling, implemented in the make.simmap function in the phytools R package. Briefly, each genome was assigned to a geographical region on the basis of country of isolation (or presumed region of inoculation for travel-associated isolates), these regions were treated as discrete tip states on the H58 maximum-likelihood tree and a Markov model for the evolution of this state (i.e., geographical region) was fitted to the tree, as proposed in ref. 58. Entries in the resulting transition matrix were interpreted as likely geographical transfers between regions, drawn as arrows on the map in **Figure 4** (directionality inferred from visual confirmation of the phylogeny).

Temporal analysis. To investigate temporal signal in the maximum-likelihood phylogeny for *S.* Typhi, we used Path-O-Gen (see URLs) to extract root-to-tip dates and analyzed their linear relationship with year of isolation using R. To assess the robustness of the H58 temporal signal, analysis of the H58 subtree was repeated 100 times with randomly permuted tip dates. The evolutionary dynamics of the H58 lineage were investigated via Bayesian analysis with BEAST (v1.6)[31]. Initial analyses were conducted on 114 isolates from across the H58 maximum-likelihood tree, covering the full temporal and geographical range of H58. Bias in isolate distribution was reduced by selecting a maximum of eight isolates from each geographical location, with as much temporal diversity as possible within that location. For countries represented by fewer than eight H58 isolates, all isolates at that location were included in the BEAST analysis. The concatenated SNP alignments of these 114 isolates were subjected to multiple BEAST analyses using both constant population size and Bayesian skyline models of changes in population size, in combination with either a strict molecular clock or a relaxed clock (uncorrelated lognormal distribution), to identify the model that best fitted the data[44]. For the BEAST analysis, the GTR+Γ substitution model was selected, and tip dates were defined as the year of isolation. For all model combinations, 3 independent chains of 100 million generations each were run to ensure convergence, with sampling every 1,000 iterations. The 3 runs were combined with LogCombiner[31], following removal of the first 10 million steps from each as burn-in. In all cases, the relaxed, (uncorrelated lognormal) clock model, which allows evolutionary rates to vary among the branches of the tree together with the skyline demographic model, proved a much better fit for the data (Bayes factor > 200).

For the final analyses reported here, ten independent runs were conducted with different samples of isolates in the alignment but using identical substitution (GTR+Γ), clock (uncorrelated, lognormal relaxed) and demographic (Bayesian skyline) models. Each permutation contained a different set of randomly selected isolates from India, Pakistan, Nepal, Cambodia, Vietnam, Tanzania, Kenya, Laos, Malawi, Bangladesh, Iraq and Fiji. The countries of Sri Lanka, Thailand, South Africa, Afghanistan, Indonesia, Myanmar, Lebanon, Palestine and Australia had five or fewer isolates, and all these isolates were therefore included in each of the ten permutation data sets. Each set of isolates was subjected to triplicate BEAST runs. Maximum–clade credibility (MCC) trees were generated using TreeAnnotator[31]. Estimates reported as median values with 95% HPDs and posterior probability values (PPVs) were used as support for the identification of ancestral nodes and their associated geographical locations. The Bayesian skyline plot was calculated and visualized using Tracer (v1.6), to investigate changes in the effective population size of the H58 lineage over time[31]. The effective sample sizes (ESSs) of the parameters were estimated to be >200 for all 10 independent runs of the analysis.

Gene content analysis. The reads for each isolate were assembled *de novo* using the short-read assembler Velvet[59] with parameters optimized with Velvet Optimizer (see URLs) to provide the highest N50 value. Contigs that were less than 300 bp long were excluded from further analysis. The assemblies were constructed and annotated using Prokka[60] by the Pathogen Informatics team at the Wellcome Trust Sanger Institute (Cambridge, UK) using an automated pipeline.

The *de novo*–assembled contig sets were mapped iteratively to the pan-genome reference set (initialized as the concatenation of the *S.* Typhi CT18

chromosome and plasmids) using MUMmer (nucmer algorithm)[61] as previously described[19]. At each iteration stage i, sequences not aligning to the current pan-genome P_{i-1} set were incorporated into an extended pan-genome, P_i. The final pan-genome P was annotated using both annotation transfer (for *S.* Typhi reference sequences) and *de novo* annotation with Prokka[60]. Paired-end reads were then aligned to the pan-genome using bwa[62] with default mapping parameters. SAMtools[54] was used to produce a pileup for each aligned read set, and this was used to summarize, for each annotated gene in the pan-genome P, the coverage (percentage of bases covered) and the presence of inactivating mutations (nonsense SNPs or non-triplet insertions and/or deletions (indels) resulting in frameshifts) in each genome. The annotated pan-genome was specifically examined for the acquisition in H58 of phage sequences and plasmids of defined incompatibility groups (see "Plasmid analyses"). New phages identified in the pan-genome study were confirmed using analysis of high-quality *de novo*–assembled contigs with the web server PHAST (Phage Search Tool)[63].

Resistance gene analysis. Acquired antimicrobial resistance genes were detected, and their precise alleles were determined, using the mapping-based allele typer SRST2 (ref. 64) together with the ARG-Annot database[65]. SRST2 was also used to identify mutations in the *gyrA*, *gyrB*, *parC* and *parE* genes that have been associated with resistance in Gram-negative bacteria (including *Salmonella*) to quinolones[34,35,66–68]. Note, however, that our data were not suitable for assessing the presence of azithromycin resistance–related mutations in the 23S rRNA sequences, which have been occasionally reported in other bacterial pathogens[69–71]. The horizontal transfer of resistance genes associated with a transposon from the IncHI1 plasmid into the chromosome of H58 isolates was analyzed using a combination of the data from the Illumina and PacBio sequencing, inspected using a combination of tools, including Genome Browser Artemis and Artemis Comparison Tool (ACT)[72], which allowed comparison of the genome assemblies against finished sequences for IncHI1 plasmids (pAKU1 (ref. 33) and pHCM1 (ref. 41)) and the CT18 chromosome[41]. IS*1* insertion sites were investigated using ISmapper (see URLs), which uses bwa[62] to map reads to the IS*1* sequence and identify those reads that flank the IS (defined as those reads that do not map within the IS but whose paired reads do). These IS*1*-flanking reads were then mapped to the CT18 chromosome reference sequence, to identify the sites of IS*1* insertion in the chromosome.

Plasmid analyses. Presence of the IncHI1 plasmid was confirmed by (i) BLASTN search of contig sets with the sequence of conserved backbone genes (164.1 kb) and antimicrobial resistance genes within Tn*10* and a composite transposon, Tn*2067*, generated from comparative analysis of the nucleotide sequences of R27 (AF250878), pHCM1 (AL513383) and pAKU1 (AM412236)[33] (**Supplementary Table 2**) using ACT[72], and (ii) detection of an IncHI1 replicon directly from reads using SRST2 and the PlasmidFinder database (v 1.2)[73]. IncHI1 plasmid multilocus sequence typing (MLST) types[23] were also determined using SRST2 (ref. 64). H58 *S.* Typhi isolates that contained fewer than 38% (63/168) of the backbone genes were excluded from further analysis. Additional plasmid replicons were identified from the SRST2 analysis of plasmid replicons, and the location of resistance genes was determined by manual investigation of the assemblies using BLASTN, Artemis and ACT[72].

H58 SNP repertoire analysis. Nonsynonymous SNPs that were identified in >99% of the H58 isolates and in no other lineage were analyzed. The SNPs and their associated genes were studied using Genome Browser Artemis[72]. Functional categories were as annotated in the *S.* Typhi CT18 genome (AL513382.1).

Statistical methods. Simple, descriptive statistics were used to compare the geographical distributions of lineages, prevalence of plasmids, and resistance genes and mutations and to calculate 95% confidence intervals. The significance of differences between studied groups of variables was calculated using Fisher's exact test. All statistical tests were two-sided at $\alpha = 0.05$, and analyses were performed using STATA (version 12.1, StataCorp) and R.

49. Croucher, N.J. *et al.* Rapid pneumococcal evolution in response to clinical interventions. *Science* **331**, 430–434 (2011).
50. Chin, C.S. *et al.* Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **10**, 563–569 (2013).
51. Casali, N. *et al.* Evolution and transmission of drug-resistant tuberculosis in a Russian population. *Nat. Genet.* **46**, 279–286 (2014).
52. Harris, S.R. *et al.* Whole-genome analysis of diverse Chlamydia trachomatis strains identifies phylogenetic relationships masked by current clinical typing. *Nat. Genet.* **44**, 413–419 (2012).
53. Harris, S.R. *et al.* Evolution of MRSA during hospital transmission and intercontinental spread. *Science* **327**, 469–474 (2010).
54. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
55. Stamatakis, A. RAxML-VI-HPC: maximum likelihood–based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).
56. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128 (2007).
57. Letunic, I. & Bork, P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.* **39**, W475–W478 (2011).
58. Bollback, J.P. SIMMAP: stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics* **7**, 88 (2006).
59. Zerbino, D.R. & Birney, E. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829 (2008).
60. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
61. Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12 (2004).
62. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
63. Zhou, Y., Liang, Y., Lynch, K.H., Dennis, J.J. & Wishart, D.S. PHAST: a fast phage search tool. *Nucleic Acids Res.* **39**, W347–W352 (2011).
64. Inouye, M. *et al.* SRST2: rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med.* **6**, 90 (2014).
65. Gupta, S.K. *et al.* ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob. Agents Chemother.* **58**, 212–220 (2014).
66. Eaves, D.J. *et al.* Prevalence of mutations within the quinolone resistance–determining region of *gyrA*, *gyrB*, *parC*, and *parE* and association with antibiotic resistance in quinolone-resistant *Salmonella enterica*. *Antimicrob. Agents Chemother.* **48**, 4012–4015 (2004).
67. Baucheron, S., Chaslus-Dancla, E., Cloeckaert, A., Chiu, C.H. & Butaye, P. High-level resistance to fluoroquinolones linked to mutations in *gyrA*, *parC*, and *parE* in *Salmonella enterica* serovar Schwarzengrund isolates from humans in Taiwan. *Antimicrob. Agents Chemother.* **49**, 862–863 (2005).
68. Song, Y. *et al.* A multiplex single nucleotide polymorphism typing assay for detecting mutations that result in decreased fluoroquinolone susceptibility in *Salmonella enterica* serovars Typhi and Paratyphi A. *J. Antimicrob. Chemother.* **65**, 1631–1641 (2010).
69. Haanperä, M., Huovinen, P. & Jalava, J. Detection and quantification of macrolide resistance mutations at positions 2058 and 2059 of the 23S rRNA gene by pyrosequencing. *Antimicrob. Agents Chemother.* **49**, 457–460 (2005).
70. Gomes, C. *et al.* *In vitro* development and analysis of *Escherichia coli* and *Shigella boydii* azithromycin-resistant mutants. *Microb. Drug Resist.* **19**, 88–93 (2013).
71. Marvig, R.L. *et al.* Mutations in 23S rRNA confer resistance against azithromycin in *Pseudomonas aeruginosa*. *Antimicrob. Agents Chemother.* **56**, 4519–4521 (2012).
72. Carver, T. *et al.* Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* **24**, 2672–2676 (2008).
73. Carattoli, A. *et al.* *In silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* **58**, 3895–3903 (2014).

nature COMMUNICATIONS

ARTICLE

Received 23 Dec 2015 | Accepted 4 Aug 2016 | Published 5 Oct 2016

DOI: 10.1038/ncomms12827 OPEN

An extended genotyping framework for *Salmonella enterica* serovar Typhi, the cause of human typhoid

Vanessa K. Wong[1,2], Stephen Baker[3,4,5], Thomas R. Connor[6], Derek Pickard[1], Andrew J. Page[1], Jayshree Dave[7], Niamh Murphy[7], Richard Holliman[7], Armine Sefton[8], Michael Millar[8], Zoe A. Dyson[9,10], Gordon Dougan[1], Kathryn E. Holt[9,10], International Typhoid Consortium[†]

The population of *Salmonella enterica* serovar Typhi (*S.* Typhi), the causative agent of typhoid fever, exhibits limited DNA sequence variation, which complicates efforts to rationally discriminate individual isolates. Here we utilize data from whole-genome sequences (WGS) of nearly 2,000 isolates sourced from over 60 countries to generate a robust genotyping scheme that is phylogenetically informative and compatible with a range of assays. These data show that, with the exception of the rapidly disseminating H58 subclade (now designated genotype 4.3.1), the global *S.* Typhi population is highly structured and includes dozens of subclades that display geographical restriction. The genotyping approach presented here can be used to interrogate local *S.* Typhi populations and help identify recent introductions of *S.* Typhi into new or previously endemic locations, providing information on their likely geographical source. This approach can be used to classify clinical isolates and provides a universal framework for further experimental investigations.

[1] The Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK. [2] Addenbrooke's Hospital, Cambridge University Hospitals NHS Foundation Trust, Cambridge Biomedical Campus, Hills Road, Cambridge CB2 0QQ, UK. [3] The Hospital for Tropical Diseases, Wellcome Trust Major Overseas Programme, Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam. [4] Centre for Tropical Medicine and Global Health, Nuffield Department of Clinical Medicine, Oxford University, Oxford, UK. [5] Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK. [6] Cardiff University School of Biosciences, Cardiff University, Cardiff, UK. [7] Public Health Laboratory London, Public Health England, London, UK. [8] Division of Infection, Barts Health NHS Trust, London, UK. [9] Centre for Systems Genomics, University of Melbourne, Parkville, Victoria 3010, Australia. [10] Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria 3010, Australia. Correspondence and requests for materials should be addressed to V.K.W. (email: vw1@sanger.ac.uk)
†A full list of consortium members appears at the end of the paper.

Typhoid fever (typhoid), caused by *Salmonella enterica* serovar Typhi (*S.* Typhi) bacteria, is a systemic human infection that affects an estimated 20.6 million people globally each year, causing an estimated 223,000 deaths[1–3]. Typhoid remains endemic in populations with limited access to sanitation and safe water, and is a notifiable or reportable infection in many industrialized countries, where it is generally associated with travel to endemic areas. Public health laboratories have relied on techniques such as phage typing[4,5] or pulsed-field gel electrophoresis[6], which are phylogenetically naive and have limited discriminatory power to support epidemiological investigations and surveillance.

A genotyping scheme based on 88 single-nucleotide polymorphisms (SNPs) identified within a limited set of genes was previously developed for *S.* Typhi[7]. This enabled the classification of the *S.* Typhi population into 85 haplotypes (haploid genotypes) based on biallelic profiles and provided the first phylogenetic framework for epidemiological studies[8]. Subsequently, whole-genome sequencing (WGS) has been used to identify many more SNPs and other phylogenetically informative markers for discriminating within *S.* Typhi, which has limited genetic variation[9–16]. Similar progress has been made in other monophyletic clades of bacterial pathogens, such as *Mycobacterium tuberculosis*[17] and *Yersinia pestis*[18].

We have recently reported the WGS of almost 2,000 *S.* Typhi isolates sourced from 63 countries[14]. This study identified >22,000 chromosomal SNPs in the core genome, which were used to build a comprehensive phylogenetic tree. Notably, the analysis confirmed the emerging dominance of the multidrug resistance-associated H58 clade, including the recent spread of H58 *S.* Typhi into Africa, confirming the value of SNP-based WGS analysis of *S.* Typhi to understand contemporary typhoid epidemiology. Here we utilize these WGS data to define a global population framework for *S.* Typhi and to define a new genotyping scheme comprising 68 SNPs that provides extensive coverage of typhoid-causing bacteria circulating globally. Given the increasingly widespread adoption of WGS by public health laboratories for the tracking of bacterial pathogens[19,20], we further aimed to explore the utility of *S.* Typhi WGS data, analysed via genotyping, to predict the geographical source of travel-associated *S.* Typhi isolated in the United Kingdom. This approach gives greater discriminatory power and improved phylogenetic information than the earlier scheme[7], and forms a robust framework for public health surveillance, epidemiological investigations and laboratory experiments of typhoid.

