Global MLST of *Salmonella* Typhi Revisited in Post-genomic Era: Genetic Conservation, Population Structure, and Comparative Genomics of Rare Sequence Types

Kien-Pong Yap[1], Wing S. Ho[1], Han M. Gan[2], Lay C. Chai[1] and Kwai L. Thong[1]*

[1] *Institute of Biological Sciences, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia,* [2] *School of Science, Monash University Malaysia, Bandar Sunway, Malaysia*

Typhoid fever, caused by *Salmonella enterica* serovar Typhi, remains an important public health burden in Southeast Asia and other endemic countries. Various genotyping methods have been applied to study the genetic variations of this human-restricted pathogen. Multilocus sequence typing (MLST) is one of the widely accepted methods, and recently, there is a growing interest in the re-application of MLST in the post-genomic era. In this study, we provide the global MLST distribution of *S.* Typhi utilizing both publicly available 1,826 *S.* Typhi genome sequences in addition to performing conventional MLST on *S.* Typhi strains isolated from various endemic regions spanning over a century. Our global MLST analysis confirms the predominance of two sequence types (ST1 and ST2) co-existing in the endemic regions. Interestingly, *S.* Typhi strains with ST8 are currently confined within the African continent. Comparative genomic analyses of ST8 and other rare STs with genomes of ST1/ST2 revealed unique mutations in important virulence genes such as *flhB*, *sipC,* and *tviD* that may explain the variations that differentiate between seemingly successful (widespread) and unsuccessful (poor dissemination) *S.* Typhi populations. Large scale whole-genome phylogeny demonstrated evidence of phylogeographical structuring and showed that ST8 may have diverged from the earlier ancestral population of ST1 and ST2, which later lost some of its fitness advantages, leading to poor worldwide dissemination. In response to the unprecedented increase in genomic data, this study demonstrates and highlights the utility of large-scale genome-based MLST as a quick and effective approach to narrow the scope of in-depth comparative genomic analysis and consequently provide new insights into the fine scale of pathogen evolution and population structure.

Keywords: MLST, phylogenomic, *Salmonella* Typhi, sequence types, typhoid, whole genome sequencing

INTRODUCTION

Typhoid fever poses a significant health threat to many endemic countries. *Salmonella enterica* serovar Typhi (*S.* Typhi), the etiologic agent, can be transmitted through contaminated food and water via the oral-fecal route. Annually, over 21 million cases and nearly 200, 000 deaths are reported worldwide (Crump and Mintz, 2010). Despite major treatment and prevention efforts,

global typhoid cases remain very high (Lozano et al., 2010; Murray et al., 2010). The disease is also human restricted and the infected individuals could persist as long-term carriers, which in turn serve as the reservoir for new infections and outbreaks (Gonzalez-Escobedo et al., 2010).

The epidemiological investigation of *S.* Typhi is important for disease control such as during a disease outbreak to trace the potential sources. Over the last few decades, many molecular subtyping methods have been applied to genotype bacterial pathogens, and among these, multilocus sequence typing (MLST) is the most commonly used genotyping method to determine the ancestral lineages of many bacteria, including *S.* Typhi (Achtman et al., 2012; Leekitcharoenphon et al., 2012). This method allows discrete characterization of isolates using the internal fragments of housekeeping genes sequences (Urwin and Maiden, 2003; Achtman et al., 2012). The 7 loci MLST scheme of *S.* Typhi was first utilized by Kidgell et al. (2002) to determine *S.* Typhi lineages based on the known sequence types (STs). However, MLST is of limited use for monomorphic pathogen, such as *S.* Typhi, as their populations accrue very limited variations, thus hampering efforts in a population study. In the recent years, high-throughput WGS has become the ultimate approach to study bacterial population structure and phylogeny (Didelot et al., 2012; Chen et al., 2013; Sherry et al., 2013).

Based on MLST, presumably, the global most widespread *S.* Typhi is genetically characterized as ST1 and ST2 in the earlier studies (Kidgell et al., 2002; Zhang et al., 2011; Dahiya et al., 2013; Martínez-Gamboa et al., 2015). However, this conclusion was drawn from analyses of limited numbers of samples. There is also a lack of data on the emergence of uncommon STs which may have been circulating but remain undetected, nor is there a clear answer for the predominance of ST1 and ST2 globally. It is possible that these predominant and clonally related ST1 and ST2 have high transmissibility and/or enhanced virulence to circumvent killing by the innate immune system and probably acquired ability to evade host immune responses, allowing them to establish long-term carriage stage. To date, there are only a few uncommon STs reported for *S.* Typhi. These include ST8-typed isolate (422mar92 from Zaire, Africa, 1992) which was first reported by Kidgell et al. (2002) and ST3 (SARB64, Senegal, 1988; Hangzhou-31, Hang Zhou, China; Kidgell et al., 2002; Wang et al., 2009). Since then, there is no further record for ST8, at least from all available 7,089 *S. enterica* entries in the MLST database (last accessed 10 October 2015). Therefore, it is extremely challenging to gain full evolutionary insights, particularly in regards to virulence and pathogenesis of the pathogen, without first comparing the current widely disseminated populations with the seemingly "near-to-extinct" populations in the evolutionary timescales of *S.* Typhi; much less if these missing-link populations hardly exist or being identified. To test this hypothesis, we performed MLST on both local and global data, either experimentally through conventional MLST or WGS-derived, and scan for any uncommon STs, and to establish the population structure of this pathogen from the global perspective of *S.* Typhi population. Upon identification, the genomes of the uncommon STs were compared against the predominant ST1- and ST2-typed genomes (three reference genomes and six of our previously sequenced genomes). Other publicly available *S.* Typhi genomes were used as background comparison to provide an accurate framework to discern any unique variations carried by these rare populations. Our findings provide valuable information that is important for the understanding of the poor adaptation of *S.* Typhi with uncommon STs, which may otherwise be capable of disseminating globally.

