HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2016 June 26.

Published in final edited form as:

Nat Genet. 2015 June ; 47(6): 632–639. doi:10.1038/ng.3281.

Phylogeographical analysis of the dominant multidrug-resistant H58 clade of *Salmonella* Typhi identifies inter- and intracontinental transmission events

A full list of authors and affiliations appears at the end of the article.

Abstract

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].

Correspondence should be addressed to V.K.W. (vw1@sanger.ac.uk).

[52]These authors contributed equally to this work.

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.

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.

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 ($n = 140$) 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 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 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 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.

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. *S*AMtools[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.

Authors

Vanessa K Wong[1,2], Stephen Baker[3,4,5], 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[12,13,14], 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[29,30,31], 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], and Gordon Dougan[1,52]

Affiliations

[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

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.

References

1. Parry CM, Hien TT, Dougan G, White NJ, Farrar JJ. Typhoid fever. N Engl J Med. 2002; 347:1770–1782. [PubMed: 12456854]
2. Connor BA, Schwartz E. Typhoid and paratyphoid fever in travellers. Lancet Infect Dis. 2005; 5:623–628. [PubMed: 16183516]
3. Crump JA, Mintz ED. Global trends in typhoid and paratyphoid fever. Clin Infect Dis. 2010; 50:241–246. [PubMed: 20014951]
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. 2014; 2:e570–e580. [PubMed: 25304633]
5. 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; 9:841–845. [PubMed: 3438648]
6. Fraser A, Paul M, Goldberg E, Acosta CJ, Leibovici L. Typhoid fever vaccines: systematic review and meta-analysis of randomised controlled trials. Vaccine. 2007; 25:7848–7857. [PubMed: 17928109]
7. Khan MI, Ochiai RL, Clemens JD. Population impact of Vi capsular polysaccharide vaccine. Expert Rev Vaccines. 2010; 9:485–496. [PubMed: 20450323]
8. Bhutta ZA. Current concepts in the diagnosis and treatment of typhoid fever. Br Med J. 2006; 333:78–82. [PubMed: 16825230]
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. 1973; 4:597–601. [PubMed: 4602828]
10. Anderson ES. The problem and implications of chloramphenicol resistance in the typhoid bacillus. J Hyg (Lond). 1975; 74:289–299. [PubMed: 1091698]
11. Mirza SH, Beeching NJ, Hart CA. Multi-drug resistant typhoid: a global problem. J Med Microbiol. 1996; 44:317–319. [PubMed: 8636944]
12. Holt KE, et al. Emergence of a globally dominant IncHI1 plasmid type associated with multiple drug resistant typhoid. PLoS Negl Trop Dis. 2011; 5:e1245. [PubMed: 21811646]
13. Chau 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; 51:4315–4323. [PubMed: 17908946]