Results

Defining phylogenetically informative genotypes for *S.* Typhi. In order to develop a comprehensive genotyping system, we used WGS data from >1,800 globally representative *S.* Typhi[14] to identify phylogenetically informative clades and subclades based on SNP architecture[21]. A summary of the isolates is shown in Table 1 and full details are provided in Supplementary Data 1 and Supplementary Table 1. Using a combination of phylogenetic tree topology and population genetic methods (using BAPS; Bayesian Analysis of Population Structure[21]), we defined 16 *S.* Typhi clades that could be further divided into 49 subclades (Fig. 1, see Methods). Most of the clades could be grouped into four nested clusters (1–4, which we refer to as 'primary clusters'), each with 100% bootstrap support and defined by >20 SNPs (coloured branches in Fig. 1a). The median pairwise distances between isolates were as follows: 25 SNPs within subclades, 109 SNPs within clades and 243 SNPs between clades. We labelled these primary clusters, clades and subclades using a structured hierarchical nomenclature system similar to that used for *M. tuberculosis*[17], whereby cluster 1 is subdivided into clades 1.1 and 1.2; clade 1.1 is further subdivided into subclades 1.1.1, 1.1.2, 1.1.3 and so on (see Fig. 1b, Methods). An interactive version of the global phylogeny, with strains labelled by genotype, country of origin and year of isolation, is available at http://microreact.org/project/styphi[22].

Under the new genotype nomenclature, the globally disseminated multidrug resistant clone commonly referred to as H58 (which actually includes haplotype H58 and eight other H58-derived haplotypes under the original Roumagnac *et al.* scheme[7]), constitutes a single subclade (4.3.1). No other subclades were identified within clade 4.3. The CT18 reference genome (H1 in Roumagnac *et al.* scheme) belonged to subclade 3.2.1, while the laboratory strain Ty2 and its attenuated mutant BRD948 (H10 under the Roumagnac scheme) belonged to clade 4.1 (with no further differentiation to subclade level by BAPS). The backbone of the minimum spanning tree of Roumagnac *et al.* haplotypes was broadly consistent with the backbone structure of the whole-genome phylogeny (Supplementary Fig. 1a). However, mapping the Roumagnac haplotypes to the whole-genome phylogeny showed that the older scheme provides highly uneven resolution across the *S.* Typhi phylogeny (Supplementary Fig. 1b), with a lack of resolution in some cases (11 Roumagnac haplotypes span two or more distinct subclades each; for example, H52 comprises clades 3.4, 3.5, 4.1 and 4.2) and excessive resolution in others (24 subclades are further divided into two or more haplotypes in the Roumagnac scheme).

A new SNP-based genotyping framework for *S.* Typhi. We identified a minimum set of 68 SNPs that can be used to genotype *S.* Typhi into the four primary clusters, 16 clades and 49 subclades. For each of these groups, we identified all SNPs that were unique to members of the group, and selected one such SNP to be used for genotyping. We prioritized the inclusion of synonymous intragenic SNPs (that is, located within a protein-coding sequence, but with no change to the encoded amino acid), within genes that showed evidence of genetic stability within the *S.* Typhi population (that is, nucleotide diversity <1% and dN/dS <0.7 across the global data set, with no inactivating mutations identified). Details of the genotyping SNPs are given in Supplementary Table 2. This genotyping scheme has greater discriminatory power than the original Roumagnac haplotyping scheme ($D = 0.96$ versus 0.78), is phylogenetically informative by design and the hierarchical nomenclature of genotypes is intrinsically informative with respect to phylogenetic relationships between clades and subclades.

Geographical distribution of *S.* Typhi clades and subclades. Next, we examined the geographical distribution of *S.* Typhi genotypes. For these analyses, isolates of the same subclade, country and year were collapsed to a single representative to reduce the impact of localized outbreaks on our collection; this resulted in 541 unique isolates for analysis. Primary clusters 2, 3 and 4 were broadly distributed across continents (greens, blues and reds, respectively, in Fig. 1c), likely reflecting the relatively ancient spread of *S.* Typhi across the globe. Isolates outside these clusters, which result from deep branching closer to the root of the *S.* Typhi whole-genome tree, were rare in our collection ($n = 24$ unique isolates) and mostly found in Africa ($n = 16$). While the three common clusters (2–4) were present in most regions we analysed, cluster 2 predominated among American isolates ($n = 18/23$ unique isolates, 78%). Most clades were detected on multiple continents ($n = 11/16$) and included isolates from Asia ($n = 13/16$) and/or Africa ($n = 10/16$), which together

Table 1 | Summary of 1,831 *S.* Typhi.

Continent (*region*)	Country of origin ($n \geq 5$)	Range of isolation dates (years)	Number of isolates
Asia		*1972–2012*	*1068 (58.3%)*
Southeast Asia		1976–2012	719
	Cambodia	2007–2012	210
	Indonesia	1976–2012	129
	Laos	2000–2010	138
	Vietnam	1972–2011	221
	Malaysia	2005–2011	6
	Other (Philippines, Thailand, East Timor, Myanmar)	2002–2012	15
South Asia		1977–2012	321
	Bangladesh	1998–2012	51
	India	1977–2012	174
	Nepal	1999–2012	47
	Pakistan	2003–2012	45
	Other (Sri Lanka, Afghanistan)	2001–2012	4
Western Asia		1997–2011	25
	Iraq	2006–2011	11
	Lebanon	2001–2011	7
	Other (Armenia, Palestine, Turkey, Western Asia)	1997–2011	7
Eastern Asia		2002–2011	3
	Other (China)	2002–2011	3
Africa		*1958–2013*	*374 (20.4%)*
North Africa		1961–2009	24
	Algeria	1999–2009	7
	Morocco	1999–2000	9
	Other (Sudan, Egypt, Tunisia)	1961–2008	8
East Africa		1980–2010	115
	Kenya	1998–2009	56
	Tanzania	2006–2010	52
	Other (Comoros, Madagascar)	1980–2002	7
Central Africa		1958–2011	49
	Cameroon	1958–2009	27
	DRC	1976–2011	17
	Other (Angola, Central African Republic)	2001–2009	5
West Africa		1998–2009	30
	Other (Burkina Faso, Cape Verde, Benin, Guinea, Ivory Coast, Gabon, Liberia, Mali, Niger, Nigeria, Mauritania, Senegal, Togo)	1998–2009	30
Southern Africa		2004–2013	153
	Malawi	2004–2013	112
	South Africa	2004–2012	41
Africa		2009–2012	3
	Unknown	2009–2012	3
Europe		*1916–2009*	*7 (0.4%)*
Eastern Europe		1916–1996	2
	Other (Russia)	1916–1996	2
Western Europe		2009	4
	Other (France (suspected African origin of infection))	2009	4
Southern Europe		2009	1
	Other (Malta)	2009	1
Australia and Oceania		*1980–2012*	*342 (18.7%)*
Australia		2010–2012	3
	Australia	2010–2012	3
Oceania		1980–2012	339
	Fiji	1981–2012	170
	Samoa	1992–2012	117
	Papua New Guinea	1980–2012	47
	Other (Tonga, Vanuatu)	1980–2003	5
North America		1958–2011	5 (0.3%)
	Other (USA, Mexico)	1958–2011	5
Central America		2012	1 (0.05%)
	Other (El Salvador)	2012	1
South America		1905–2012	17 (0.9%)
	Argentina	1905–2006	10
	Other (French Guiana, Peru, South America)	2002–2012	6
Unknown origin		1939–2012	19 (1.0%)

S. Typhi, *Salmonella enterica* serovar Typhi.
Typhi isolates from the global collection, which were used to define genotypes. Countries with fewer than five isolates were grouped into the category 'Other'; *n* indicates the number of such countries in each region.

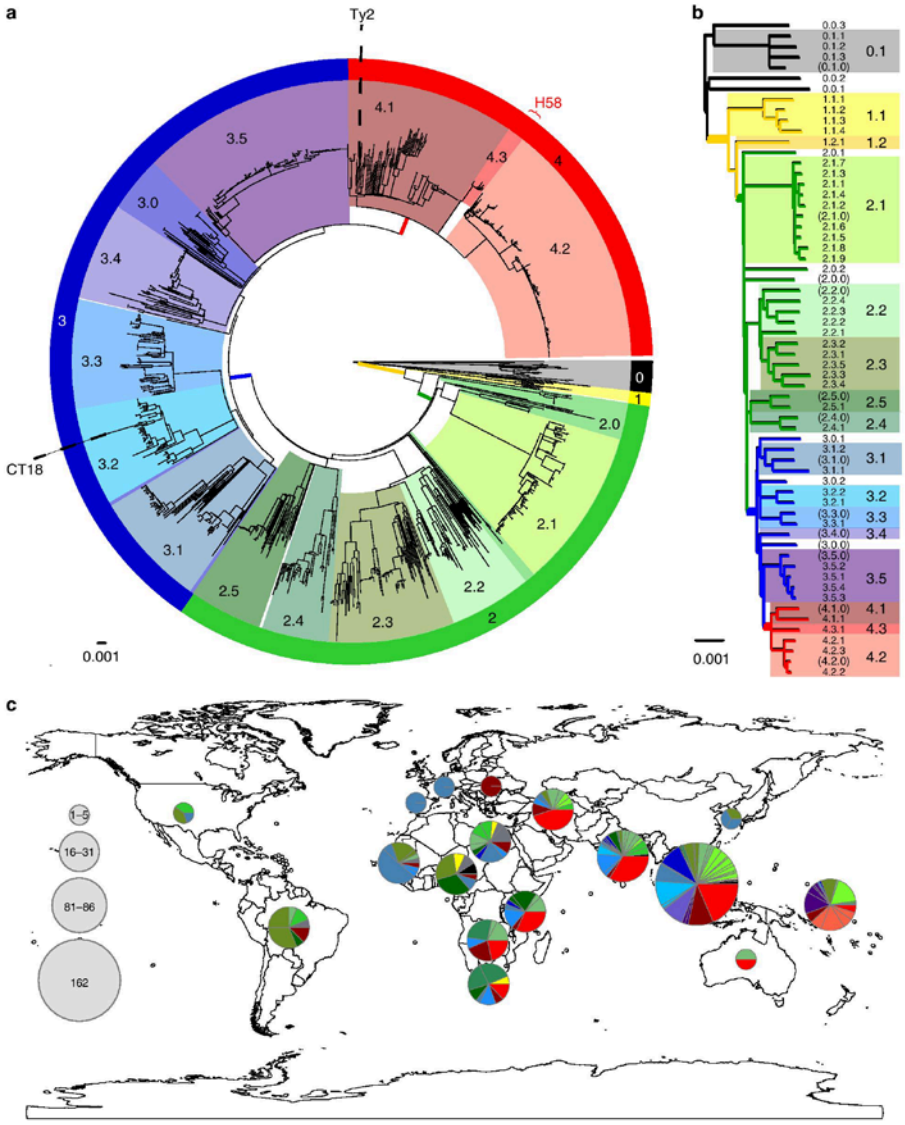

Figure 1 | Population structure of *S. Typhi* based on genome-wide SNPs. (**a**) Whole-genome tree of 1,831 global *S.* Typhi isolates. Primary clusters 1–4 are indicated in the outer coloured ring; branches defining these groups are coloured in the tree. These groups are further divided into clades, which are shaded and labelled. The location of *S.* Typhi reference genomes CT18 (accession number AL513382) and Ty2 (accession number AE014613) are indicated on the tree. Subclade 4.3.1 (H58, marked in red), which comprises half of the global collection, is represented by just 50 (6%) randomly selected isolates out of the total 852 belonging to this subclade, so that the relationships between other clades can be visualized. (**b**) Tree backbone showing further division of 16 *S.* Typhi clades (shaded) into 49 subclades (labelled; note 12 undifferentiated clade groups shown in brackets). Branches are coloured by primary cluster. (**c**) Map of the world showing subclade diversity of *S.* Typhi isolates in the global collection, by region[22]. Where groups of isolates from the same country and year belonged to the same subclade, this was classified as an 'outbreak' and the group is only represented once in the pie graphs. Pies are sized to indicate number of isolates; slices are coloured by clade; multiple slices of the same colour indicate multiple subclades belonging to the same clade.

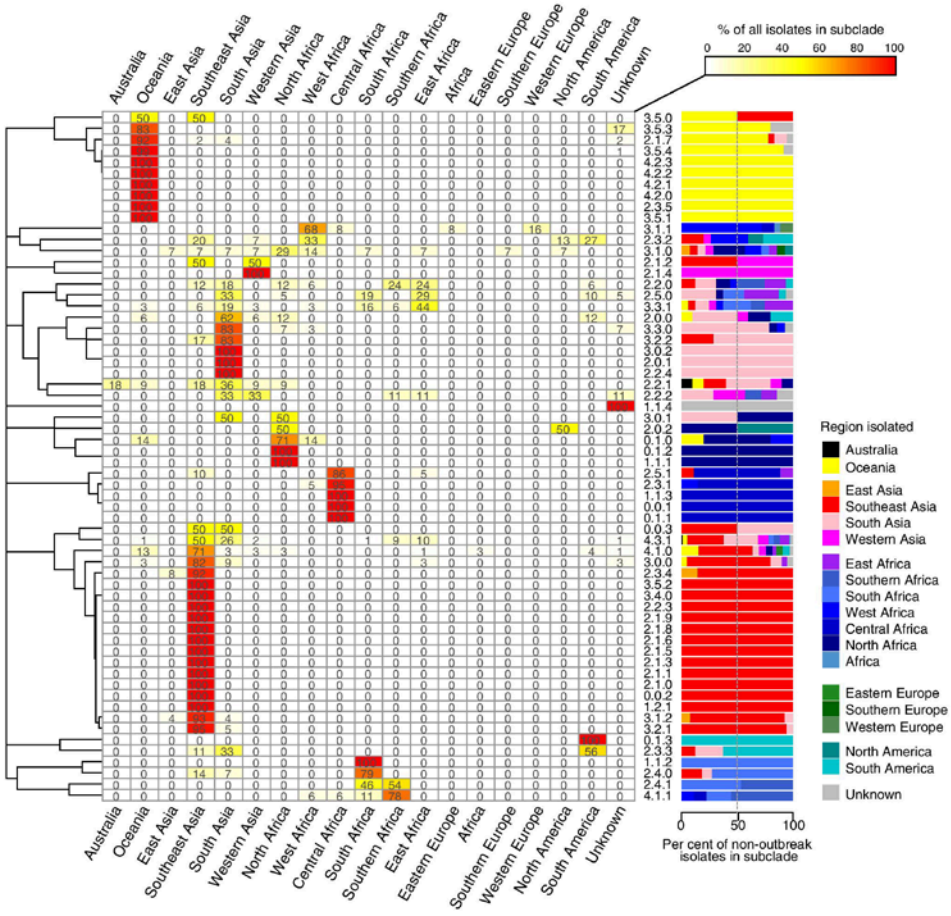

Figure 2 | Geographical clustering of *S.* Typhi subclades. Heatmap shows, for each subclade, the percentage of unique isolates originating from each of the geographical regions. Where groups of isolates from the same country and year belonged to the same subclade, this was classified as an 'outbreak' and the group is only represented once. The same data are represented as a scaled bar graph to the right. The full list of isolates by country and subclade is provided in Supplementary Data 1.

made up 78% of our isolate collection (Table 1). However, there were differences in the geographic distributions of clades, with most clades being dominated by unique isolates from a single continent (Asia, Africa or Oceania; see Supplementary Fig. 2).