MATERIALS AND METHODS

Local and Regional Bacterial Strains Selection

To determine the STs of *S.* Typhi in this region, we have randomly selected 19 representative strains isolated from various clinical outcomes (outbreaks, sporadic cases, and human carriers). The *S.* Typhi strains were obtained over three decades (34 years; 1983–2008) from 13 distinct geographical locations in three endemic countries (Malaysia, Chile, and Papua New Guinea), which represent three large continents (Southeast Asia, Oceania, and South America). These strains include two each from Papua New Guinea and Chile while the remaining strains are from Malaysia. These strains have been previously described (Thong et al., 1995, 1996a,b, 2002; Supplementary File 1). Although the numbers of strains are relatively small, they were selected from the best available representatives from our collection. The bacterial cultures were further cryopreserved as glycerol stocks at −80°C in 50% glycerol for storage.

Genomic DNA Extraction and MLST

Genomic DNA extracted from overnight cell cultures in Lysogenic Broth (Oxoid, Hampshire, UK) according to the manufacturer's protocol (Promega Corporation, Madison, WI, USA) was used as PCR template. Housekeeping genes, *thrA* (aspartokinase+homoserine dehydrogenase*)*, *purE* (phosphoribosylaminoimidazole carboxylase), *sucA* (alpha-ketoglutarate dehydrogenase), *hisD* (histidinol dehydrogenase), *aroC* (chorismate synthase), *hemD* (uroporphyrinogen III cosynthase), and *dnaN* (DNApolymerase III beta subunit) were targeted for the MLST scheme (Kidgell et al., 2002). PCR assays were carried out with ∼50 ng of DNA template, 150 μM (each) deoxynucleoside triphosphates, 1× PCR colorless buffer, 1.2 mM MgCl$_2$, 0.2 μM of primer, and 0.5 U of Go Taq Flexi DNA Polymerase (Promega, Madison, WI, USA) in 25 μl reaction mixtures. PCR conditions used are as follows; initial denaturation at 94°C for 30 s, 30 cycles of denaturation at 95°C for 30 s, primer annealing at 55°C for 30 s, and extension at 76°C for 30 s; and a final extension at 75°C at 2 min. Reactions were performed using a PCR Master Cycler (Eppendorf AG, Hamburg, Germany). Products were separated by 1.5% agarose gel electrophoresis and visualized with GelRed (Biotium, Inc., Hayward, CA, USA) staining and UV illumination with a gel documentation system (Gel

Doc 2000; Bio-Rad, Hercules, CA, USA). Primers used for the PCR assays are listed in Supplementary File 2. PCR products obtained were purified using PCR purification Kit (MEGAquick-spin[TM], iNtRON Biotechnology, Seongnam, Korea) according to manufacturer's instruction. The purified PCR products were submitted to a commercial sequencing facility for sequencing. Primers used for the sequencing reactions are listed in Supplementary File 2. The seven sequenced loci of 19 local and regional *S*. Typhi strains were first trimmed, edited and aligned using MEGA 6 (Tamura et al., 2013). Sequences were subsequently submitted to the MLST database[1] and assigned existing or novel allele type numbers. The composite sequence STs were defined by the database based on the set of allelic profiles derived from each of the seven loci. Allele sequences for each strain were then concatenated to a final sequence length of 3,336 bp.

Mlst using Whole Genome Sequences of Global Strains

To perform MLST for global strains, all accessible 1,814 whole genome sequences (WGSs) of *S*. Typhi (as of 21 September 2015) and 2 *S*. Paratyphi A strains and their background data were retrieved from NCBI Genome databases via anonymous file transfer protocol (FTP) at ftp://ftp.ncbi.nih.gov/genomes/ or manually downloaded from NCBI. The accession numbers of the reference, our previously sequenced, and published genomes are listed (Supplementary Files 1 and 3). The newly released genomes sequences, raw sequence data, and accession number can be accessed as a batch via European Nucleotide Archive accession number ERP001718 (Wong et al., 2015). The results of the ST assignment and unconfirmed allelic profiles were manually evaluated. In brief, all WGSs were first re-annotated and validated as previously described (Yap et al., 2012a,b) for standardization. The nucleotide sequences were then aligned against MLST database[1] and assigned existing or novel allele type numbers. For validation, the WGSs were also submitted to Centre for Genomic Epidemiology MLST 1.8[2] (Larsen et al., 2012) to assign ST based on the closest matches against the MLST databases. The composite STs (STs) were defined by the database based on the set of allelic profiles derived from each of the seven loci.