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. 2010; 48:2171–2176. [PubMed: 20392916]
15. Chiou CS, et al. Antimicrobial resistance in *Salmonella enterica* serovar Typhi from Bangladesh, Indonesia, Taiwan and Vietnam. Antimicrob Agents Chemother. 2014; 58:6501–6507. [PubMed: 25136011]
16. Hendriksen RS, et al. Genomic signature of multi-drug resistant *Salmonella* Typhi related to a massive outbreak in Zambia during 2010 and 2012. J Clin Microbiol. 2015; 53:262–272. [PubMed: 25392358]
17. Roumagnac P, et al. Evolutionary history of *Salmonella* Typhi. Science. 2006; 314:1301–1304. [PubMed: 17124322]
18. Kidgell C, et al. Salmonella Typhi, the causative agent of typhoid fever, is approximately 50,000 years old. Infect Genet Evol. 2002; 2:39–45. [PubMed: 12797999]
19. Holt KE, et al. High-throughput sequencing provides insights into genome variation and evolution in *Salmonella* Typhi. Nat Genet. 2008; 40:987–993. [PubMed: 18660809]
20. Holt KE, et al. Pseudogene accumulation in the evolutionary histories of *Salmonella enterica* serovars Paratyphi A and Typhi. BMC Genomics. 2009; 10:36. [PubMed: 19159446]
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. 2008; 46:1741–1746. [PubMed: 18322069]
22. Holt KE, et al. High-throughput bacterial SNP typing identifies distinct clusters of *Salmonella* Typhi causing typhoid in Nepalese children. BMC Infect Dis. 2010; 10:144. [PubMed: 20509974]
23. Phan MD, et al. Variation in *Salmonella enterica* serovar Typhi IncHI1 plasmids during the global spread of resistant typhoid fever. Antimicrob Agents Chemother. 2009; 53:716–727. [PubMed: 19015365]
24. Le TA, et al. Clonal expansion and microevolution of quinolone-resistant *Salmonella enterica* serotype Typhi in Vietnam from 1996 to 2004. J Clin Microbiol. 2007; 45:3485–3492. [PubMed: 17728470]
25. Holt KE, 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. [PubMed: 21245916]
26. Baker S, et al. Combined high-resolution genotyping and geospatial analysis reveals modes of endemic urban typhoid fever transmission. Open Biol. 2011; 1:110008. [PubMed: 22645647]
27. Holt KE, 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:e1490. [PubMed: 22303491]
28. Lutterloh E, et al. Multidrug-resistant typhoid fever with neurologic findings on the Malawi-Mozambique border. Clin Infect Dis. 2012; 54:1100–1106. [PubMed: 22357702]
29. Phoba MF, et al. Multidrug-resistant *Salmonella enterica*, Democratic Republic of the Congo. Emerg Infect Dis. 2012; 18:1692–1694. [PubMed: 23017665]
30. Breiman RF, et al. Population-based incidence of typhoid fever in an urban informal settlement and a rural area in Kenya: implications for typhoid vaccine use in Africa. PLoS ONE. 2012; 7:e29119. [PubMed: 22276105]
31. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 2007; 7:214. [PubMed: 17996036]
32. Wain J, et al. Quinolone-resistant *Salmonella* Typhi in Viet Nam: molecular basis of resistance and clinical response to treatment. Clin Infect Dis. 1997; 25:1404–1410. [PubMed: 9431387]
33. Holt KE, et al. Multidrug-resistant *Salmonella enterica* serovar Paratyphi A harbors IncHI1 plasmids similar to those found in serovar Typhi. J Bacteriol. 2007; 189:4257–4264. [PubMed: 17384186]
34. Hooper DC. Quinolone mode of action—new aspects. Drugs. 1993; 45(suppl 3):8–14. [PubMed: 7689456]
35. Menezes GA, Harish BN, Khan MA, Goessens WH, Hays JP. Antimicrobial resistance trends in blood culture positive *Salmonella* Typhi isolates from Pondicherry, India, 2005–2009. Clin Microbiol Infect. 2012; 18:239–245. [PubMed: 21714829]