In contrast, at the subclade level, only 22% of subclades ($n = 11$) were found on more than one continent, and most were dominated by unique isolates from a single country or region: 40 subclades (82%) had $\geq$50% of non-outbreak isolates from a single country (Fig. 2) and 44 subclades (90%) had $\geq$50% of non-outbreak isolates from a single region (Fig. 2). A total of 28 subclades comprised five or more non-outbreak isolates each, and of these common subclades, 12 (43%) were detected in a single region only (six in Oceania, five in Southeast Asia and one in South Asia; Fig. 2). In total, 16 common subclades (57%) were highly restricted to a region (>90% of isolates drawn from a single region) and 20 (71%) were generally associated with one region (>70% of isolates drawn from a single region; Fig. 2). These data suggest that most *S.* Typhi subclades represent localized bacterial subpopulations with barriers to geographical dispersion, and that transfers to new locations rarely result in long-term establishment of local populations. In contrast with this general pattern, subclade 4.3.1 (previously H58) was found in nine different regions across Africa, Asia and Oceania. Only 10 other subclades (20%) were found on more than one continent, and the majority of these were dominated either by Asian, African or Oceanian isolates (Fig. 2). Thus, the recent global dissemination of subclade 4.3.1, which spread out of South Asia ~30 years ago and has established successful local clonal

expansions in dozens of countries[14], likely represents a comparatively rare event in the evolutionary history of *S.* Typhi.

Genomic prediction of the geographical origins of *S.* Typhi by comparison with the global framework. Since most *S.* Typhi subclades were associated with a narrow geographical source, we hypothesized that genotyping of *S.* Typhi isolates could be used to predict the likely geographical origins of typhoid cases. As this is clearly challenging for the more widely distributed subclades, we also sought to examine whether specific SNPs could be used to predict origins down to the country level. For 1,501 out of 1,831 (82%) isolates in our global collection, the genetically closest isolate was from the same country. Where the closest isolate was 0–1 SNPs away, this frequency was 95% and for <10 SNPs, 90% (Supplementary Fig. 3).

Since our current global genome collection includes groups of isolates that were frequently collected from the same time and place, this should not be taken as a reliable measure of the general predictive power of SNP distance for *S.* Typhi. In order to further explore the power of our global genomic framework to predict geographic origins of travel-associated typhoid, we sequenced and genotyped 99 novel *S.* Typhi that were isolated from patients attending a hospital in East London, United Kingdom between 2005 and 2010 (Table 2). A total of 13 genotypes were identified. Epidemiological interviews were able to link 81 of these cases with travel to a specific country; the remaining 18 cases were not associated with travel. The median SNP distance between these novel isolates and genomes in our global collection was 21 SNPs (interquartile range, 18–25 SNPs), posing a challenge for prediction of their geographical origin. Among the 81 travel-associated UK isolates, 53 were genotyped as 4.3.1; these were all linked to travel to countries within South Asia (Table 3), and clustered along with South Asian isolates from the global collection (Fig. 3 and Supplementary Fig. 4n). For the 28 non-4.3.1 travel-associated UK isolates, the location of travel generally matched the geographical origin of the closest isolate (in terms of number of SNPs) in the global collection: travel location and closest global isolate source matched at the region level in all cases, and at the country level in most cases ($n = 20$, 71%). That is, prediction of geographical origin based on the

Table 2 | Summary of 99 East London travel-associated *S.* Typhi isolates used in the study.

Country of origin	Range of isolation dates (year)	Number of isolates
Bangladesh	2006–2012	38
India	2006–2012	22
Pakistan	2006–2012	13
Nepal	2007	1
India/Pakistan*	2008	1
India/Kuwait*	2008	2
Bangladesh/India*	2010	2
Nigeria	2009	1
Ghana	2007	1
No known travel	2005–2011	18
Total	2005–2012	99

S. Typhi, *Salmonella enterica* serovar Typhi.
The country of origin and range of isolation dates (years) for the isolates are described.
*For five patients, multiple countries of travel were recorded, and it was not possible to confirm in which country the *S.* Typhi infection originated.

Table 3 | Summary of genotyping and SNP results for travel-associated *S.* Typhi isolates with known country of travel.

Subclade	Country of travel	*N* (travel)	Closest genome country match	*N* (global)	Region frequencies (excluding outbreaks)
2.0.1	Bangladesh	1	1	9	*South Asia (100%)
2.1.7	India	2	2	49	Oceania (92%)
2.2.0	Pakistan	1	1	17	Southern Africa (24%) East Africa (24%) South Asia (18%)
2.2.2	India	1	1	9	South Asia (33%) Western Asia (33%)
2.3.3	Bangladesh	2	2	9	South America (56%) South Asia (33%)
3.1.1	Nigeria, Ghana	2	0	25	*West Africa (68%)
3.2.2	Bangladesh, Pakistan	3	2	12	*South Asia (83%)
3.3.0	Bangladesh, Pakistan, India	14	13	30	*South Asia (83%)
3.3.1	Pakistan	1	1	32	East Africa (44%) South Asia (19%)
4.1.0	India	1	1	78	Southeast Asia (71%)
4.3.1	Bangladesh, India, Pakistan, Nepal, Kuwait	53	—	853	Southeast Asia (50%) South Asia (26%)

Closest genome country match, number of travel-associated isolates whose country of travel matched that of the closest genome in the global collection (based on lowest number of SNPs); *N* (global), number of isolates in the global collection that were assigned to this subclade; *N* (travel), number of travel-associated isolates that were assigned to the subclade; Region frequencies, frequency of each geographic region among isolates of this subclade from the global collection (note groups of isolates from the same subclade, country and year were classified as outbreaks and represented only once per group in the frequency calculations); SNP, single-nucleotide polymorphism; *S.* Typhi, *Salmonella enterica* serovar Typhi.
*Highlights the most frequent region for this subclade among the global collection, where this matches the region of travel.

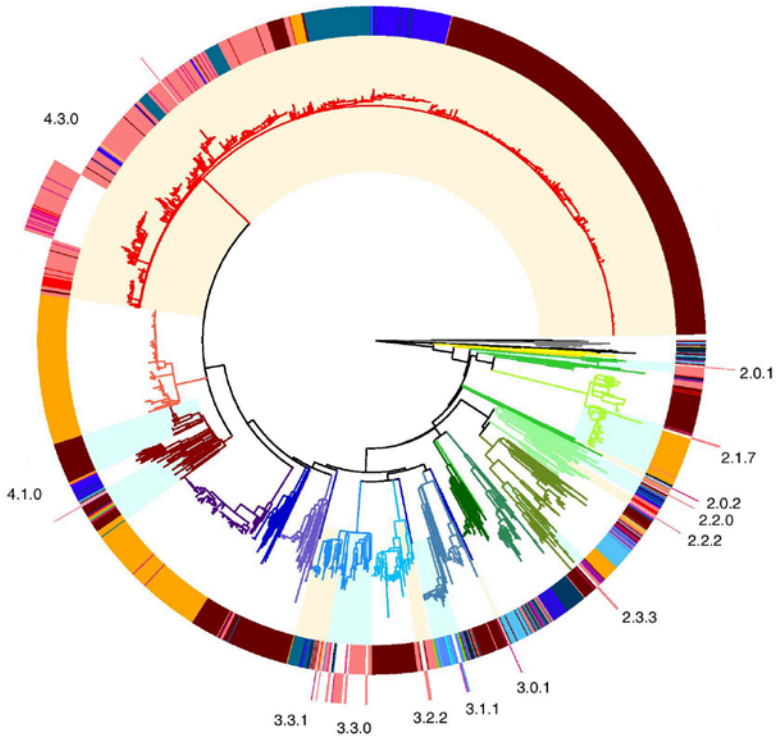

Figure 3 | Phylogeny of 99 travel-associated *S.* Typhi in comparison with the global genomic framework containing 1,831 isolates. Whole-genome SNP tree is shown in the centre and branches are coloured by clade. Rings indicate region of origin: inner ring, global collection; outer ring, travel-associated isolates. Subclades that contain travel-associated isolates are highlighted within the tree (shaded in alternating colours) and labelled around the outside; intrasubclade phylogenies are provided in Supplementary Fig. 4.

closest strain of known location in the current global framework would have yielded the correct region of origin in all cases, and the correct country of origin in 71% of cases (95% confidence interval (CI), 66–76%). Furthermore, for non-4.3.1 subclades, genotyping alone was predictive of geographical origin at the regional level for the same proportion of isolates (71%).

It is likely that power to predict the geographical sources of UK isolates would be improved by wider geographical coverage in the reference genome collection. Two of these isolates were genotyped as subclade 3.1.1 and linked with travel to Ghana and Nigeria; the closest isolates in our global collection were 16–17 SNPs away and were not from these precise locations, but likely originated from bordering countries in West Africa (Supplementary Fig. 4e). It is likely that a deeper coverage of West African isolates in our global framework would provide greater power to resolve geographic associations within this region, which comprised less than 2% of our current global collection ($n = 30$ isolates). Similarly, for the other travel-associated isolates for which the recorded country of travel did not match the closest genome in the global tree, the closest genome was also from a neighbouring country (for example, Pakistan, India, Bangladesh; see Supplementary Fig. 4).

Genomic predictions of the geographical origins of 18 non-travel-associated UK isolates are shown in Table 4 and Supplementary Fig. 4. Thirteen isolates were 4.3.1 and clustered together with travel-associated isolates from South Asia, within a broader group of South Asian 4.3.1 isolates (Fig. 3). This suggests that *S.* Typhi imported into the United Kingdom from these regions have likely been transmitted onwards within the United Kingdom to individuals with no recent travel history (Supplementary Fig. 4n). Two additional isolates were from subclades that were dominated by a single region in our global collection—3.1.1 (68% West Africa) and 3.3.0 (83% South Asia). Notably, while the 4.3.1 isolates were closely related to travel-associated isolates recently obtained in London, they were ≥ 17 SNPs away from any isolates in the global collection. Thus, the diversity captured by the global collection does not provide the resolution to precisely identify the origin of these isolates[23].

Discussion

Our data show that the global *S.* Typhi population consists of 49 distinct subclades that are strongly geographically clustered, with many locations harbouring subpopulations of *S.* Typhi established over long periods of time. We show how these

Table 4 | Summary of genotyping and SNP results for travel-associated *S.* Typhi isolates of unknown origin.

Isolate	Country of closest SNP match	Distance (#SNPs)	Subclade	Subclade distribution
H06434426	Mexico	146	2.0.2	North America (50%) North Africa (50%)
H06156550	Pakistan	92	3.0.1	South Asia (50%) North Africa (50%)
H05272442	Ghana	17	3.1.1	*West Africa (68%)
H09176223	Bangladesh	11	3.3.0	*South Asia (83%)
H10182335	India	18	3.3.1	East Africa (44%) South Asia (19%)
H10046338	Bangladesh†	13	4.3.1	Southeast Asia (50%) South Asia (26%)
H05406403	Bangladesh†	4		
H06136379	Bangladesh†	19		
H06136380	Bangladesh†	17		
H05118260	Bangladesh/India†	19		
H10382491	India†	9		
H10394694	India†	9		
H06016481	India†	9		
H05196407	India†	14		
H05196408	India†	12		
H05212226	India†	17		
H11372598	Pakistan†	10		
H09266336	Pakistan†	9		

SNP, single-nucleotide polymorphism; *S.* Typhi, *Salmonella enterica* serovar Typhi.
For each London isolate, the closest isolate in the global collection was determined (closest = smallest SNP distance, that is, smallest number of core genome SNPs); the country and SNP distance are recorded.
*Highlights the most frequent region for this subclade among the global collection, where this matches the region of the closest isolate in the global collection.
†Location of closest travel-associated isolates from London (unresolvable beyond 'South Asia' based on the global collection alone).

subclades can be identified through a simple genotyping scheme consisting of 68 SNPs. Importantly, while we show that this scheme is highly phylogenetically informative, it can be readily inferred from raw sequence data without the need for multiple genome comparisons, phylogenetic analysis or any other complex or computationally intensive steps. Such properties make this universal SNP-based system a valuable tool upon which researchers can develop future studies. The *S.* Typhi genome is highly stable and exhibits minimal genetic variation and virtually no recombination[9,14], and we recently estimated the substitution rate to be slower than one SNP per genome per year[14]; therefore, the genotyping framework is expected to be robust to future evolution.

Owing to the strong geographical clustering of the various subclades, whole-genome comparison of novel *S.* Typhi isolates to the existing global population framework is strongly predictive of geographic origin at the regional level and has the potential to accurately predict origins to the country level. This has important public health implications for typhoid surveillance and control in endemic and non-endemic areas; however, ongoing updates to the global genomic framework will be important to ensure the utility of genomic surveillance for typhoid. For example, we found that the origin of travel-associated 4.3.1 isolates could not be resolved using the prior global framework alone, but benefitted from updated information provided by other recent travel-associated isolates of known geographical origin. This illustrates the importance of expanding and updating the global genomic framework through sequencing of novel isolates and suggests that, while ongoing surveillance in endemic areas is undoubtedly important, the use of clinically well-characterized travel-associated organisms isolated in non-endemic countries may also provide a valuable source for improving the granularity of data in the framework for genome-based surveillance of *S.* Typhi[23]. In addition, it will be important to expand the current global framework to include more recent isolates (the most recent in our current collection was from 2013) as well as isolates from regions that are currently under-represented (including Africa, the Americas and northeast Asia).

WGS-equipped reference laboratories provide a highly accessible source to expand the global genomic framework for typhoid, with potential benefits to local but also global typhoid control. For example, in England, Wales and Northern Ireland ~520 typhoid cases are reported annually to the national reference laboratory (Public Health England). These cases are investigated in order to determine whether they are associated with travel to typhoid endemic regions[23]. However, approximately one-fifth of typhoid cases in the United Kingdom cannot be traced to a country of origin. At present, Public Health England provides molecular typing, which since April 2015 includes WGS as well as antimicrobial susceptibility profiling, for *S.* Typhi isolated from such cases. The resulting data are considered important for local epidemiology. However, we propose that this could also serve as a proxy for informal surveillance of typhoid molecular epidemiology in endemic regions. This may prove particularly valuable when supported by our genotyping framework for simplified attribution.

Methods

Bacterial isolates and WGS. A total of 1,930 *S.* Typhi isolates were analysed in this study (Supplementary Data 1), including a collection of 1,831 globally distributed isolates contributed by members of the International Typhoid Consortium[14] and 99 novel *S.* Typhi isolated in East London, UK. *S.* Typhi comprising the global collection were isolated between 1905 and 2013 and originate from 65 countries spanning six continents (Asia, Africa, North and South America, Europe, and Australia and Oceania) as previously described[14].

An additional 99 novel *S.* Typhi isolates were obtained from returning travellers with a febrile illness who presented at The Royal London Hospital, Barts Health NHS Trust in East London, UK, between 2005 and 2012. Travel history, available for 81 of the travellers, included visits to seven countries within the continents of Asia and Africa. DNA was extracted using the Wizard Genomic DNA Kit (Promega, Madison, WI, USA) as per the manufacturer's instructions. Index-tagged paired end Illumina sequencing libraries were prepared as previously described[24]. These were combined into pools, each containing 96 uniquely tagged libraries, and were sequenced on the Illumina Hiseq2500 platform (Illumina, San

Diego, CA, USA) according to the manufacturer's protocols to generate tagged 100 base pair (bp) paired-end reads.

SNP analysis. For analysis of SNPs, the paired-end reads were mapped to the reference genome of *S.* Typhi CT18 (ref. 25), using SMALT (version 0.7.4; http://www.sanger.ac.uk/resources/software/smalt/). SNPs were identified as previously described[14], using *samtools mpileup*[26] and filtering with a minimum mapping quality of 30 and a quality ratio cutoff of 0.75 (ref. 24). SNPs located within phage regions, repetitive sequences or recombinant regions were excluded as previously described[14], resulting in a final set of 22,143 chromosomal SNPs in an alignment length of 4,275,037 bp for the global collection of 1,831 *S.* Typhi isolates. An expanded alignment comprising 22,673 SNPs from *S.* Typhi isolates from the global collection (1,831) plus 99 traveller-associated UK isolates was generated using the same procedures as above. Pairwise SNP distances between isolates (that is, the number of core genome SNP loci at which pairs of isolates had discordant alleles) were extracted from each alignment *i* using the *ape* package[27] for R (v3.2; function call: *dist.dna(i,model = "N",pairwise.deletion = T)*).