MLST Phylogenetic and Data Analyses

All 1,783 concatenated sequences of MLST (local/regional strains and genome-derived; Supplementary File 3) with 100% allele recovery, including the allelic sequences of 2 outgroups sequences extracted from *S*. Paratyphi A complete genomes [*S*. Paratyphi A AKU_12601 (FM200053.1) and *S*. Paratyphi A ATCC 9150 (CP000026.1)] were aligned with MAFFT (Katoh and Standley, 2013) with parameter E-INS-i and a phylogenetic tree was constructed by approximate maximum-likelihood using FastTree2 with the generalized time-reversible model and Bayesian mixture model (GTR+CAT; Price et al., 2010). The

statistical significance of phylogeny was estimated by bootstrap analysis with 1000 pseudoreplicates. Allele numbers and STs of this study were deposited in the publicly accessible *S*. *enterica* MLST database[1]. To determine the genetic polymorphism of *hemD* among *S*. *enterica*, all 333 *hemD* (864 bp) trimmed gene sequences were retrieved from MLST *S*. *enterica* database[1]. A multiple sequence alignment was then built with CLUSTALW in MEGA 6 (Tamura et al., 2013).

Comparative Genomics Analysis of *S*. Typhi with Uncommon and Predominant STs

After the MLST analyses, the eight genomes subtyped with uncommon ST8 (76-1292, E01-5741, 05-8683, 206926, MDUST177, 627334), ST2233 (np45), ST2359 (ST821/98) were retrieved and studied in detail. The assembled genomes were annotated and mapped to the reference genomes (CT18) as previously described (Ho et al., 2012; Yap et al., 2012a). The tRNA and tmRNA were predicted using Aragorn (Laslett and Canback, 2004) whereas the rRNA was predicted with rRNAMMer (Lagesen et al., 2007), and manually validated as described earlier (Osama et al., 2012; Suhaimi et al., 2014). To compare, we have included three reference genomes and six of our own published sequenced genomes (described earlier), representing ST1 (Ty2, P-stx-12, Ty21a, BL196, CR0044) and ST2 populations (CT18, CR0063, UJ308A, UJ816A, ST0208). The pan-genome data of these genomes was obtained using PGAP (Zhao et al., 2012) and compared as previously described (Yap et al., 2014). The phage regions of the pan-genomes identified were predicted and confirmed with web server PHAST (Phage Search Tool; Zhou et al., 2011) and intact phage regions were manually examined. For plasmid analyses, the plasmid replicons were detected using PlasmidFinder 1.3 database (Carattoli et al., 2014) and plasmid MLST was determined using pMLST1.4 (Carattoli et al., 2014) and validated with manual inspection. The acquired antimicrobial genes were identified and determined using ResFinder 2.1 (Zankari et al., 2012). To understand other possible genetic events that shaped the rare STs populations, we performed CRISPRs (clustered regularly interspaced short palindromic repeats) sequences and Restriction–Modification (RM) system prediction analyses using CRISPRFinder (Grissa et al., 2007) and Restriction-ModificationFinder1.1[3], respectively. For virulence genes and *Salmonella* Pathogenicity Islands (SPIs) analyses, the annotated sequences were mapped against Virulence Factor Database (VFDB) and Pathogenicity Island Database (PAIDB) and KEGG pathway (Chelvam et al., 2015), respectively using BLASTn (>98% identity and 60% coverage, E-value < $1x^{-10}$) and Artemis (Carver et al., 2008).

Robust Phylogenomic Analyses using 1,808 Global *S*. Typhi Genomes

Phylogenomic analyses were conducted to better understand the phylogeny of the uncommon STs in the perspective of all *S*. Typhi genomes. To compute large genome samples is

---

[1]http://mlst.warwick.ac.uk

[2]https://cge.cbs.dtu.dk/services/MLST/

[3]https://cge.cbs.dtu.dk/services/Restriction-ModificationFinder/

computationally demanding, time-consuming and technically challenging. Therefore, we utilized a newer approach of alignment-free algorithm using andi v1.4 (Haubold et al., 2015) to rapidly compute large-scale evolutionary distances between 1,808 *S.* Typhi (∼8.6 Gbases; six genomes were excluded from phylogenetic analyses due to low quality) and bias from mapping against reference genome can be avoided. The approach based on a new distance measure, d_a, which approximates local alignments by anchoring them with long, unique matches of a minimal length. The matches are equidistant in the query and the subject is equivalent to restricting the analysis to ungapped alignments. These exact matches were then searched efficiently using enhanced suffix arrays. The anchor distances and arrays were computed using the multithreaded UNIX command-line in the andi v1.4 package (Haubold et al., 2015). The resulting data was aligned, analyzed, transformed and the phylogenetic tree (NJ method) was constructed with SplitsTree4.12.6 (Huson and Bryant, 2005). A rooted tree was inferred using the *S.* Paratyphi A AKU 12601 and *S.* Paratyphi A ATCC 9150 as outgroup. The approach was repeated for 1,806 genomes without the outgroup to generate an unrooted tree. To enhance the visualization of the gigantic tree, the tree was re-rooted with an *S.* Typhi strain from the earliest *S.* Typhi cluster to the outgroup. The node to *S.* Paratyphi A was also collapsed (very long branch length) to improve visualization of the massive tree.