36. Keddy KH, Smith AM, Sooka A, Ismail H, Oliver S. Fluoroquinolone-resistant typhoid, South Africa. Emerg Infect Dis. 2010; 16:879–880. [PubMed: 20409393]
37. Thomson N, et al. The role of prophage-like elements in the diversity of *Salmonella enterica* serovars. J Mol Biol. 2004; 339:279–300. [PubMed: 15136033]
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. 1997; 24:155–167. [PubMed: 9140973]
39. Teplitski M, Goodier RI, Ahmer BM. Pathways leading from BarA/SirA to motility and virulence gene expression in *Salmonella*. J Bacteriol. 2003; 185:7257–7265. [PubMed: 14645287]
40. Lawhon SD, et al. Global regulation by CsrA in *Salmonella typhimurium*. Mol Microbiol. 2003; 48:1633–1645. [PubMed: 12791144]
41. Parkhill J, et al. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. Nature. 2001; 413:848–852. [PubMed: 11677608]
42. Fu Y, Galan JE. A *Salmonella* protein antagonizes Rac-1 and Cdc42 to mediate host-cell recovery after bacterial invasion. Nature. 1999; 401:293–297. [PubMed: 10499590]
43. Ashton PM, et al. MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. Nat Biotechnol. 2015; 33:296–300. [PubMed: 25485618]
44. Holt KE, et al. Shigella sonnei genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe. Nat Genet. 2012; 44:1056–1059. [PubMed: 22863732]
45. Le Hello S, et al. The global establishment of a highly-fluoroquinolone resistant *Salmonella enterica* serotype Kentucky ST198 strain. Front Microbiol. 2013; 4:395. [PubMed: 24385975]
46. Baker S, et al. Fitness benefits in fluoroquinolone-resistant *Salmonella* Typhi in the absence of antimicrobial pressure. eLife. 2013; 2:e01229. [PubMed: 24327559]
47. Bwakura-Dangarembizi M, et al. A randomized trial of prolonged co-trimoxazole in HIV-infected children in Africa. N Engl J Med. 2014; 370:41–53. [PubMed: 24382064]
48. Guy L, Kultima JR, Andersson SG. genoPlotR: comparative gene and genome visualization in R. Bioinformatics. 2010; 26:2334–2335. [PubMed: 20624783]
49. Croucher NJ, et al. Rapid pneumococcal evolution in response to clinical interventions. Science. 2011; 331:430–434. [PubMed: 21273480]
50. Chin CS, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods. 2013; 10:563–569. [PubMed: 23644548]
51. Casali N, et al. Evolution and transmission of drug-resistant tuberculosis in a Russian population. Nat Genet. 2014; 46:279–286. [PubMed: 24464101]
52. Harris SR, et al. Whole-genome analysis of diverse Chlamydia trachomatis strains identifies phylogenetic relationships masked by current clinical typing. Nat Genet. 2012; 44:413–419. [PubMed: 22406642]
53. Harris SR, et al. Evolution of MRSA during hospital transmission and intercontinental spread. Science. 2010; 327:469–474. [PubMed: 20093474]
54. Li H, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078–2079. [PubMed: 19505943]
55. Stamatakis A. RAxML-VI-HPC: maximum likelihood–based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006; 22:2688–2690. [PubMed: 16928733]
56. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics. 2007; 23:127–128. [PubMed: 17050570]
57. Letunic I, Bork P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res. 2011; 39:W475–W478. [PubMed: 21470960]
58. Bollback JP. SIMMAP: stochastic character mapping of discrete traits on phylogenies. BMC Bioinformatics. 2006; 7:88. [PubMed: 16504105]
59. Zerbino DR, Birney E. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. 2008; 18:821–829. [PubMed: 18349386]
60. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014; 30:2068–2069. [PubMed: 24642063]

61. Kurtz S, et al. Versatile and open software for comparing large genomes. Genome Biol. 2004; 5:R12. [PubMed: 14759262]
62. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010; 26:589–595. [PubMed: 20080505]
63. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. Nucleic Acids Res. 2011; 39:W347–W352. [PubMed: 21672955]
64. Inouye M, et al. SRST2: rapid genomic surveillance for public health and hospital microbiology labs. Genome Med. 2014; 6:90. [PubMed: 25422674]
65. Gupta SK, et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother. 2014; 58:212–220. [PubMed: 24145532]
66. Eaves DJ, 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. 2004; 48:4012–4015. [PubMed: 15388468]
67. Baucheron S, Chaslus-Dancla E, Cloeckaert A, Chiu CH, 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. 2005; 49:862–863. [PubMed: 15673791]
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. 2010; 65:1631–1641. [PubMed: 20511368]
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. 2005; 49:457–460. [PubMed: 15616336]
70. Gomes C, et al. In vitro development and analysis of *Escherichia coli* and *Shigella boydii* azithromycin-resistant mutants. Microb Drug Resist. 2013; 19:88–93. [PubMed: 23176550]
71. Marvig RL, et al. Mutations in 23S rRNA confer resistance against azithromycin in *Pseudomonas aeruginosa*. Antimicrob Agents Chemother. 2012; 56:4519–4521. [PubMed: 22644032]
72. Carver T, et al. Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. Bioinformatics. 2008; 24:2672–2676. [PubMed: 18845581]
73. Carattoli A, et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother. 2014; 58:3895–3903. [PubMed: 24777092]