Phylogenetic analyses. The maximum likelihood (ML) phylogenetic tree shown in Fig. 1 was built from the 22,143-SNP alignment of all 1,831 isolates using RAxML (version 7.8.6)[28] with the generalized time-reversible model and a Gamma distribution to model site-specific rate variation (the GTR + γ substitution model; GTRGAMMA in RAxML). The tree was outgroup-rooted by including a pseudo-sequence comprising *S. Paratyphi* A alleles in the alignment. Support for the ML phylogeny was assessed via 100 bootstrap pseudo-analyses of the alignment data.

The backbone topology of the global ML tree, showing relationships between subclades (Fig. 1b), was recovered by randomly selecting one isolate from each subclade to retain, and removing all other tips from the tree (using *drop.tips()* in the *ape* package[27] for R (v3.2)). A ML phylogenetic tree was also generated separately from 22,673 SNPs of *S.* Typhi isolates from the global collection (1,831) plus the 99 East London traveller-associated isolates, using the same procedures as above. All ML trees were visualized and annotated using Python (https://github.com/katholt/plotTree/#python-code).

Identification of phylogenetically informative clades and subclades. In addition to the whole-genome phylogenetic analysis outlined above, we investigated the population structure of the global *S.* Typhi collection using a phylogeny-free population genetics approach, implemented in BAPS v.6.0 (ref. 21). Hierarchal clustering analyses were conducted on identified clusters until single-member clusters were obtained, thus allowing the discovery of nested genetic population structures[21]. Ten nested levels of molecular variation were fitted to the data using 10 independent runs of the stochastic optimization algorithm with the *a priori* upper bound of the number of clusters varying over the interval 50–300 across the runs[30].

As our goal was to identify genotypes that were both phylogenetically and epidemiologically informative, we explored the homogeneity (1-Simpson's diversity) of geographical source within BAPS clusters (as an indicator of the potential power of genotyping to identify geographical origin of travel-associated isolates) at different levels of clustering (Supplementary Fig. 5). This showed that within-cluster homogeneity increased up to the sixth level of clustering and then reached a plateau, with deeper clustering providing no greater resolution of geographical origin (Supplementary Fig. 5). The third level of clustering resulted in most clusters being dominated by a single continent (14/17 clusters with >80% of isolates from one continent), while sixth-level clustering resulted in most clusters containing isolates from a single country (60/89 clusters with >80% of isolates from one country; Supplementary Fig. 6). We therefore used the BAPS clusters to guide the definition of clades (BAPS level 3) and subclades (BAPS level 6).

In order to maintain compatibility with the phylogeny, some minor modifications of the raw BAPS clusters were required (this consisted of subdividing some BAPS clusters and merging others, but not reassigning members between clusters; see Supplementary Fig. 7). The modified level-3 BAPS clusters were designated 'clades' and were assigned labels of the form [x].[y], where [x] indicates to which major cluster each clade belongs and [y] designates sister clades within each major cluster. The modified level-6 BAPS clusters were designated 'subclades' and assigned labels of the form [x].[y].[z], where [x].[y] indicate to which clade each subclade belongs and [z] designates sister subclades within each clade. Thus, genotype names indicate relationships between genotypes; for example, 2.1.1 and 2.1.2 are sister subclades within clade 2.1, while 2.2.1 is a member of the distinct clade 2.2.

Some BAPS clusters were polyphyletic and consisted of isolates belonging to rare phylogenetic lineages whose common ancestor in the phylogenetic tree coincided with the common ancestor of an entire clade ($n = 9$) or primary cluster ($n = 2$). These groups contain isolates that, given increased numbers, may emerge as distinct BAPS clusters that form sister taxa within the parent clade (or primary cluster), and were thus designated [z] = 0 (or [y] = 0) to indicate non-equivalence with the properly differentiated sister clades ($n = 16$) or subclades ($n = 49$). For example, while the genotypes 2.1 and 2.2 represent distinct sister clades that are each monophyletic, isolates assigned to 2.0 are paraphyletic and include multiple lineages that could not be further subdivided by BAPS analysis (Supplementary Fig. 7).

Subclade 4.3.1, which is the only subclade of Clade 4.3, corresponds to the group referred to as H58, based on the haplotyping scheme of Roumagnac *et al.* in which it is defined by the presence of a single SNP *glpA*-C1047T (position 2,348,902 in *S.* Typhi CT18, BiP33 (ref. 7)). BAPS clustering at any level could not further subdivide subclade 4.3.1 (H58).

SNP-based genotyping. We identified a minimum set of 68 SNPs with which to rapidly genotype *S.* Typhi into the 16 clades and 49 subclades, as described above (Supplementary Table 2). Short read alignment (BAM) files, generated by mapping Illumina reads to the CT18 reference genome (accession AL513382), were used to assign genotypes for each novel read set using a custom Python script (available at https://github.com/katholt/genotyphi). Briefly, the script uses *samtools mpileup* to extract from each BAM file the consensus base calls at the SNP loci. The resulting variant call format file is then processed to identify the presence of cluster-, clade- and/or subclade-defining SNP alleles (defined in Supplementary Table 2) that pass a minimum quality threshold (default consensus base Phred score ≥20) and uses these to assign the read set to a cluster, clade and subclade. Discriminatory power was calculated using the method outlined in ref. 31.

Data availability. Raw sequence data are available in the European Nucleotide Archive under accession ERP001718. Supplementary Data 1 lists accession numbers for each isolate. The software for Microreact interactive tree viewer is available at: http://microreact.org/project/styphi[22]. SMALT is available at: http://www.sanger.ac.uk/resources/software/smalt/. Python script to visualize and annotate trees is available at https://github.com/katholt/plotTree/#python-code. Python script to call SNPs is downloadable at https://github.com/katholt/genotyphi.

References

1. Parry, C. M., Hien, T. T., Dougan, G., White, N. J. & Farrar, J. J. Typhoid fever. *N. Engl. J. Med.* **347,** 1770–1782 (2002).
2. Crump, J. A. & Mintz, E. D. Global trends in typhoid and paratyphoid Fever. *Clin. Infect. Dis.* **50,** 241–246 (2010).
3. Mogasale, V. *et al.* Burden of typhoid fever in low-income and middle-income countries: a systematic, literature-based update with risk-factor adjustment. *Lancet Global Health* **2,** e570–e580 (2014).
4. Sen, B. *et al.* Phage typing, biotyping & antimicrobial resistance profile of *Salmonella enterica* serotype Typhi from Kolkata. *Indian J. Med. Res.* **125,** 685–688 (2007).
5. Kim, S. *et al.* Clustering analysis of *Salmonella enterica* serovar Typhi isolates in Korea by PFGE, ribotying, and phage typing. *Foodborne Pathog. Dis.* **6,** 733–738 (2009).
6. Dutta, S. *et al.* Antimicrobial resistance, virulence profiles and molecular subtypes of *Salmonella enterica* serovars Typhi and Paratyphi A blood isolates from Kolkata, India during 2009-2013. *PLoS ONE* **9,** e101347 (2014).
7. Roumagnac, P. *et al.* Evolutionary history of *Salmonella Typhi*. *Science* **314,** 1301–1304 (2006).
8. Baker, S. *et al.* High-throughput genotyping of *Salmonella enterica* serovar Typhi allowing geographical assignment of haplotypes and pathotypes within an urban District of Jakarta, Indonesia. *J. Clin. Microbiol.* **46,** 1741–1746 (2008).
9. Holt, K. E. *et al.* High-throughput sequencing provides insights into genome variation and evolution in *Salmonella* Typhi. *Nat. Genet.* **40,** 987–993 (2008).
10. Holt, K. E. *et al.* Temporal fluctuation of multidrug resistant *Salmonella* Typhi haplotypes in the Mekong river delta region of Vietnam. *PLoS Negl. Trop. Dis.* **5,** e929 (2011).
11. Holt, K. E. *et al.* High-throughput bacterial SNP typing identifies distinct clusters of *Salmonella* Typhi causing typhoid in Nepalese children. *BMC Infect. Dis.* **10,** 144 (2010).
12. Baker, S. *et al.* Combined high-resolution genotyping and geospatial analysis reveals modes of endemic urban typhoid fever transmission. *Open Biol.* **1,** 110008 (2011).
13. Holt, K. E. *et al.* Emergence of a globally dominant IncHI1 plasmid type associated with multiple drug resistant typhoid. *PLoS Negl. Trop. Dis.* **5,** e1245 (2011).
14. Wong, V. K. *et al.* Phylogeographical analysis of the dominant multidrug-resistant H58 clade of *Salmonella* Typhi identifies inter- and intracontinental transmission events. *Nat. Genet.* **47,** 632–639 (2015).
15. Pham Thanh, D. *et al.* Identification of *Salmonella enterica* serovar Typhi genotypes by use of rapid multiplex ligation-dependent probe amplification. *J. Clin. Microbiol.* **51,** 2950–2958 (2013).
16. Baltazar, M. *et al.* Multidrug-resistant *Salmonella enterica* serotype Typhi, Gulf of Guinea Region, Africa. *Emerg. Infect. Dis.* **21,** 655–659 (2015).
17. Coll, F. *et al.* A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat. Commun.* **5,** 4812 (2014).
18. Morelli, G. *et al. Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. *Nat. Genet.* **42,** 1140–1143 (2010).

19. Holmes, A. *et al.* The utility of whole genome sequencing of *Escherichia coli* O157 for outbreak detection and epidemiological surveillance. *J. Clin. Microbiol.* **53,** 3565–3573 (2015).
20. Gardy, J. L. *et al.* Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N. Engl. J. Med.* **364,** 730–739 (2011).
21. Corander, J., Marttinen, P., Siren, J. & Tang, J. Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics* **9,** 539 (2008).
22. Aanensen, D. M. *et al.* Whole-genome sequencing for routine pathogen surveillance in public health: a population snapshot of invasive *Staphylococcus aureus* in Europe. *MBio* **7 (2016).**
23. Public Health England. *Health Protection Report: Enteric fever surveillance quarterly report (England, Wales and Northern Ireland): first quarter 2015 Infection reports,* Volume 9 Number 16, Published on 8 May 2015.
24. Croucher, N. J. *et al.* Rapid pneumococcal evolution in response to clinical interventions. *Science* **331,** 430–434 (2011).
25. Parkhill, J. *et al.* Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* **413,** 848–852 (2001).
26. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
27. Paradis, E., Claude, J. & Strimmer, K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20,** 289–290 (2004).
28. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22,** 2688–2690 (2006).
29. Corander, J. & Tang, J. Bayesian analysis of population structure based on linked molecular information. *Math Biosci.* **205,** 19–31 (2007).
30. Casali, N. *et al.* Evolution and transmission of drug-resistant tuberculosis in a Russian population. *Nat. Genet.* **46,** 279–286 (2014).
31. Hunter, P. R. Reproducibility and indices of discriminatory power of microbial typing methods. *J. Clin. Microbiol.* **28,** 1903–1905 (1990).

Acknowledgements

We thank the members of the Pathogen Informatics Team and the core sequencing teams at the Wellcome Trust Sanger Institute (Cambridge, UK). We are grateful to David Harris for his work in managing the sequence data. We also thank David Aanensen and Silvia Argimon for use of the data visualization tool Microreact.

This work was supported by a number of organizations. The Wellcome Trust Sanger Institute authors were funded by Wellcome Trust Award #098051. V.K.W. was supported by the Wellcome Trust (#098051) and the National Institute of Health Research (NIHR) Cambridge Biomedical Research Centre (BRC). N.F. was supported by the Wellcome Trust Research Fellowship #WT092152MA. N.F., R.S.H. and this work were supported by a strategic award from the Wellcome Trust for the MLW Clinical Research Programme (#101113/Z/13/Z). C.P. was funded by The Wellcome Trust Mahidol University Oxford Tropical Medicine Research Programme, supported by the Wellcome Trust of Great Britain (Major Overseas Programmes—Thailand Unit Core Grant), the European Society for Paediatric Infectious Diseases and University of Oxford-Li Ka Shing Global Health Foundation. D.D., P.N. and V.D. were supported by the Wellcome Trust (core grant #089275/H/09/Z). Z.A.D. was supported by the Wellcome Trust (Strategic award #106158). K.E.H. was supported by the NHMRC of Australia (fellowship #1061409) and the Victorian Life Sciences Computation Initiative (VLSCI; grant #VR0082). C.A.M. was supported by a Clinical Research Fellowship from GlaxoSmithKline and PJH by a UK Medical Research Council PhD studentship. This work forms part of an EU FP7 Marie Curie Actions Industry Academia Partnerships and Pathways (IAPP) Consortium Programme, entitled GENDRIVAX (Genome-driven vaccine development for bacterial infections), involving the Wellcome Trust Sanger Institute, KEMRI Nairobi and Novartis Vaccines Institute for Global Health. The Institut Pasteur (IP) authors were funded by the IP, the Institut de Veille Sanitaire and by the French Government 'Investissement d'Avenir' programme (Integrative Biology of Emerging Infectious Diseases Laboratory of Excellence, grant #ANR-10-LABX-62-IBEID). C.H.W. was supported by the UK Medical Research Council (MRC; #MR/J003999/1). C.O. was supported by Society in Science, The Branco Weiss Fellowship, administered by the ETH Zurich. A.K.C. was supported by the MRC (#G1100100/1). J.J. was supported by the antibiotic resistance surveillance project in DR Congo, funded by Project 2.01 of the Third Framework Agreement between the Belgian Directorate General of Development Cooperation and the Institute of Tropical Medicine, Antwerp, Belgium. F.M. was supported by a research grant from the Bill and Melinda Gates Foundation. J.A.C. was supported by the joint US National Institutes of Health-National Science Foundation Ecology and Evolution of Infectious Disease program (#R01 TW009237) and the UK Biotechnology and Biological Sciences Research Council (BBSRC; #BB/J010367/1), and by UK BBSRC Zoonoses in Emerging Livestock Systems awards #BB/L017679, #BB/L018926 and #BB/L018845. S.K. was supported by the NIH Grant Number R01 AI099525-02. S.B. is a Sir Henry Dale Fellow, jointly funded by the Wellcome Trust and the Royal Society (#100087/Z/12/Z). S.O. was supported by the National Institute Of Allergy And Infectious Diseases of the National Institutes of Health (#R01AI097493). C.D. was supported by the University of Oxford-Li Ka Shing Global Health Programme. A.E.M. was supported by a Biotechnology and Biological Sciences Research Council award (#BB/M014088/1). P.T. was supported by the Wellcome Trust of Great Britain (Major Overseas Programmes—Thailand Unit Core Grant) and University of Oxford-Li Ka Shing Global Health Foundation.

Author contributions

Study design: V.K.W., S.B., K.E.H. and G.D. Sequencing data generation: A.J.P. and J.A.K. Data analysis: V.K.W., K.E.H., T.R.C., Z.A.D. and A.J.P. Isolate acquisition and processing and clinical data collection: D.J.P., S.B., N.F., N.R.T., F-X.W., P.J.H., N.T.V.T., R.F.B., C.H.W., S.K., M.G., R.S.H., J.J., O.I., W.J.F., C.M., J.A.C., M.K., K.J., S.D., F.M., J.C., C.T., S.O., C.A.M., C.D., K.H.K., A.M.S., C.M.P., A.K., E.K.M., J.I.C., S.D., B.B., M.D., D.B., T.N.T., S.P.S., M.H., P.N., R.S.O., L.I., D.D., V.D., G.T., L.W., J.A.C., E.D.P., S.N., E.J.N., D.P.T., P.T., S.S., M.V., J.P., K.D., G.H., J.F., J.D., R.H., N.M., A.S., M.M. and K.E.H. Manuscript writing: V.K.W., K.E.H. and G.D. All authors contributed to manuscript editing. Project oversight: K.E.H. and G.D.

Disclaimer

The findings and conclusions contained within this publication are those of the authors and do not necessarily reflect positions or policies of the Bill and Melinda Gates Foundation, and the content do not necessarily represent the official views of the National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Wong, V.K. *et al.* An extended genotyping framework for *Salmonella enterica* serovar Typhi, the cause of human typhoid. *Nat. Commun.* 7:12827 doi: 10.1038/ncomms12827 (2016).