RESULTS

High Allele Recovery Rate in Currently Sequenced *S.* Typhi Genomes and STs Identification from Both Global and Local/Regional Data

Composite allelic profiles were successfully recovered from 1,762 (97%) *S.* Typhi genomes with 100% sequence coverage and 43 (2.3%) genomes using the top BLAST hits criteria (>99% sequence coverage, E-value cut-off <1 × 10^{-10}). The remaining profiles were predicted using the top BLAST hits criteria (coverage less than 99%) against MLST database, which accounted for only less than 0.5% of the genomes studied. Out of the 1,827 global and local/regional strains, 1282 (70.2%) and 536 (29.3%) were subtyped as ST1 and ST2, respectively (one unknown ST strain; Supplementary File 3). Our findings showed that ST1 and ST2 have been homogenously circulating for over the period of 109 years and represent the current predominant populations globally. Both STs were detected as early as in the year 1905, the oldest strain included in our study. Likewise, our local/regional strains exhibited the same pattern of ST1/2 dominance throughout a period of 30 years. We also attempted to detect the geographical structuring of both STs in three levels (continents, region within continents and country), but a very limited evidence was found except for the archipelago of Samoa (116 ST1 and 1 ST2) and Papua New Guinea (1 ST1, 48 ST2) of the Oceania regions, in which an obvious skew toward one ST was observed. We also found no association between STs and

source of isolations in both global and local/regional collection. Interestingly, only 6 strains (76-1292, E01-5741, 05-8683, 206926, MDUST177, 627334) were subtyped as ST8, a rare ST but common in the African regions (Central, North, and South Africa). Remarkably, ST8 was undetected outside of African continent for more than a century. Although, this population was geographically-restricted and seemingly unsuccessful in its dissemination to other parts of the world, the latest ST8-typed strain identified in our study was dated 2012. Since the oldest strain was recovered as earlier as the year 1976, hence, these strains have been circulating in the community for at least 37 years, indicating an establishment of the long-term local reservoir in the African region. It is challenging to confirm whether or not this population has long existed before the aforementioned periods, or represents the ancestral population to ST1 and ST2. Nevertheless, the containment of ST8-typed population only within the African continent but nowhere else is consistent with the notion of the early migration of humans out of Africa, and thus the spread of the pathogens, as proposed by Kidgell et al. (2002). As the bacterium accrues very limited variation, it is highly possible that these two populations (ST1/ST2 and ST8) shared a very recent common ancestor which later diverged into two distinct populations. Apart from ST8, we also detected two uncommon STs, ST2233 (np45) and ST2359 (ST821/98) from South Asia (Nepal) and South America (Argentina), respectively.

High Sequence Conservation of *S.* Typhi from Various Endemic Regions of Typhoid

All seven MLST loci were successfully recovered for the 1,781 *S.* Typhi strains that were obtained from diverse geographical regions (7 continents, 65 countries) and sources of isolation (blood, stool, urine, and environment) over a period of 109 years. The sizes of the trimmed sequenced alleles of all seven loci ranged from 399 to 501 bp. The concatenated sequence amounted to 3336 bp, an approximate 0.07% of the size of *S.* Typhi complete genome (with reference to the CT18 complete genome). The pairwise allelic alignments showed that the average sequence divergence of these loci was very low, which were contributed mainly by *hemD*, which delineates two large populations (ST1 and ST2). The *hemD* gene exhibited only two allelic profiles, 1 and 2 (single synonymous nucleotide polymorphism). The SNP of the *hemD* occurs at position 129 of the trimmed partial sequence where the thymine base of *hemD*1 is replaced by cytosine in *hemD*2. Further investigation on the diversity of *hemD* genes by including a total of 333 *hemD* genes from *S. enterica* (retrieved from MLST database[4] revealed that the polymorphic site (C129T) of the *hemD* is conserved across all members of the group, except for the polymorphism carried by *hemD*1 of *S.* Typhi (Supplementary File 4). Our global phylogenetic analysis of *hemD* gene from *Enterobacteriaceae* members indicated that the *hemD*1 allele may have diverged simultaneously or later from the MRCA, given the fact that