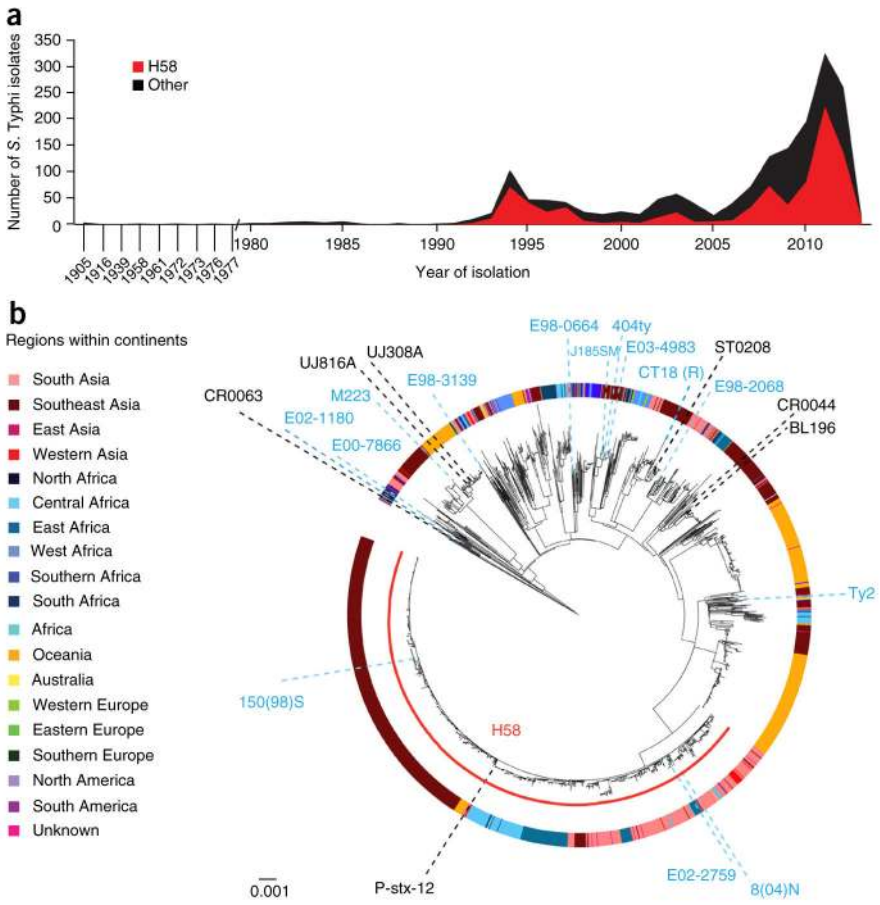

Figure 1.

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.

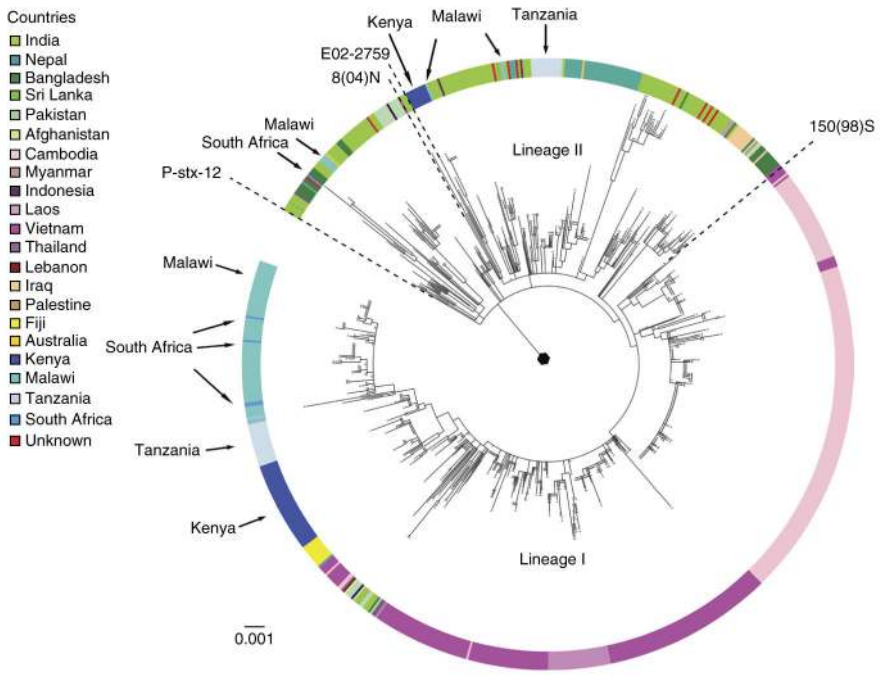

Figure 2.

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.