Members of the International Typhoid Consortium

Julian Parkhill[1], Nicholas A Feasey[11], Robert A Kingsley[1,12], Nicholas R Thomson[1,5], Jacqueline A Keane[1], François-Xavier Weill[13], Simon Le Hello[13], Jane Hawkey[9,10,14], David J Edwards[9,10], Simon R Harris[1], Amy K Cain[1], James Hadfield[1], Peter J Hart[15], Nga Tran Vu Thieu[3], Elizabeth J Klemm[1], Robert F Breiman[16,17,18], Conall H Watson[19], W John Edmunds[19], Samuel Kariuki[1,16], Melita A Gordon[20,21], Robert S Heyderman[21,22], Chinyere Okoro[1,2], Jan Jacobs[23,24], Octavie Lunguya[25,26], Chisomo Msefula[21,27], Jose A Chabalgoity[28], Mike Kama[29], Kylie Jenkins[30], Shanta Dutta[31], Florian Marks[32], Josefina Campos[33], Corinne Thompson[3,4],

Stephen Obaro[34,35,36], Calman A MacLennan[1,15,37], Christiane Dolecek[3,4], Karen H Keddy[38], Anthony M Smith[38], Christopher M Parry[39,40], Abhilasha Karkey[41], Sabina Dongol[41], Buddha Basnyat[41], Amit Arjyal[41], E Kim Mulholland[5,42], James I Campbell[3,4], Muriel Dufour[43], Don Bandaranayake[44], Take N Toleafoa[45], Shalini Pravin Singh[46], Mochammad Hatta[47], Paul N Newton[4,48], David Dance[4,48], Viengmon Davong[48], Robert S Onsare[16], Lupeoletalalelei Isaia[49], Guy Thwaites[3,4], Lalith Wijedoru[50,51], John A Crump[52], Elizabeth De Pinna[53], Satheesh Nair[53], Eric J Nilles[54], Duy Pham Thanh[3], Paul Turner[4,50,55], Sona Soeng[55], Mary Valcanis[56], Joan Powling[56], Karolina Dimovski[56], Geoff Hogg[56], Jeremy Farrar[3,4], Alison E Mather[57], Ben Amos[58]

[11]Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK; [12]Institute of Food Research, Norwich Research Park, Colney, Norwich NR4 7UA, UK; [13]Institut Pasteur, Unité des Bactéries Pathogènes Entériques, Paris, France; [14]Faculty of Veterinary and Agricultural Sciences, University of Melbourne, Parkville, Victoria 3052, Australia; [15]Institute of Biomedical Research, School of Immunity and Infection, College of Medicine and Dental Sciences, University of Birmingham, Birmingham, UK; [16]Kenya Medical Research Institute (KEMRI), PO Box 43640-00100, Nairobi, Kenya; [17]Centers for Disease Control and Prevention, 1600 Clifton Road, Atlanta, Georgia 30329-4027, USA; [18]Emory Global Health Institute, 1599 Clifton Road, NE 1599-001-1AH, Atlanta, Georgia 30322, USA; [19]Centre for the Mathematical Modelling of Infectious Diseases, Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK; [20]Institute of Infection and Global Health, University of Liverpool L69 7BE, UK; [21]Malawi-Liverpool-Wellcome-Trust Clinical Research Programme, College of Medicine, University of Malawi, PO Box 30096, Chichiri, Blantyre 3, Malawi; [22]Division of Infection and Immunity, University College London, London, UK; [23]Department of Clinical Sciences, Institute of Tropical Medicine, Antwerp, Belgium; [24]KU Leuven, University of Leuven, Department of Microbiology and Immunology, Belgium; [25]National Institute for Biomedical Research, Kinshasa, Democratic Republic of the Congo; [26]University Hospital of Kinshasa, Kinshasa, Democratic Republic of the Congo; [27]Microbiology Department, College of Medicine, University of Malawi, Malawi; [28]Departamento de Desarrollo Biotecnologico, Instituto de Higiene, Facultad de Medicina, Avda A Navarro 3051, Montevideo, CP 11600, Uruguay; [29]Ministry of Health, Toorak, PO Box 2223, Suva, Fiji; [30]Fiji Health Sector Support Program, PO Box 14986, Suva, Fiji; [31]National Institute of Cholera and Enteric Diseases, P-33 CIT Road, Scheme XM, Beliaghata, Kolkata 700 010, India; [32]International Vaccine Institute, Department of Epidemiology, Kwanak PO Box 14, Seoul 151-600, Republic of Korea; [33]Enteropathogen Division, ANLIS-Carlos G Malbran Institute, CABA, Argentina; [34]Division of Pediatric Infectious Diseases, University of Nebraska Medical Center, Omaha, NE 68198, USA; [35]University of Abuja Teaching Hospital, Gwagwalada, FCT, Nigeria; [36]Bingham University, Karu, Nassarawa State, Nigeria; [37]The Jenner Institute, Nuffield Department of Medicine, University of Oxford, Oxford, UK; [38]Centre for Enteric Diseases, National Institute for Communicable Diseases, Division in the National Health Laboratory Service and Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa; [39]Department of Clinical Research, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK; [40]Graduate School of Tropical Medicine and Global Health, Nagasaki University, Nagasaki, Japan; [41]Patan Academy of Health Sciences, Wellcome Trust Major Overseas Programme, Oxford University Clinical Research Unit, Kathmandu, Nepal; [42]Murdoch Childrens Research Institute, Melbourne, Australia; [43]Enteric and Leptospira Reference Laboratory, Institute of Environmental Science and Research Limited (ESR), New Zealand; [44]National Centre for Biosecurity and Infectious Disease, Institute of Environmental Science and Research, Porirua, New Zealand; [45]Samoa Ministry of Health, Apia, Samoa; [46]National Influenza Center, World Health Organization, Center for Communicable Disease Control, Suva, Fiji; [47]Department of Microbiology, Hasanuddin University, Makassar, Indonesia; [48]Lao-Oxford-Mahosot Hospital-Wellcome Trust Research Unit, Microbiology Laboratory, Mahosot Hospital, Vientiane, Lao People's Democratic Republic; [49]National Health Services, Tupua Tamasese Meaole Hospital, Samoa; [50]Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand; [51]Paediatric Emergency Medicine, Chelsea and Westminster Hospital, London, UK; [52]Centre for International Health, University of Otago, PO Box 56, Dunedin 9054, New Zealand; [53]Salmonella Reference Service, Public Health England, Colindale, London NW9 5EQ, UK; [54]Emerging Disease Surveillance and Response, Division of Pacific Technical Support, World Health Organization, PO Box 113, Suva, Fiji; [55]Cambodia-Oxford Medical Research Unit, Angkor Hospital for Children, Siem Reap, Cambodia; [56]Microbiological Diagnostic Unit - Public Health Laboratory, Department of Microbiology and Immunology at the Peter Doherty Institute for Infection and Immunity, The University of Melbourne, Victoria 3010, Australia; [57]Department of Veterinary Medicine, University of Cambridge, Cambridge, UK; [58]St Augustine's Hospital, Muheza, Tanzania.

PLOS | NEGLECTED TROPICAL DISEASES

RESEARCH ARTICLE

Whole Genome Sequence Analysis of *Salmonella* Typhi Isolated in Thailand before and after the Introduction of a National Immunization Program

Zoe A. Dyson[1,2]*, Duy Pham Thanh[3], Ladaporn Bodhidatta[4], Carl Jeffries Mason[4], Apichai Srijan[4], Maia A. Rabaa[3,5], Phat Voong Vinh[3], Tuyen Ha Thanh[3], Guy E. Thwaites[3,5], Stephen Baker[3,5,6‡], Kathryn E. Holt[1,2‡]*

1 Centre for Systems Genomics, University of Melbourne, Parkville, Victoria, Australia, **2** Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria, Australia, **3** The Hospital for Tropical Diseases, Wellcome Trust Major Overseas Programme, Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam, **4** Department of Enteric Diseases, Armed Forces Research Institute of Medical Sciences, Bangkok, Thailand, **5** Centre for Tropical Medicine and Global Health, Oxford University, Oxford, United Kingdom, **6** The London School of Hygiene and Tropical Medicine, London, United Kingdom

‡ These authors are joint senior authors on this work.
* kholt@unimelb.edu.au (KEH); zoe.dyson@unimelb.edu.au (ZAD)

OPEN ACCESS

Citation: Dyson ZA, Thanh DP, Bodhidatta L, Mason CJ, Srijan A, Rabaa MA, et al. (2017) Whole Genome Sequence Analysis of *Salmonella* Typhi Isolated in Thailand before and after the Introduction of a National Immunization Program. PLoS Negl Trop Dis 11(1): e0005274. doi:10.1371/journal.pntd.0005274

Editor: Edward T. Ryan, Massachusetts General Hospital, UNITED STATES

Received: October 11, 2016

Accepted: December 20, 2016

Published: January 6, 2017

Data Availability Statement: Raw sequence data have been submitted to the European Nucleotide Archive (ENA) under project PRJEB5281.

Funding: This project was funded by the Wellcome Trust of Great Britain (106158/Z/14/Z); SB is a Sir Henry Dale Fellow, jointly funded by the Wellcome Trust and the Royal Society (100087/Z/12/Z) and ZAD is funded by strategic award #106158. KEH is supported by fellowship #1061409 from the NHMRC of Australia. DTP is a leadership fellow

Abstract

Vaccines against *Salmonella* Typhi, the causative agent of typhoid fever, are commonly used by travellers, however, there are few examples of national immunization programs in endemic areas. There is therefore a paucity of data on the impact of typhoid immunization programs on localised populations of *S.* Typhi. Here we have used whole genome sequencing (WGS) to characterise 44 historical bacterial isolates collected before and after a national typhoid immunization program that was implemented in Thailand in 1977 in response to a large outbreak; the program was highly effective in reducing typhoid case numbers. Thai isolates were highly diverse, including 10 distinct phylogenetic lineages or genotypes. Novel prophage and plasmids were also detected, including examples that were previously only reported in *Shigella sonnei* and *Escherichia coli*. The majority of *S.* Typhi genotypes observed prior to the immunization program were not observed following it. Post-vaccine era isolates were more closely related to *S.* Typhi isolated from neighbouring countries than to earlier Thai isolates, providing no evidence for the local persistence of endemic *S.* Typhi following the national immunization program. Rather, later cases of typhoid appeared to be caused by the occasional importation of common genotypes from neighbouring Vietnam, Laos, and Cambodia. These data show the value of WGS in understanding the impacts of vaccination on pathogen populations and provide support for the proposal that large-scale typhoid immunization programs in endemic areas could result in lasting local disease elimination, although larger prospective studies are needed to test this directly.

funded through the Oak Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The view expressed in this article are those of the author(s) and do not reflect the official policy of the Department of the Army, Department of Defense, or the US government.

Competing Interests: The authors have declared that no competing interests exist.

Author Summary

Typhoid fever is a systemic infection caused by the bacterium *Salmonella* Typhi. Typhoid fever is associated with inadequate hygiene in low-income settings and a lack of sanitation infrastructure. A sustained outbreak of typhoid fever occurred in Thailand in the 1970s, which peaked in 1975–1976. In response to this typhoid fever outbreak the government of Thailand initiated an immunization program, which resulted in a dramatic reduction in the number of typhoid cases in Thailand. To better understand the population of *S.* Typhi circulating in Thailand at this time, as well as the impact of the immunization program on the pathogen population, we sequenced the genomes of 44 *S.* Typhi obtained from hospitals in Thailand before and after the immunization program. The genome sequences showed that isolates of *S.* Typhi bacteria isolated from post-immunization era typhoid cases were likely imported from neighbouring countries, rather than strains that have persisted in Thailand throughout the immunization period. Our work provides the first historical insights into *S.* Typhi in Thailand during the 1970s, and provides a model for the impact of immunization on *S.* Typhi populations.

Introduction

Salmonella enterica subspecies *enterica* serovar Typhi (*S.* Typhi) is a human restricted bacterial pathogen and the etiological agent of typhoid fever. *S.* Typhi is transmitted faeco-orally and can establish asymptomatic carriage in a small subset of an exposed population [1]. Recent estimates [2–4] place the global burden of typhoid fever at 25–30 million cases annually, of which 200,000 are associated with deaths. Typhoid fever occurs most commonly in industrialising countries, specifically in locations with limited sanitation and related infrastructure [5]; children and young adults are among the most vulnerable populations in these settings [6–8]. Antimicrobial therapy together with water sanitation and hygiene (WASH) interventions are the major mechanisms by which typhoid fever is controlled [9, 10]. However, none of these approaches are optimal and resistance against antimicrobials has become increasingly common in *S.* Typhi since the 1970s [11–13]. A number of typhoid vaccines are licenced for use [14–18], however, they are not widely used as a public health tools in endemic areas, with the exception of controlling severe outbreaks such as those following natural disasters [19–22].

A sustained typhoid fever outbreak occurred in Thailand in the 1970s. A sharp increase in cases was observed in 1973–1974, which finally peaked in 1975–1976. In response, the government of Thailand established a national typhoid immunization program, which represented the first programmatic use of a typhoid vaccine in the country [14, 22, 23]. The immunization program targeted over 5 million school aged children (7–12 years) each year in Bangkok between 1977 and 1987 (80% of the eligible population). Thus, Thai school children were eligible to receive a single locally produced heat/phenol-inactivated subcutaneous dose of 2.5 x 10^8 *S.* Typhi organisms annually [14, 22, 23], before the program was halted in the early 1990s because of high rates of adverse reactions caused by the vaccine [22]. To our knowledge this is the only such programmatic use of a vaccine for controlling Typhoid fever in children in Thailand. Data from four teaching hospitals in Bangkok showed a 93% reduction in blood culture confirmed infections with *S.* Typhi between 1976 (n = 2,000) and 1985 (n = 132) [14, 23]. Notably, no significant decline was observed in isolation rates of *Salmonella* Paratyphi A (*S.* Paratyphi A), a *Salmonella* serovar distinct from *S.* Typhi that causes a clinical syndrome indistinguishable from typhoid fever, but for which *S.* Typhi vaccines

provide little or no cross-protection [14]. This observation suggests that the reduction in *S.* Typhi infections was not attributable to improvements in infrastructure and hygiene practices only [5, 14, 20, 23]. While the inactivated *S.* Typhi vaccine was found to be highly efficacious [22, 23], it is no longer used as a consequence of being overly reactogenic [14, 16, 22, 23, 24]. A Vi capsular polysaccharide vaccine [15] and live-attenuated oral vaccine of strain Ty21a [16] have since replaced this vaccine for travellers to endemic locations [5, 21, 24].

The typhoid immunization program in Thailand provided a unique opportunity to investigate the impact of immunization on *S.* Typhi populations circulating within an endemic area. Here we present an analysis of a historical collection of 44 *S.* Typhi isolates obtained from patients in Thailand between 1973 and 1992 (before and during the immunization program). As *S.* Typhi populations demonstrate little genetic diversity, we used whole genome sequencing (WGS) to characterise these isolates, and core genome phylogenetic approaches to compare the historic isolates from Thailand to a recently published global *S.* Typhi genomic framework [4].

Materials and Methods

Ethics statement

This is a retrospective study of bacterial isolates unlinked to patient information and was not subject to IRB approval.