---

[4]http://mlst.warwick.ac.uk

the *hemD*2 allele is conserved throughout the *Salmonella* sp. and its closely related species, *Escherichia coli* and *Shigella* sp. Although, polymorphisms were also detected in *hisD, thrA,* and *dnaN* of the uncommon STs (ST 8, ST2233, ST2359), but these allelic frequencies were extremely low, in which they differed by only one SNP each (ST8-*hisD*1 to *hisD*3; ST2233-*dnaN*1 to *dnaN*478; ST2359-*thrA*5 to *thrA*545). To understand the phylogeny of these STs, an MLST phylogenetic tree of all the *S.* Typhi strains was inferred using the maximum-likelihood approach from the concatenated sequences. We observed two main clonal clade harboring large numbers of ST1 and ST2 strains, which appeared to be a very recent divergence from its common ancestor. The ST8-typed population is phylogenetically close to ST2, whereas ST2359 forms a polytomy with ST1. However, the node containing ST2233 that is closer to the root is polytomically unresolved from ST1/ST2 clades. Although the bifurcation of ST1/ST2 may be the result of the addition of ST2233-typed strain, but the internal nodes of ST8 and ST2359 remain unresolved, most likely the result of short divergence and/or low discriminatory power of MLST (**Figure 1**). Such clonal relationships would be expected for a recent bottleneck that allowed only a few clones to survive, possibly through purifying selection, such as those commonly observed in other monomorphic bacteria.
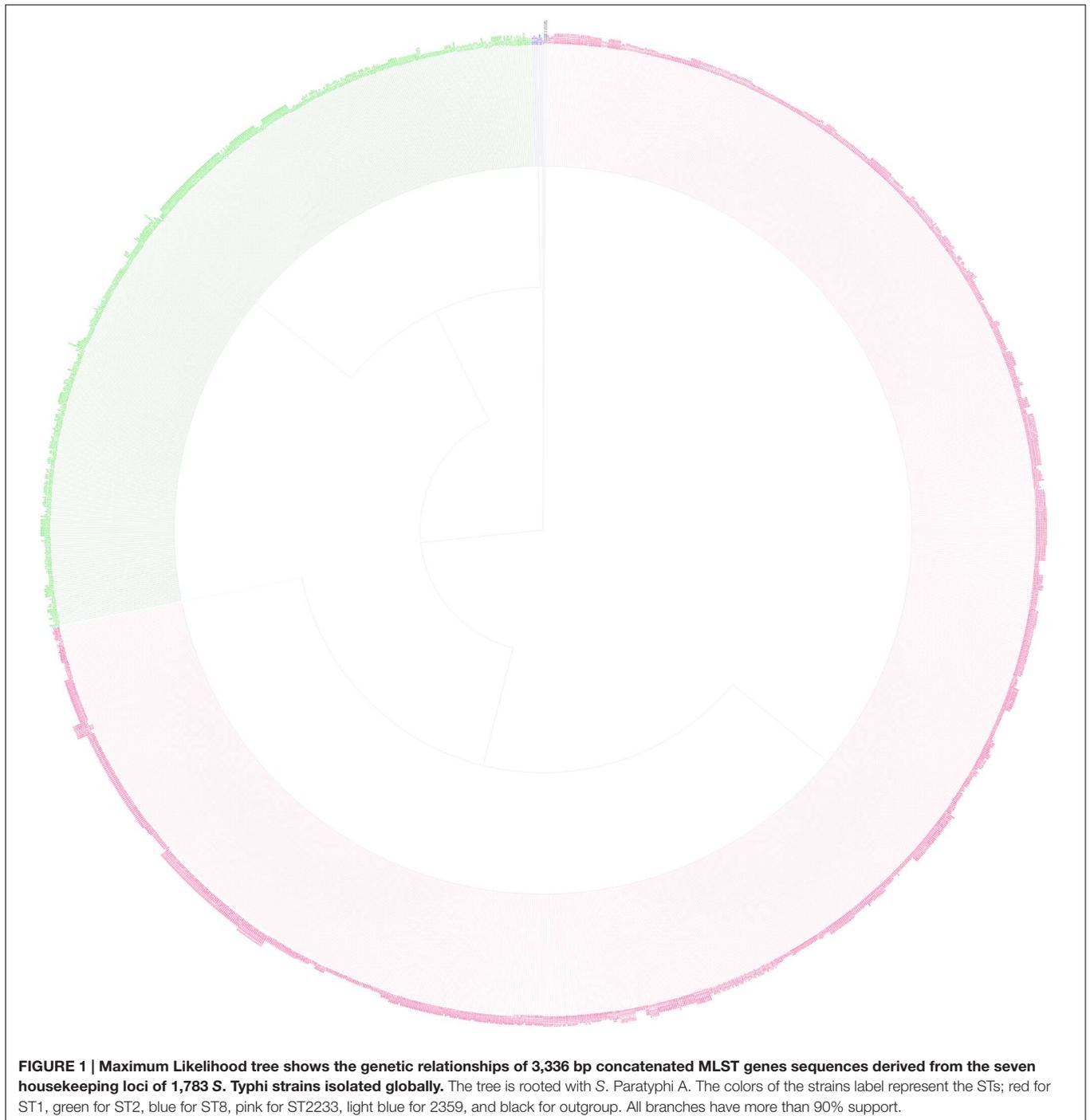
Phylogenomic Analyses Revealed Strong Local Phylogeographical Signals