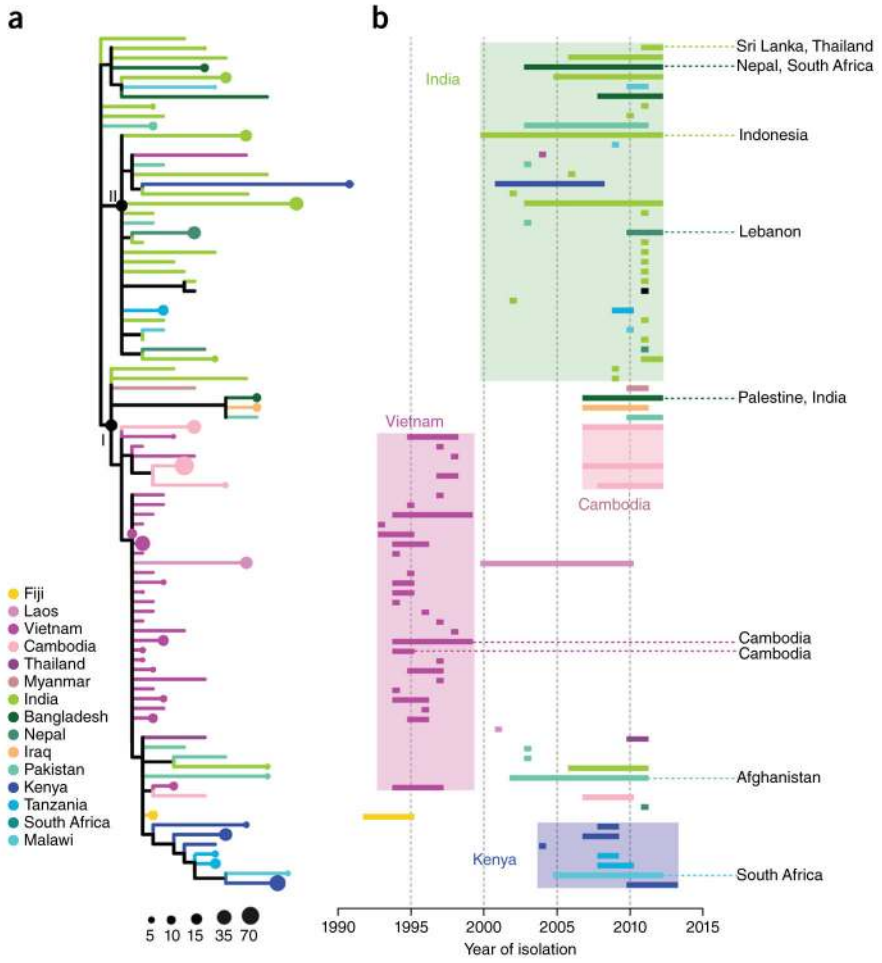

Figure 3.

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.

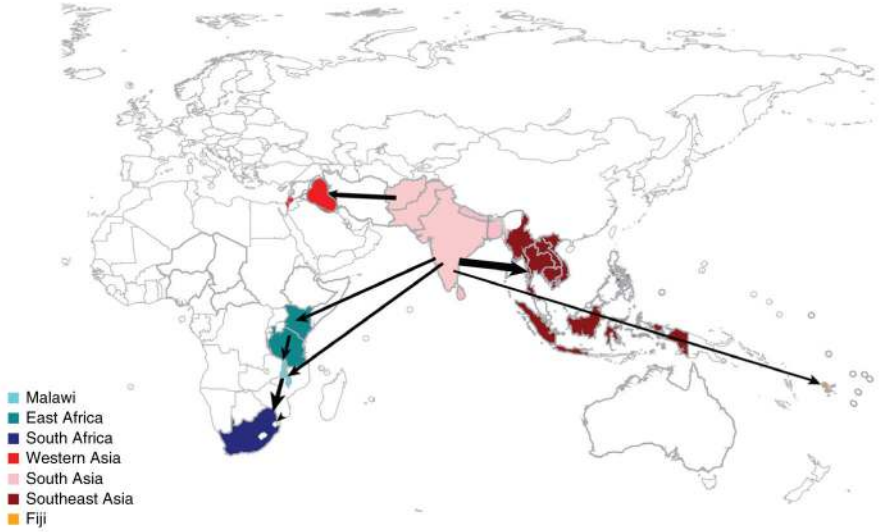

Figure 4.

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.