Bacterial isolation and antimicrobial susceptibility testing

Forty-four *S.* Typhi isolated from patients with suspected typhoid fever attending hospitals in Bangkok, Nonthaburi, Loi, and Srakaew, in Thailand between 1973 and 1992 were available for genome sequencing in this study (Fig 1 and S1 Table). At the time of original isolation, bacterial cultures were transferred on nutrient agar slants to the department of Enteric Diseases, Armed Forces Research Institute of Medical Sciences (AFRIMS), Bangkok, Thailand for identification and antimicrobial susceptibility testing. At AFRIMS, bacterial isolates were subcultured on Hektoen Enteric agar (HE) and identification was performed by biochemical testing on Kligler iron agar slants, tryptone broth for indole, lysine decarboxylase medium, ornithine decarboxylase medium, urease test, mannitol and motility media (Becker Dickenson, Thailand). Serological agglutination was performed using *Salmonella* O antisera and *Salmonella* Vi antiserum (Difco, USA). Bacterial strains were stored frozen at -70˚C in 10% skimmed milk or lyophilised in 10% skimmed milk; lyophilized ampoules were stored at 2–8˚C. Prior to DNA extraction for sequencing, lyophilized bacteria were rehydrated with trypticase soy broth, inoculated on McConkey agar and incubated at 37˚C for 18–24 hours. If bacteria were stored frozen in skimmed milk, organisms were inoculated directly onto McConkey agar after thawing and then incubated at 37˚C for 18–24 hours.

Antimicrobial susceptibility testing against ampicillin, chloramphenicol, cephalothin, gentamicin, kanamycin, neomycin, sulfisoxazole, trimethoprim/sulfamethoxazole, and tetracycline was performed by disk diffusion according to Clinical and Laboratory Standards Institute (CLSI) [25–28].

Genome sequencing and SNP analysis

Genomic DNA from the 44 *S.* Typhi from Thailand was extracted using the Wizard Genomic DNA Extraction Kit (Promega, Wisconsin, USA). Two μg of genomic DNA was subjected to indexed WGS on an Illumina Hiseq 2000 platform at the Wellcome Trust Sanger Institute, to generate 100 bp paired-end reads. For analysis of SNPs, paired end Illumina reads were

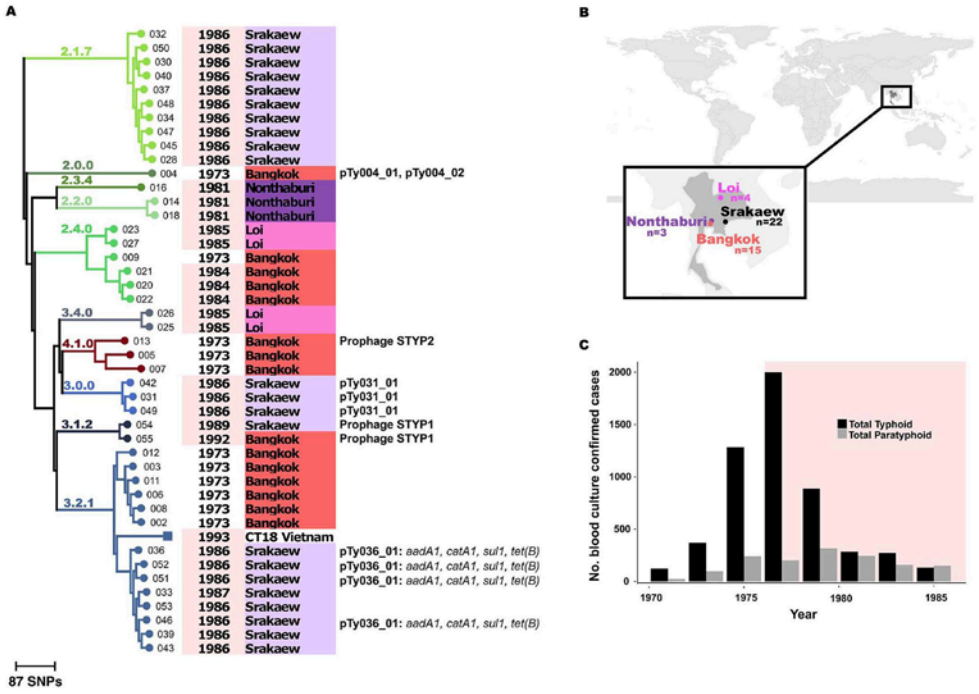

Fig 1. Genomic analysis of Thai S. Typhi. (A) Maximum likelihood phylogenetic tree (outgroup rooted). Strains are labelled with their three digit name code, year of isolation (pink shading indicates post-vaccine isolates); source location (shaded by city, as indicated in panel B); and plasmid content (any antibiotic resistance genes are indicated in italics). Branch lengths are indicative of the number of SNPs. (B) Locations from which S. Typhi were isolated in Thailand. (C) Total number of positive blood cultures of S. Typhi (black) and Paratyphi A (grey) between 1970 and 1985; immunization period is indicated in pink; reproduced using data from reference (14).

doi:10.1371/journal.pntd.0005274.g001

mapped to the reference sequence of *S.* Typhi CT18 (accession no: AL513382) [29] using the RedDog (v1.4) mapping pipeline, available at https://github.com/katholt/reddog. RedDog uses Bowtie (v2.2.3) [30] to map reads to the reference sequence, then high quality SNPs called with quality scores above 30 are extracted from the alignments using SAMtools (v0.1.19) [31]. SNPs were filtered to exclude those with less than 5 reads mapped or with greater than 2.5 times the average read depth (representing putative repeated sequences), or with ambiguous base calls. For each SNP that passed these criteria in any one isolate, consensus base calls for the SNP locus were extracted from all genomes (ambiguous base calls and those with phred quality scores less than 20 were treated as unknowns and represented with a gap character). SNPs with confident homozygous allele calls (i.e. phred score >20) in >95% of the *S.* Typhi genomes (representing a 'soft' core genome of common *S.* Typhi sequences) were concatenated to produce an alignment of alleles at 45,893 variant sites. The resultant allele calls for 68 of these SNPs were used to assign isolates to previously defined lineages according to an extended *S.* Typhi genotyping framework [32] code available at https://github.com/katholt/genotyphi).

SNPs called in phage regions, repetitive sequences (354 kb; ~7.4% of bases in the CT18 reference chromosome, as defined previously [33] or recombinant regions (~180kb; <4% of the CT18 reference chromosome, identified using Gubbins (v1.4.4) [34]) were excluded, resulting in a final set of 1,850 SNPs identified in an alignment length of 4,275,037 bp for the 44 isolates. SNP alleles from Paratyphi A strain 12601 [35] were also included as an outgroup to root the tree. For global context, raw read data [4] were also subjected to genotyping analysis and those isolates sharing the genotypes that were observed in the Thai collection (n = 340; details in S2 Table) were subjected to the same SNP analyses, resulting in a final set of 9,700 SNPs for a total of 386 isolates.

Phylogenetic and SNP analysis

Maximum likelihood (ML) phylogenetic trees (Figs 1 and 2) were constructed using the 1,850 and 9,700 bp SNP alignments, respectively, using RAxML (v 8.1.23) [36] with a generalized time-reversible model and a gamma distribution to model site specific recombination (GTR+Γ substitution model; GTRGAMMA in RAxML), with Felsenstein correction for ascertainment bias. Support for ML phylogenies was assessed via 100 bootstrap pseudoanalyses of the alignments. For the larger tree containing global isolates, clades containing only isolates from only a single country were collapsed manually in R using the drop.tip() function in the *ape* package [37]. Subtrees were extracted for each subclade, which are therefore each rooted by the other subclades. Pairwise SNP distances between isolates were calculated from the SNP alignments using the dist.gene() function in the *ape* package for R [37].

Accessory genome analysis

Acquired antimicrobial resistance (AMR) genes were detected, and their precise alleles determined, by mapping to the ARG-Annot database [38] of known AMR genes using SRST2 v0.1.5 [39]. Plasmid replicon sequences were identified using SRST2 to screen reads for replicons in the PlasmidFinder database [40, 41]. Raw read data was assembled *de novo* with SPAdes (v 3.5.0) [42] and circular contigs were identified visually and extracted using the assembly graph viewer Bandage (v0.7.0) [43]. These putative plasmid sequences were annotated using Prokka (v1.10) [44] followed by manual curation. Where IncHI1 plasmid replicons were identified using SRST2, and their presence confirmed by visual inspection of the assembly graphs, IncHI1 plasmid MLST (pMLST) sequence types were determined using SRST2 [13, 39, 45, 46]. Where resistance genes were detected from short read data, Bandage was used to inspect their location in the corresponding *de novo* assembly graph in order to determine whether they were encoded in the bacterial chromosome or on a plasmid. Assembled contigs were concatenated and putative prophage genomes were identified with the PHAge Search Tool (PHAST) [47], and their novelty determined by BLASTN analysis against the GenBank database. Pairwise alignments between novel and known prophage sequences were visualised using the *genoPlotR* package for R [48].

Nucleotide sequence and sequence read data accession numbers

Raw sequence data have been submitted to the European Nucleotide Archive (ENA) under project PRJEB5281; individual sample accession numbers are listed in S1 and S2 Tables. Assembled phage and protein sequences were deposited in GenBank, accession numbers are listed in Table 1.

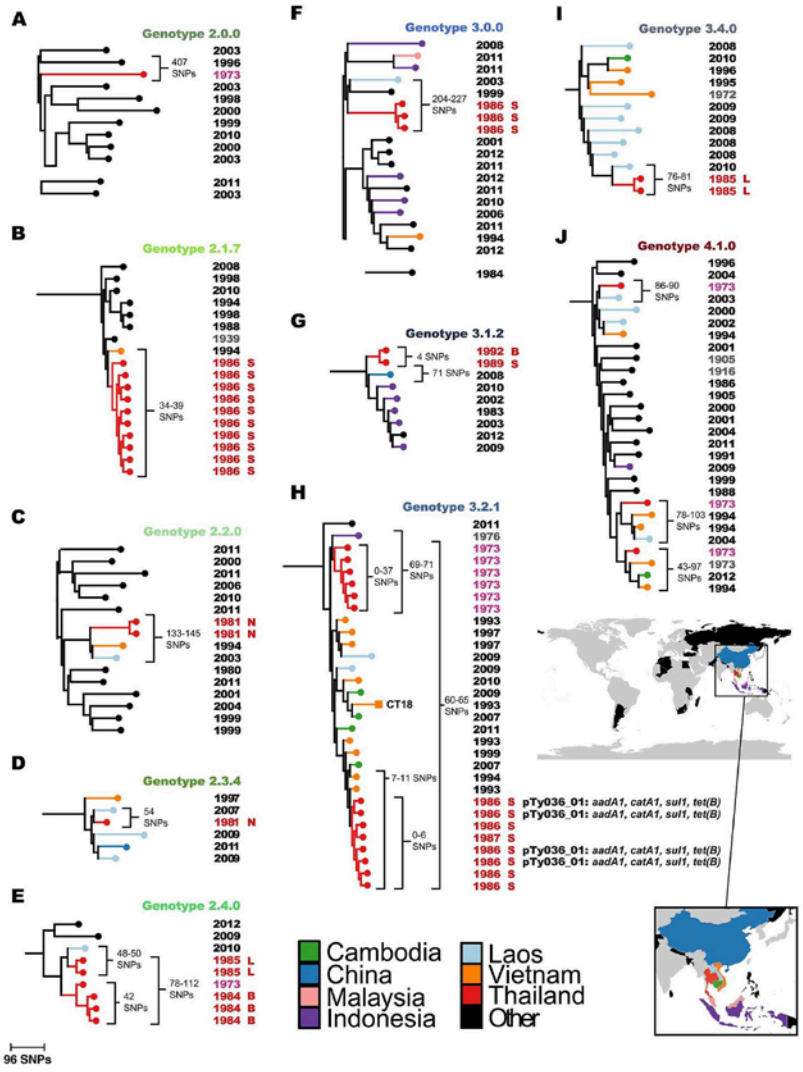

Fig 2. Zoomed in phylogenies showing relationships of Thai S. Typhi to global isolates. Maximum likelihood trees including S. Typhi isolates from the Thai and global collections are shown, for each genotype that was observed amongst the Thai isolates. (A) Genotype 2.0.0 tree. (B) Genotype 2.1.7 tree. (C) Genotype 2.2.0 tree. (D) Genotype 2.3.4. tree (E) Genotype 2.4.0. tree (F) Genotype 3.0.0 tree (G) Genotype 3.1.2. tree (H) Genotype 3.2.1 tree. (I) Genotype 3.4.0 tree. (J) Genotype 4.1.0 tree. Colored branches and nodes indicate country of origin, according to the inset legend. Year of isolation is shown to the right; pink and red, Thai isolates obtained before and after the introduction of the immunization program;

grey and black, non-Thai isolates obtained before and after the introduction of the immunization program. Thai isolates are also labelled to indicate their city of origin: L, Loi; B, Bangkok; S, Srakaew; N, Nonthaburi. SNP distances between isolates as well as AMR plasmids are labelled, with any resistance genes indicated in italics. Branch lengths are indicative of the number of SNPs.

doi:10.1371/journal.pntd.0005274.g002

Results

The population structure of *S.* Typhi in Thailand

All 44 *S.* Typhi isolates collected between 1973 and 1992 were subjected to WGS and SNP analysis. Genome-wide SNPs were used to construct a ML phylogeny and isolates were assigned to previously defined genotypes [32] using a subset of SNPs (see Methods). These analyses subdivided the population into ten distinct genotypes, each corresponding to a specific lineage in the ML phylogeny (Fig 1). Genotype 3.2.1 (which includes the reference genome CT18, isolated from Vietnam in 1993 [29]) was the most common (n = 14, 32%), followed by genotype 2.1.7 (n = 10, 23%). Genotypes 2.0 (n = 1, 2%) and 4.1 (n = 3, 7%) were observed only in 1973 (pre-vaccine period). Genotypes 2.1.7 (n = 10, 23%), 2.3.4 (n = 1, 2%), 3.4.0 (n = 2, 5%), 3.0.0 (n = 3, 7%), 3.1.2 (n = 2, 5%), were observed only after 1981 (post-vaccine period). Each of these post-immunization genotypes was from a single location and time period (Fig 1), consistent with short-term localised transmission. The only exceptions were the two *S.* Typhi 3.1.2 isolates, that were from Srakaew in 1989 and Bangkok in 1992 and separated by just 4 SNPs. Genotypes 3.2.1 and 2.4.0 were observed amongst both pre- and post-vaccine isolates.

Thai *S.* Typhi in the context of a global genomic framework

Based on the Thai *S.* Typhi genotyping results we hypothesised that the post-immunization typhoid infections in Thailand resulted from occasional re-introduction of *S.* Typhi from outside the country, as opposed to long-term persistence of *S.* Typhi lineages within Thailand. To explore this possibility, and to provide a global context for our analysis, we examined 1,832 *S.* Typhi genomes from a recently published global collection that included isolates from 63

Table 1. Summary of mobile genetic elements observed in *S.* Typhi isolates from Thailand.