To view the phylogenetic relationship of these strains with much higher resolution, a whole-genome based phylogenetic tree was built from 1,808 *S.* Typhi genomes (**Figure 2**). This robust phylogenomic tree shows that all *S.* Typhi is monophyletic and distinct from its closest counterpart, *S.* Paratyphi A with long branch length, reiterating the classical phylogeny of these two serovars, which diverged from each other earlier through convergent evolution (Holt et al., 2009). In this phylogeny, there is a clear separation from early diverging basal groups and the core groups that diverge later. Remarkably, these earliest basal groups closest to the root consisting of *S.* Paratyphi A are almost entirely isolated from Africa, which supporting the "theory of human-pathogen co-evolution" since the early human migration out of Africa. Moving further from the roots, the pathogen seems to have widely propagated across various regions, evidenced by several sub-clusters isolated from Africa, South America, Asia, and Australia/Oceania. The frequencies of these basal groups are relatively low, suggesting that these groups may have been under-sampled or reflect true disappearance, plausibly as a result of purifying selection. In contrast, the more recent core clade experienced a radiation that resulted in at least three large clusters with many sub-clusters observed. Interestingly, we detected strong signals of phylogeographical structuring in which some sub-clusters were recovered from the same geographical regions, despite being separated for years, highlighting the importance of local reservoir in maintaining endemicity in those regions. This observation corroborates our earlier study (Yap et al., 2014) in which a human carrier strain (CR0044-isolated from a food handler) was indeed highly related to the outbreak strain (BL196-isolated from a large outbreak in Malaysia), emphasizing the

propensity of human carriers to trigger future outbreaks in the endemic regions. Notably, the ST8-typed population appeared to be ancestrally related to the African and Asia sub-cluster, in which it diverged from the common ancestor to form a distinct terminal cluster. Interestingly, it was observed that ST8-typed population shared the MRCA with a sub-cluster exclusive to Australia/Oceania strains, suggesting that clonally related local strains are circulating in the regions. In contrast, the ST2359-typed ST821/98 was found highly related to the reference genome Ty2 (isolated from Russia) and its derivative vaccine strain Ty21a. Other strains in the sub-cluster of ST821/98 were in fact geographically related, such as those few isolated from Russia and Europe. To note, the close relative of ST821/98; the attenuated Ty21a and its parental strain, Ty2 were both reported to share the same genetic mutant that partly contribute to attenuation of the live vaccine. The ST2230-typed np45 strain was in the major cluster I together with the CT18 and ST0208 (isolated from Southeast Asia). The sub-cluster in which np45 located was geographically restraint to mainly South Asian regions (India and Bangladesh). This sub-cluster was ancestrally related to some other smaller clusters that contain uniformly Asian and African strains, to some extent, indicative of population displacement by the relatively recent clones. In comparing the major clusters of the phylogeny, the strains were separated by such short internal branch length that they appeared to have diverged nearly simultaneously. Most of the sub-clusters are separated by an extremely shallow branch, indicating shared common ancestors, and apparently were quite successful in its dissemination, illustrated by its appearance over several continents. Notably, few strains from the Australia/Oceania regions were detected in the terminal of the relatively longer branching of the deep node, suggesting divergence of variants from the existing nodes, perhaps arose in response to intense selection pressures.

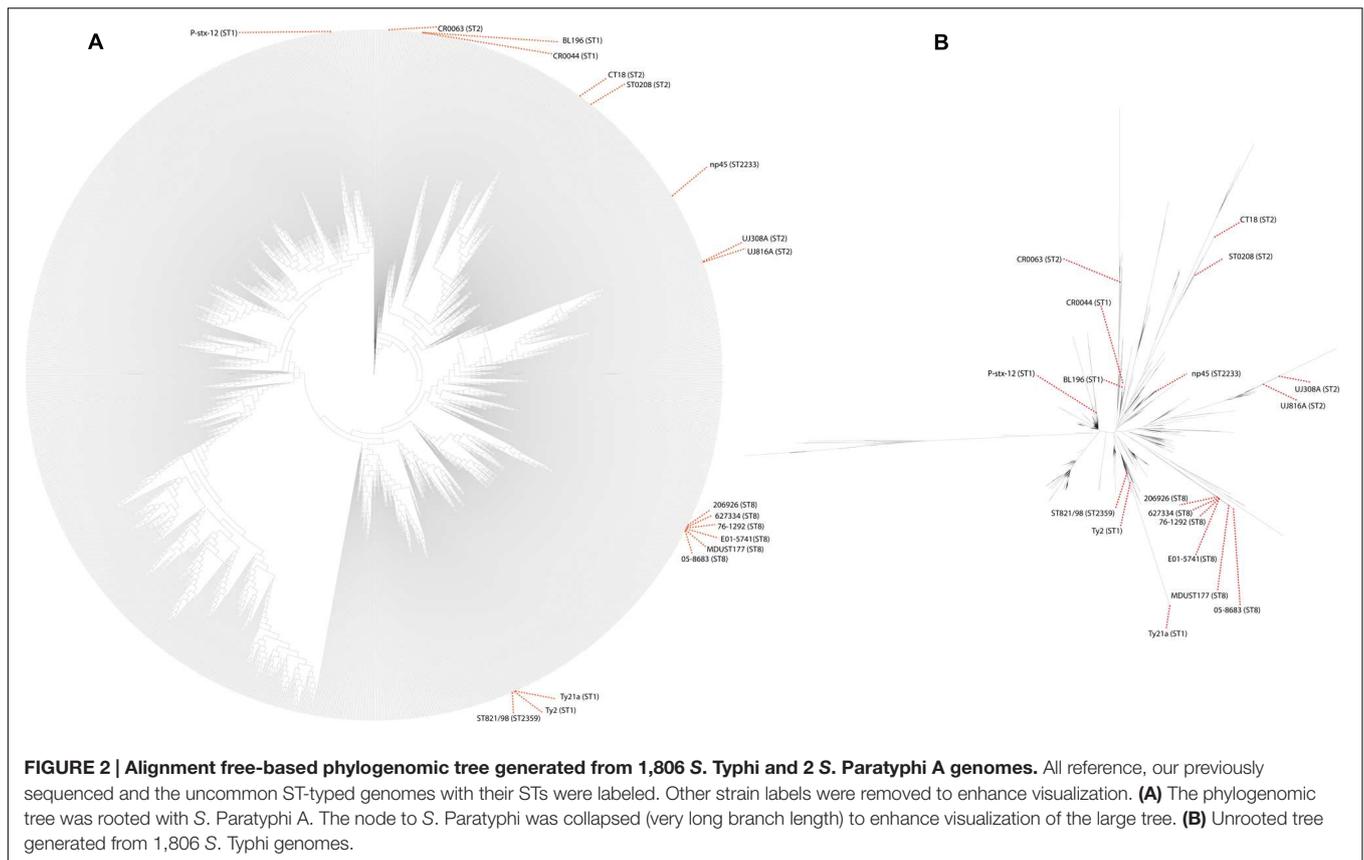## Comparative Genomics of Rare STs Strains Revealed Unique Gene Signatures