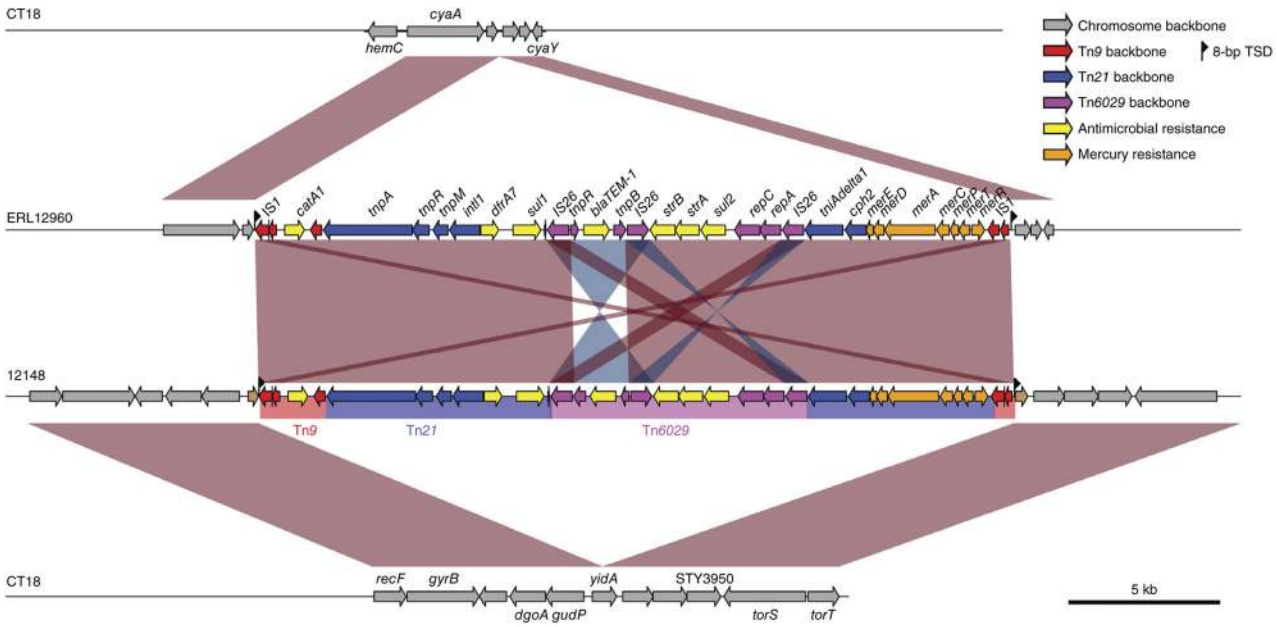

Figure 5.

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.

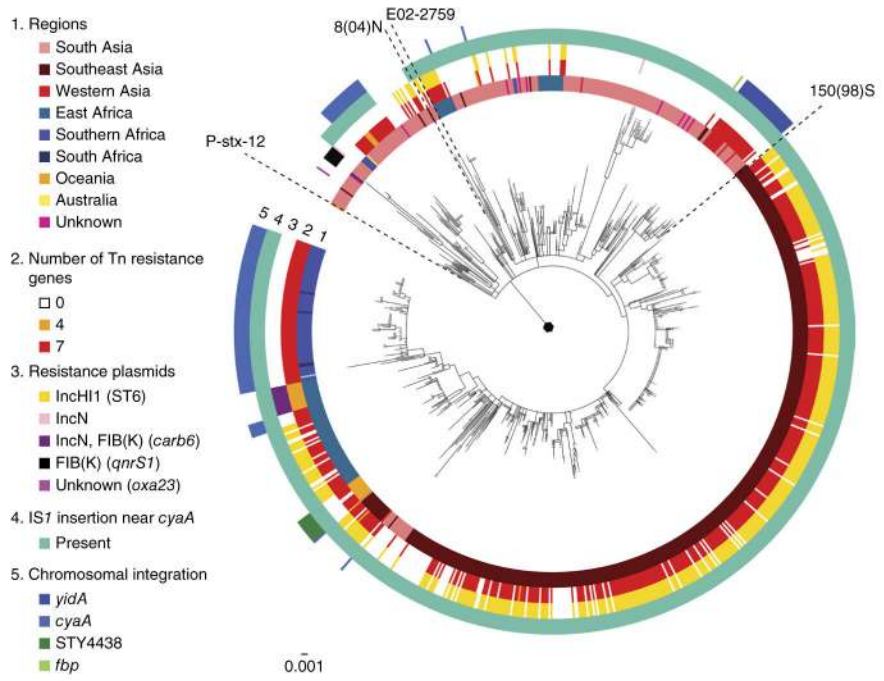

Figure 6.

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.