Isolate	Genotype	Name	Replicons detected and/or attachment sites	Size (no. putative genes)	Accession number	Function
004	2.0.0	pTy004_01	FIB (pHCM2)	108, 998 bp (133)	KX833209	Cryptic, Phage defence (Rha protein)
		pTy004_02	X1	38, 266 bp (49)	KX833212	Phage defence (Abortive Infection)
031	3.0.0	pTy031_01	N/A	40, 835 bp (53)	KX833210	Phage defence (Restriction Modification)
042						
049						
036	3.2.1	pTy036_01	HI1	~215 kbp	N/A.	AMR (*sul1*, *catA1*, *tet(B)*, *aadA1*)
046						
051						
052						
054	3.1.2	Prophage STYP1	*attL* CAAGCTGGTCAG *attR* CAAGCTGGTCAG	28,946 bp (39)	KX833211	Cryptic
055						
013	4.1.0	Prophage STYP2	*attL* ATTCGTAATGCGAAGGTCGTAGGTTC GACTCCTATTATCGGCACCAT *attR* ATTCGTAATGCGAAGGTCGTAGGTT CGACTCCTATTATCGGCACCA	34, 780 bp (50)	KX833213	Cryptic

doi:10.1371/journal.pntd.0005274.t001

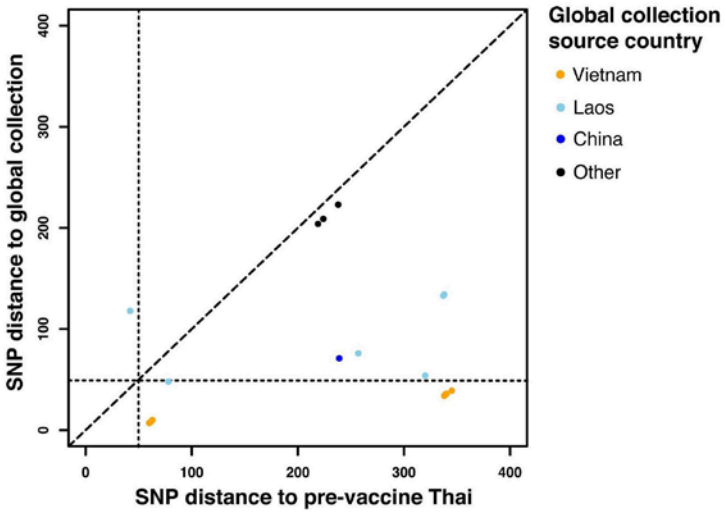

Fig 3. SNP distances for Thai and global collection isolates. SNP distance between post-vaccine Thai isolates and their closest pre-vaccine Thai and post-vaccine global collection relatives, colored points indicate country of origin.

doi:10.1371/journal.pntd.0005274.g003

countries [4]. Genome-wide SNP-based ML trees for each of these genotypes, showing the relationships between Thai and global isolates, are shown in Fig 2. In general, post-vaccine Thai isolates were closely related to recent isolates sourced from neighbouring countries including Vietnam, Laos and Cambodia (Fig 2), consistent with regional endemic circulation. In contrast, most pre-vaccine isolates had no close neighbours in the global collection, particularly 2.0.0 strains (Fig 2A), suggesting they may have been Thailand-specific lineages that have died out following the vaccine program. The *S.* Typhi genomes in the global collection were mainly isolated 2–3 decades after the Thai isolates as we did not have access to contemporaneous isolates from these countries that could identify specific transfer events. However, all but three of the post-vaccine Thai isolates shared shorter SNP distances with isolates from neighbouring countries than they did with pre-vaccination Thai isolates (see Fig 3), consistent with these cases being caused by occasional re-introduction of genotypes circulating in the region. Notably, Thai *S.* Typhi 3.2.1 that were isolated in 1986–7 clustered separately from the 1973 pre-vaccine isolates ($\geq$60 SNPs apart), but closely with isolates from Vietnam and Cambodia (differing by as few as 7 SNPs; Fig 2H). Post-vaccine Thai *S.* Typhi 2.4 formed two distinct groups that were not consistent with direct descendance from earlier isolates (Fig 2E). These data are therefore consistent with transfer of *S.* Typhi into Thailand from neighbouring countries during the post-immunization program era, although the long-term circulation of ancestral populations in Thailand remains an unlikely alternative explanation.

Acquired antimicrobial resistance

We identified acquired AMR genes in the genomes of four *S.* Typhi genotype 3.2.1 that were isolated in Srakaew in 1986 (Fig 1, Table 1). These isolates shared the same four AMR genes:

sul1 (sulphonamides), *catA1* (chloramphenicol), *tet(B)* (tetracyclines), and *aadA1* (aminoglycosides) which were carried on near-identical plasmids of IncHI1 plasmid sequence type 2 (PST2). Although the presence of insertion sequences (IS) in these plasmids prevented the complete sequences from being assembled, the regions of these plasmids encoding the AMR genes were identical in all assemblies. This commonality suggests they are a single plasmid (referred to as pTy036_01 in Fig 1 and Table 1) that was likely acquired in a common ancestor of this clade. The chromosomal and IncHI1 plasmid sequences for these four isolates were very closely related to those of a 1993 Vietnamese isolate (Viety1-60_1993) in the global *S.* Typhi collection [4, 45], consistent with regional transfer.

Other plasmids and mobile genetic elements

We identified three non-AMR related plasmids amongst the Thai isolates (Fig 1, Table 1). Ty004 (genotype 2.2) carried two novel plasmids that assembled into circular sequences, pTy004_01 and pTy004_02. The largest, pTy004_01, was a novel variant of the cryptic plasmid pHCM2 [29, 49] (Fig 4). Ty004 was isolated in Bangkok in 1973, making pTy004_01 the earliest example of a pHCM2-like plasmid reported to date. pTy004_01 was distant from other pHCM2-like plasmids in the global *S.* Typhi genome collection, sharing 92% coverage and 99% nucleotide identity with the reference sequence pHCM2 of *S.* Typhi CT18 (genotype 3.2.1) which was isolated approximately 20 years later in Vietnam [29]. The pTy004_01 sequence (Fig 4) appears to be ~2 kbp larger than pHCM2, and encodes an additional tRNA-Lys as well as an insertion of a hypothetical protein (*orf17*) into a putative DNA polymerase gene (HCM2.0015c in pHCM2, divided into *orf16* and *orf18* in pTy004_01). Plasmid pTy004_02 was ~38 kbp in size and similar to *E. coli* plasmid pEQ2 (65% coverage, 98% nucleotide identity), encoding genes for conjugation, chromosomal partitioning, addiction systems and an abortive infection protein (*orf44*). Three isolates (Ty031, Ty042, and Ty049) all of genotype 3.0.0 and obtained from Srakaew in 1986, carried a ~40 kbp cryptic plasmid that we named pTy031_01. This plasmid was similar to that carried by *Enterobacter hormaechei* strain CAV1176 (83% coverage, 96% identity) and encoded genes for chromosomal partitioning, addiction systems, and a putative restriction modification system (*orf33-orf34*).

PHAST analysis revealed the presence of novel intact prophages in three Thai *S.* Typhi isolates (Fig 1, Table 1). Two *S.* Typhi 3.1.2, isolated from Srakaew in 1989 and Bangkok in 1992,

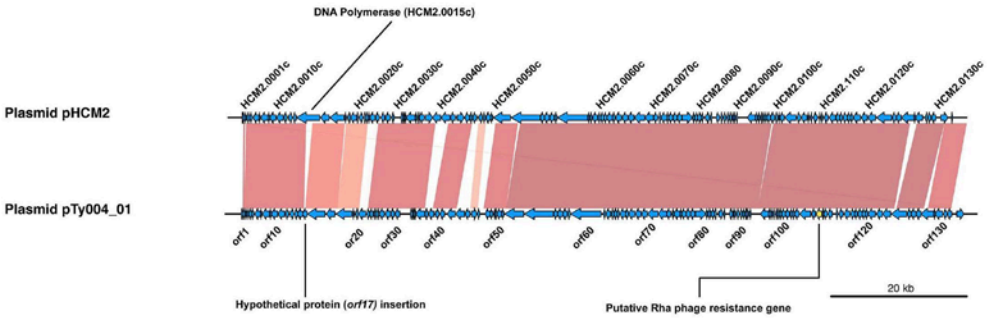

Fig 4. Blast comparison of novel plasmid pTy004_01 with pHCM2 (AL513383). Shaded regions indicate areas of sequence homology, intensity of shading indicates relative nucleotide similarity. Arrows represent protein coding genes, direction indicates coding strand.

doi:10.1371/journal.pntd.0005274.g004

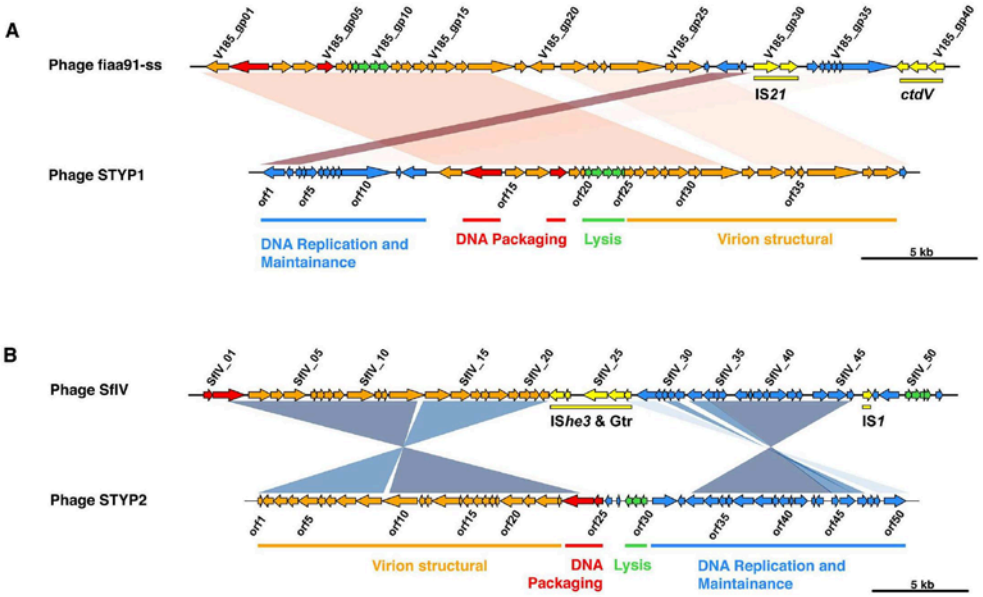

Fig 5. Blast comparison of novel phages observed in Thai S. Typhi isolates to nearest known phage sequences. (A) Novel phage STYP1 compared to Shigella sonnei phage fiAA91-ss (NC_022750). (B) Novel phage STYP2 compared to Shigella flexneri phage SfIV (NC_022749). Shaded regions indicate areas of sequence homology, intensity of shading indicates relative nucleotide similarity. Arrows represent protein coding genes (direction indicates coding strand), colored by encoded protein functions: red, DNA packaging module; orange, virion morphogenesis module; yellow, cargo genes; blue, DNA replication and lysogenic cycle maintenance; green, lysis module.

doi:10.1371/journal.pntd.0005274.g005

shared a novel phage STYP1 that was similar to fiAA91-ss infective for *Shigella sonnei* (Fig 5A). However, the *S.* Typhi phage lacked the cytolethal distending toxin *cdt* genes and the IS*21* element found in phage fiAA91-ss [50]. This prophage sequence had a mosaic architecture, incorporating a number of putative insertions of phage tail fiber genes that were not present in the fiAA91-ss reference genome (Fig 5A). Additionally, a single isolate of genotype 4.1 obtained from Bangkok in 1973 contained a novel SfIV-like phage, here named STYP2, that lacked the serotype conversion gene Gtr cluster and IS*1* element of phage SfIV [51]. Again, the novel Thai phage variant also encoded novel tail fiber genes not in the SfIV reference genome, as well as a Dam methylase gene *(orf37)* (Fig 5B).

Discussion

These data provide a historical insight into the population structure of *S.* Typhi in Thailand in 1973 (pre-immunization program, n = 11) and 1981–1992 (post- immunization program, n = 33). It has been reported that the national *S.* Typhi immunization program in Thailand, which commenced in 1977, was highly effective in reducing the burden of typhoid fever [14]. Our data are consistent with the hypothesis that the vaccine program successfully depleted the endemic *S.* Typhi population to the extent that most subsequent typhoid cases resulted from sporadic introduction of non-indigenous *S.* Typhi, rather than long-term persistence of the

pre-vaccine era population. It is apparent that these introductions were sometimes accompanied by limited local transmissions, resulting in small, localized outbreaks, but we found no evidence to suggest that these result in the establishment of stable local source populations. Notably, the post-immunization *S.* Typhi isolates from Loi (in the north of Thailand near the border with Laos, from which it is separated by the Mekong river) were most closely related to Laos isolates, whilst those from the capital Bangkok and nearby Nonthaburi and Srakaew districts were closely related to other isolates from across Southeast Asia (Fig 2), suggesting there may have been multiple routes of import into Thailand.

Our study is limited by the sample of isolates available for analysis, which was small and reflects opportunistic sampling of sporadic local cases in the four sites and historical storage. A larger collection of historical isolates from Thailand and neighboring countries in the 1970s and 1980s would help to further elucidate the epidemiological patterns of *S.* Typhi before and after the vaccination program. However, from our data, it is notable that the Thai isolates cluster according to site, consistent with limited local transmission rather than dissemination of lineages between locations. The only exception to this was two genotype 3.1.2 isolates, which were collected from Srakaew in 1989 and Bangkok in 1992 and differed by only 4 SNPs. This is consistent with either transfer between these cities in Thailand following an initial introduction into the country, or two independent transfers into Thailand from a common source. The phylogenetic structure is most suggestive of the latter, but denser samples from Thailand and/or potential source populations would be required to resolve this with confidence.

While our sample is small, this study is nevertheless the largest to date exploring genetic diversity amongst *S.* Typhi from Thailand. An earlier global haplotyping study that included seven Thai isolates [52] identified five distinct haplotypes in Thailand (H3, 1989; H42, 1990; H50, 2002; Vi- H52, 1990; H79, 2002), three of which are related to genotypes that we identified amongst Thai strains in this study (H79, 2.3.4; H52, 3.4; H42, 3.1.2) [32]. Genotype 4.3.1 (H58) was not found amongst our historical Thai isolates. This is consistent with previously published spatiotemporal analyses of the global isolate collection, which showed this rapidly expanding clone only began spreading throughout Asia after 1990 [4]. To our knowledge the only evidence to date of the presence of 4.3.1 (H58) in Thailand comes from the global study [4], in which three isolates were identified from 2010–2011, most likely introduced from India. Therefore, our genomic snapshot of the Thai *S.* Typhi population is consistent with previous insights and is likely reasonably representative for the study period. In the years following the vaccination program the prevalence of Typhoid fever in Thailand has continued to decline [53, 54]. The vaccination program has been credited with reducing disease incidence in Thailand and was followed by increased economic development in the region as well as improvements to both water and sanitation systems that have likely improved the control of such outbreaks [53, 54]. Consequently, Typhoid fever is no longer considered a serious public health threat in Thailand [53].

The presence of novel plasmids and prophages in the Thai isolates is also noteworthy. While small plasmids of unknown function have been observed in *S.* Typhi previously [55], they are infrequent compared to the IncHI1 MDR plasmid and the cryptic plasmid pHCM2 [33]. Presumably, such plasmids are ephemeral; possibly because their maintenance imposes a fitness burden on the host cells so a strong selective advantage is required for retention [56, 57]. It is also possible that the lack of previous reports regarding the diversity of small plasmids in *S.* Typhi reflects a technological complexity, however, this is bypassed with high-throughput WGS and we detected negligible small plasmid content in the global collection of 1,832 genomes using the same screening approach [4, 32, 58]. Notably, few of the Thai plasmids share nucleotide sequence homology with those previously described in *S.* Typhi, but were closely related to those found in other *Enterobacteriaceae*. The novel pHCM2-like plasmid

(pTy004_01) and two additional plasmids (pTy004_02 and pTy031_01) harbored genes associated with phage resistance, which could provide protection against phage predation [59–62]. We also observed two novel prophages integrated into Thai genomes, which both showed variation in their phage tail structural regions compared to close neighbors found in *Shigella/E. coli*. These regions are typically responsible for binding of phage to host receptors [63–65], thus the variation in these regions may be associated with recent adaptations to the *S.* Typhi host. While genomic data from more recent *S.* Typhi collections shows limited evidence for genetic exchange with other organisms [4], the detection amongst older Thai isolates of both phage and plasmids that have been previously associated with *E. coli/Shigella* suggests that genetic exchange may have been more common in the past or in certain localized populations.

Overall, these data provide valuable historical insights into the *S.* Typhi populations circulating in Thailand during the 1970s and 1980s, and early examples of the two most common *S.* Typhi plasmids, as well as other mobile elements identified within the *S.* Typhi population. Importantly, while genomic epidemiology has been applied to study typhoid transmission, antimicrobial resistance evolution and antibiotic treatment failure in various settings [66–68], this study provides an important proof-of-principle demonstration that this approach can also provide useful insights into the impact of typhoid vaccines on circulating bacterial populations. This should motivate the adoption of WGS methods to monitor *S.* Typhi populations during future immunization programs and other large-scale interventions, which could potentially identify differential impacts on distinct genotypes.