Because of the limited dissemination of the rare STs, the ST8-typed strains, in particular, which are not found in other countries outside the African region, we attempted to elucidate the genomic signatures, which might present potential key factors contributing to their poor dissemination. The differences between the seemingly unsuccessful and successful population may explain the evolutionary events that are required for successful infection, transmission, and/or maintenance of carriage state. As the reservoir of *S.* Typhi is mainly human carrier, the low prevalence of this rare STs may reflect its poor adaptation in the host (carriage state), which is required for wide dissemination. This postulation is supported by the fact that the ST 8-typed population was not identified outside of African region, consistent with the route of early human migration out of Africa. In contrast, although ST2233 and ST2359 were rarely identified but their presence outside of African region may reflect new emerging variants. As the current knowledge of *S.* Typhi is solely based on the predominant STs, therefore,

FIGURE 1 | Maximum Likelihood tree shows the genetic relationships of 3,336 bp concatenated MLST genes sequences derived from the seven housekeeping loci of 1,783 *S.* Typhi strains isolated globally. The tree is rooted with *S.* Paratyphi A. The colors of the strains label represent the STs; red for ST1, green for ST2, blue for ST8, pink for ST2233, light blue for 2359, and black for outgroup. All branches have more than 90% support.

we attempted to perform comparative genomics to dissect any genetic signatures of this important population. In terms of basic genomic features, the genome sizes of ST8-, ST2233-, and ST2359-typed isolates ranged from 4.69 to 4.85 Mb with an average GC content of 52.1%. The CDS contents of these genomes ranged from 4839 to 5046 with an average tRNA and rRNA of 71 and 5, respectively. Every genome also carries one tmRNA each. The basic genetic features of these genomes are similar to many previously described genomes (Parkhill et al., 2001;

Deng et al., 2004; Holt et al., 2008). High genomic synteny and conservation, with limited evidence of recombination, were also observed in the uncommon STs-typed strains, which shared similar findings with other previous studies (Holt et al., 2008; Yap et al., 2014).

Salmonella Pathogenicity Islands are unique pathogenicity regions with high numbers of virulence factors (VFs) carried by *S. enterica*, which facilitate the bacterium in host colonization, invasion and maintenance of virulence in the host (Coburn et al.,

FIGURE 2 | Alignment free-based phylogenomic tree generated from 1,806 *S.* Typhi and 2 *S.* Paratyphi A genomes. All reference, our previously sequenced and the uncommon ST-typed genomes with their STs were labeled. Other strain labels were removed to enhance visualization. **(A)** The phylogenomic tree was rooted with *S.* Paratyphi A. The node to *S.* Paratyphi was collapsed (very long branch length) to enhance visualization of the large tree. **(B)** Unrooted tree generated from 1,806 *S.* Typhi genomes.

2007; Dougan and Baker, 2014). SPI-7, the unique virulence region of *S.* Typhi was found intact in all uncommon ST-typed strains. Further, all the strains commonly shared the 227 out of 242 VFs compared. The remaining VFs were shared by at least two genomes except for the strain 76-1292, which carry a unique gene, *astA,* which encode for heat stable enterotoxin. All the compared genomes carried the hypothetical virulence gene, *t0576*, which was absent in CT18. Interestingly, the *tsaC* was only present in the ST8-typed strains and CT18 but lacks in other genomes. Five out of six genomes of ST8-typed strain lack *pilV* but carry the *pilV2*, an alternate and duplicated gene of *pilV*. Notably, four out of six ST 8-typed genomes carried a *seD* gene which was absent in all other genomes. On the contrary, *STY2517* was the only unique gene present in CT18 but absent in others. Notably, Np45 genome which was subtyped as ST2233 lacks VFs needed for secretion machineries such as *sciT*, *ssaV*, and *sthB* although other virulence genes in the respective operons were fully intact, which likely indicates a true absence rather than sequencing artifacts. The VFs with the highest variation were noted in an outer fimbriae gene, *sefC*, although other VFs are highly conserved.