Supporting Information

S1 Table. Isolate and sequencing details.
(DOCX)

S2 Table. Global isolate and sequencing details.
(DOCX)

Author Contributions

Conceptualization: SB KEH LB CJM DPT MAR.

Formal analysis: ZAD KEH.

Funding acquisition: GET KEH SB.

Investigation: DPT LB CJM AS MAR PVV THT.

Resources: LB CJM SB KEH.

Supervision: DPT.

Visualization: ZAD KEH.

Writing – original draft: ZAD.

Writing – review & editing: KEH SB.

References

1. Parry CM, Hien TT, Dougan G, White NJ, Farrar JJ. Typhoid fever. N Engl J Med. 2002; 347(22):1770–82. doi: 10.1056/NEJMra020201 PMID: 12456854
2. Mogasale V, Maskery B, Ochiai RL, Lee JS, Mogasale VV, Ramani E, 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(10):e570–e80. doi: 10.1016/S2214-109X(14)70301-8 PMID: 25304633

3. Crump JA, Mintz ED. Global trends in typhoid and paratyphoid Fever. Clin Infect Dis. 2010; 50(2):241–6. doi: 10.1086/649541 PMID: 20014951
4. Wong VK, Baker S, Pickard DJ, Parkhill J, Page AJ, Feasey NA, et al. Phylogeographical analysis of the dominant multidrug-resistant H58 clade of Salmonella Typhi identifies inter- and intracontinental transmission events. Nat Genet. 2015; 47(6):632–9. doi: 10.1038/ng.3281 PMID: 25961941
5. Connor BA, Schwartz E. Typhoid and paratyphoid fever in travellers. Lancet Infect Dis. 2005; 5 (10):623–8. doi: 10.1016/S1473-3099(05)70239-5 PMID: 16183516
6. Crump JA, Luby SP, Mintz ED. The global burden of typhoid fever. Bull World Health Organ. 2004; 82 (5):346–53. PMID: 15298225
7. Sur D, von Seidlein L, Manna B, Dutta S, Deb AK, Sarkar BL, et al. The malaria and typhoid fever burden in the slums of Kolkata, India: data from a prospective community-based study. Trans R Soc Trop Med Hyg. 2006; 100(8):725–33. doi: 10.1016/j.trstmh.2005.10.019 PMID: 16455118
8. Karkey A, Arjyal A, Anders KL, Boni MF, Dongol S, Koirala S, et al. The Burden and Characteristics of Enteric Fever at a Healthcare Facility in a Densely Populated Area of Kathmandu. PLoS ONE. 2010; 5 (11):e13988. doi: 10.1371/journal.pone.0013988 PMID: 21085575
9. Cairncross S, Cumming O, Jeandron A, Rheingans R, Ensink J, Brown J, et al. Water, Sanitation and Hygiene Evidence Paper. London: Department for International Development. 2013.
10. Bhutta ZA. Current concepts in the diagnosis and treatment of typhoid fever. BMJ. 2006; 333(7558):78–82. doi: 10.1136/bmj.333.7558.78 PMID: 16825230
11. Chau TT, Campbell JI, Galindo CM, Van Minh Hoang N, Diep TS, Nga TT, et al. Antimicrobial Drug Resistance of Salmonella enterica Serovar Typhi in Asia and Molecular Mechanism of Reduced Susceptibility to the Fluoroquinolones. Antimicrob Agents Chemother. 2007; 51(12):4315–23. doi: 10.1128/AAC.00294-07 PMID: 17908946
12. Kariuki S, Revathi G, Muyodi J, Mwituria J, Munyalo A, Mirza S, et al. Characterization of Multidrug-Resistant Typhoid Outbreaks in Kenya. J Clin Microbiol. 2004; 42(4):1477–82. doi: 10.1128/JCM.42.4.1477-1482.2004 PMID: 15070992
13. Holt KE, Phan MD, Baker S, Duy PT, Nga TVT, Nair S, et al. Emergence of a Globally Dominant IncHI1 Plasmid Type Associated with Multiple Drug Resistant Typhoid. PLoS Negl Trop Dis. 2011; 5(7):e1245. doi: 10.1371/journal.pntd.0001245 PMID: 21811646
14. Bodhidatta L, Taylor DN, Thisyakorn U, Echeverria P. Control of Typhoid Fever in Bangkok, Thailand, by Annual Immunization of Schoolchildren with Parenteral Typhoid Vaccine. Rev Infect Dis. 1987; 9 (4):841–5. PMID: 3438648
15. Klugman K, Koornhof H, Schneerson R, Cadoz M, Gilbertson I, Robbins J, et al. Protective activity of Vi capsular polysaccharide vaccine against typhoid fever. Lancet. 1987; 330(8569):1165–9.
16. Levine M, Black R, Ferreccio C, Germanier R. Large-scale field trial of Ty21A live oral typhoid vaccine in enteric-coated capsule formation. Lancet. 1987; 329(8541):1049–52.
17. Ashcroft MT, Singh B, Nicholson CC, Ritchie JM, Sobryan E, Williams F. A seven-year field trial of two typhoid vaccines in Guyana. Lancet. 1967; 290(7525):1056–9.
18. Waddington CS, Darton TC, Jones C, Haworth K, Peters A, John T, et al. An Outpatient, Ambulant-Design, Controlled Human Infection Model Using Escalating Doses of Salmonella Typhi Challenge Delivered in Sodium Bicarbonate Solution. Clin Infect Dis. 2014; 58(9):1230–40. doi: 10.1093/cid/ciu078 PMID: 24519873
19. Scobie HM, Nilles E, Kama M, Kool JL, Mintz E, Wannemuehler KA, et al. Impact of a targeted typhoid vaccination campaign following cyclone Tomas, Republic of Fiji, 2010. Am J Trop Med Hyg. 2014; 90 (6):1031–8. doi: 10.4269/ajtmh.13-0728 PMID: 24710618
20. Thompson CN, Kama M, Acharya S, Bera U, Clemens J, Crump JA, et al. Typhoid fever in Fiji: a reversible plague? Trop Med Int Health. 2014; 19(10):1284–92. doi: 10.1111/tmi.12367 PMID: 25066005
21. Whitaker JA, Franco-Paredes C, del Rio C, Edupuganti S. Rethinking typhoid fever vaccines: implications for travelers and people living in highly endemic areas. J Travel Med. 2009; 16(1):46–52. doi: 10.1111/j.1708-8305.2008.00273.x PMID: 19192128
22. DeRoeck D, Ochiai RL, Yang J, Anh DD, Alag V, Clemens JD. Typhoid vaccination: the Asian experience. Expert Rev Vaccines. 2008; 7(5):547–60. doi: 10.1586/14760584.7.5.547 PMID: 18564010
23. Guzman CA, Borsutzky S, Griot-Wenk M, Metcalfe IC, Pearman J, Collioud A, et al. Vaccines against typhoid fever. Vaccine. 2006; 24(18):3804–11. doi: 10.1016/j.vaccine.2005.07.111 PMID: 16278037
24. Ochiai RL, Acosta CJ, Agtini M, Bhattacharya SK, Bhutta ZA, Do CG, et al. The Use of Typhoid Vaccines in Asia: The DOMI Experience. Clin Infect Dis. 2007; 45(Supplement 1):S34–S8.

25. CLSI. Performance standards for anitmicrobial disk suseptibility tests. Approved standard. CLSI document M02. Wayne, Pa: CLSI; 1984.

26. CLSI. Performace standards for antimicrobial disk susceptibility tests. Tentative standard. CLSI document M02. Wayne, Pa: CLSI; 1979.

27. CLSI. Performance standards of antimicrobial disk susceptibility tests. Tentative standard. CLSI document M02-A4. Wayne, Pa: CLSI; 1988.

28. CLSI. Performace standards for anitmicrobial disk susceptibility tests. Approved Standard. CLSI document M02-A4. Wayne, Pa: CLSI; 1990.

29. Parkhill J, Dougan G, James KD, Thomson NR, Pickard D, Wain J, et al. Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18. Nature. 2001; 413(6858):848–52. doi: 10.1038/35101607 PMID: 11677608

30. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Meth. 2012; 9(4):357–9.

31. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N. The sequence alignment/Map format and SAMtools. Bioinformatics. 2009; 25.

32. Wong VK, Baker S, Connor TR, Pickard D, Page AJ, Dave J, et al. An extended genotyping framework for Salmonella enterica serovar Typhi, the cause of human typhoid. Nat Commun. 2016; 7:12827. doi: 10.1038/ncomms12827 PMID: 27703135

33. Holt KE, Parkhill J, Mazzoni CJ, Roumagnac P, Weill F-X, Goodhead I. High-throughput sequencing provides insights into genome variation and evolution in Salmonella Typhi. Nat Genet. 2008; 40.

34. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res. 2014.

35. Holt KE, Thomson NR, Wain J, Langridge GC, Hasan R, Bhutta ZA, et al. Pseudogene accumulation in the evolutionary histories of Salmonella enterica serovars Paratyphi A and Typhi. BMC genomics. 2009; 10:36. doi: 10.1186/1471-2164-10-36 PMID: 19159446

36. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006; 22(21):2688–90. doi: 10.1093/bioinformatics/btl446 PMID: 16928733

37. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 2004; 20(2):289–90. PMID: 14734327

38. Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, et al. ARG-ANNOT, a New Bioinformatic Tool To Discover Antibiotic Resistance Genes in Bacterial Genomes. Antimicrob Agents and Chemother. 2014; 58(1):212–20.

39. Inouye M, Dashnow H, Raven L-A, Schultz MB, Pope BJ, Tomita T, et al. SRST2: rapid genomic surveillance for public health and hospital microbiology labs. Genome Med. 2014; 6.

40. Carattoli A, Bertini A, Villa L, Falbo V, Hopkins KL, Threlfall EJ. Identification of plasmids by PCR-based replicon typing. J Microbiol Meth. 2005; 63(3):219–28.

41. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, et al. In Silico Detection and Typing of Plasmids using PlasmidFinder and Plasmid Multilocus Sequence Typing. Antimicrob Agents Chemother. 2014; 58(7):3895–903. doi: 10.1128/AAC.02412-14 PMID: 24777092

42. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comp Biol. 2012; 19(5):455–77.

43. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualisation of de novo genome assemblies. Bioinformatics. 2015.

44. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014.

45. Phan M-D, Kidgell C, Nair S, Holt KE, Turner AK, Hinds J, et al. Variation in Salmonella enterica Serovar Typhi IncHI1 Plasmids during the Global Spread of Resistant Typhoid Fever. Antimicrob Agents Chemother. 2009; 53(2):716–27. doi: 10.1128/AAC.00645-08 PMID: 19015365

46. Jolley K, Maiden M. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics. 2010; 11(1):595.

47. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. Nucleic Acids Res. 2011; 39(Web Server issue):W347–52. doi: 10.1093/nar/gkr485 PMID: 21672955

48. Guy L, Roat Kultima J, Andersson SGE. genoPlotR: comparative gene and genome visualization in R. Bioinformatics. 2010; 26(18):2334–5. doi: 10.1093/bioinformatics/btq413 PMID: 20624783

49. Kidgell C, Pickard D, Wain J, James K, Diem Nga LT, Diep TS, et al. Characterisation and distribution of a cryptic Salmonella typhi plasmid pHCM2. Plasmid. 2002; 47(3):159–71. PMID: 12151231

50. Allue-Guardia A, Imamovic L, Muniesa M. Evolution of a self-inducible cytolethal distending toxin type V-encoding bacteriophage from Escherichia coli O157:H7 to Shigella sonnei. J Virol. 2013; 87 (24):13665–75. doi: 10.1128/JVI.02860-13 PMID: 24109226

51. Jakhetia R, Talukder K, Verma N. Isolation, characterization and comparative genomics of bacteriophage SfIV: a novel serotype converting phage from Shigella flexneri. BMC genomics. 2013; 14(1):677.

52. Roumagnac P, Weill FX, Dolecek C, Baker S, Brisse S, Chinh NT, et al. Evolutionary history of Salmonella typhi. Science. 2006; 314.

53. DeRoeck D, Clemens JD, Nyamete A, Mahoney RT. Policymakers' views regarding the introduction of new-generation vaccines against typhoid fever, shigellosis and cholera in Asia. Vaccine. 2005; 23 (21):2762–74. doi: 10.1016/j.vaccine.2004.11.044 PMID: 15780724

54. IHME. Global Burden of Disease Study 2015 Seattle, USA: Institute for Health Metrics and Evaluation; 2015 [cited 2016 08/12/2016]. http://ghdx.healthdata.org/gbd-results-tool.

55. Holt KE, Perkins TT, Dougan G, Kingsley RA. Genomics and pathogenesis of Salmonella enterica serovars Typhi and Paratyphi A. Salmonella: From genome to function. 2011:107–21.

56. Rychlik I, Gregorova D, Hradecka H. Distribution and function of plasmids in Salmonella enterica. Vet Microbiol. 2006; 112(1):1–10. doi: 10.1016/j.vetmic.2005.10.030 PMID: 16303262

57. Murray BE, Levine MM, Cordano AM, D'Ottone K, Jayanetra P, Kopecko D, et al. Survey of plasmids in Salmonella typhi from Chile and Thailand. J Infect Dis. 1985; 151(3):551–5. PMID: 3882852

58. Jørgensen TS, Xu Z, Hansen MA, Sørensen SJ, Hansen LH. Hundreds of Circular Novel Plasmids and DNA Elements Identified in a Rat Cecum Metamobilome. PLoS ONE. 2014; 9(2):e87924. doi: 10.1371/journal.pone.0087924 PMID: 24503942

59. Buckling A, Rainey PB. Antagonistic coevolution between a bacterium and a bacteriophage. P Roy Soc B-Biol Sci. 2002; 269(1494):931–6.

60. Koskella B, Brockhurst MA. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. FEMS microbiology reviews. 2014; 38(5):916–31. doi: 10.1111/1574-6976.12072 PMID: 24617569

61. Lenski R. Coevolution of Bacteria and Phage: Are There Endless Cycles of Bacterial Defenses and Phage Counterdefenses? J Theor Biol. 1984; 108:319–25. PMID: 6748694

62. Schrag SJ, Mittler JE. Host-parasite coexistence: The role of spatial refuges in stabilizing bacteria-phage interactions. Am Nat. 1996; 148(2):348–77.

63. Kutter E, Raya R, Carlson K. Molecular mechanisms of phage infection. In: Kutter E, Sulakvelidze A, editors. Bacteriophages Biology and Applications. Florida: CRC Press; 2005.

64. Guttman B, Raya R, Kutter E. Chapter 3—Basic Phage Biology. In: Kutter E, Sulakvelidze A, editors. Bacteriophages—Biology and Applications. New York: CRC Press; 2005. p. 29–66.

65. Thomson N, Baker S, Pickard D, Fookes M, Anjum M, Hamlin N, et al. The role of prophage-like elements in the diversity of Salmonella enterica serovars. J Mol Biol. 2004; 339(2):279–300. doi: 10.1016/j.jmb.2004.03.058 PMID: 15136033

66. International Typhoid C, Wong VK, Holt KE, Okoro C, Baker S, Pickard DJ, et al. Molecular Surveillance Identifies Multiple Transmissions of Typhoid in West Africa. PLOS Negl Trop Dis. 2016; 10(9): e0004781. doi: 10.1371/journal.pntd.0004781 PMID: 27657909

67. Pham Thanh D, Thompson CN, Rabaa MA, Sona S, Sopheary S, Kumar V, et al. The Molecular and Spatial Epidemiology of Typhoid Fever in Rural Cambodia. PLoS Negl Trop Dis. 2016; 10(6):e0004785. doi: 10.1371/journal.pntd.0004785 PMID: 27331909

68. Pham Thanh D, Karkey A, Dongol S, Ho Thi N, Thompson CN, Rabaa MA, et al. A novel ciprofloxacin-resistant subclade of H58 Salmonella Typhi is associated with fluoroquinolone treatment failure. eLife. 2016; 5:e14003. doi: 10.7554/eLife.14003 PMID: 26974227