Interestingly, we detected variations in several essential genes, which are involved in virulence and pathogenesis. Synonymous point mutations were identified in three virulence genes; *flgE* (unique SNP), *pipB* (non-unique SNP), *spaO* (non-unique SNP). We also identified shared non-synonymous point mutations in molecular chaperone *clpB* (C338R), fimbriae-like periplasmic protein *Sfm/fimF* (V64A), E3 ubiquitin–protein ligase/*sseJ* (K713E), *hilD* (K209E) and *tviE* (K266N), although not unique to six of the ST8-typed genomes but was identified only in less than 2% of the 1,808 genomes studied. Notably, we also discovered that all ST8-typed genomes shared the identical unique non-synonymous point mutations in three important virulence genes, the *flhB* (E378G), *sipC* (G176S) and *tviD* (R209Q), which were not identified in all other ST1- and ST2-typed populations. In addition, deletions were also detected in two virulence genes, *fimI* (8 bp deletion) and pseudogene *misL* (15 bp deletion), which may lead to a frameshift and premature stop codon. However, we found no unique SNPs of VFs among all eight genomes, except for the *tviD* which was shared by ST821/98 (ST2359), suggesting that they may have undergone different evolutionary paths.

Antimicrobial Resistance Genes Analyses

Of all six ST8-typed strains that were obtained from the African region, only two (76-1292 and 05-8683) acquired antimicrobial resistance genes. In fact, both the 76-1292 and 05-8683 are predicted multidrug-resistance. The 76-1292 strain harbors *sul1* (sulphonamide resistance), *aadA1* (aminoglycoside resistance), *blaOXA-1*, *catA1* (phenicol resistance) and *tetA* (tetracycline resistance) genes, which were found on IncI1 (pMLST 186) and IncP plasmids. In contrast, 05-8683 carried a rare *ereA*, a gene that is associated with azithromycin resistance (macrolide),

apart from other common antimicrobial genes of *catA1* (phenicol resistance), *sul1* (sulphonamide resistance), *tetB* (tetracycline resistance), *dfrA5* (trimethoprim), *blaTEM-1B* (beta-lactam), which were found located in a multireplicon IncF plasmids with pMLST of [F1:A-:B49]. However, both 76-1292 and 05-8683 lacks the IncHI1 plasmid, a replicon type which is commonly associated with the widely disseminated H58-haplotyped *S.* Typhi (Wong et al., 2015). On the contrary, no acquired antimicrobial genes were identified for np45 (ST2233) and ST821/98 (ST2359) strains.

Limited Evidence of Lesion in CRISPRs, Phages, and Restriction–Modification (RM) System

Phage elements were thought to be one the main driving forces of evolution in *S.* Typhi and *S. enterica* (Holt et al., 2008); however, we detected limited evidence of novel intact phages in all eight ST8-, ST2230-, and ST2359-typed strains. In fact, the shared eight regions are highly similar with only slight sequence variations (attributed mainly by the gain/loss of hypothetical and phage structural genes), as identified in the reference genomes. The CRISPRs, a known prokaryotic adaptive immune system that resists invading nucleic acids, may cause DNA mutations and damage (Li et al., 2014). In our analysis, CRISPR spacers and length, direct repeat length, and consensus sequence were commonly shared among genomes, with very little to no variation. We detected an identical six spacers with direct repeats length of 29 bp and with three variants (385, 394, or 420 bp) of CRISPR length. Although, CRISPR has been found to be correlated with STs in some study for subtyping, at least for *S. enterica* (Fabre et al., 2012), we found very limited evidence of variation in *S.* Typhi to be of valuable use in subtyping using CRISPRs. Besides, the RM systems represent known barriers against the entry of foreign DNA. Yet, the RM system is poorly understood in *S.* Typhi. In this study, we revealed extensive repertoires of RM systems, which contain four predicted RM systems (Type I, Type II, Type III, and Type IV) in all studied genomes. The Type I system of *S.* Typhi, which include genes encoding methyltransferase (*M.SptAIII*) and specificity subunits (*S.SptAIII*) were highly similar to that of *S.* Paratyphi A ATCC9150, whereas *EcoKI*, which codes for restriction enzyme is homologous to *E. coli* K-12. The genes (*M.SenAboDcm*, *M.SenSPBIII*, *M.Sen158III*) in Type II code for methyltransferases, which are commonly identified in various *S. enterica*. Type III has both the restriction and methylation domains; the genes *SenAZII* (Restriction enzyme) and *M.SenSPBII* (methyltransferase) were of homologs of *S. enterica* subsp. arizonae serovar 62:z4,z23 and *S.* Paratyphi B SPB7, respectively. For Type IV, only *StyLT2Mrr* was identified, which involves in methyl-directed restriction as carried by the closely related *S.* Typhimurium LT2. The comparative genomics data demonstrated that the RM systems were intact with high similarities, suggesting no functional lesions. Notably, two and only plasmid-bearing ST8-typed *S.* Typhi strains (05-8683 and 76-1292) carried an additional Type 1 *EcoR124I* encoding restriction enzyme (*E. coli*) and Type II *M.EcoGIX* encoding

methyltransferase (*E. coli* O104: H4 str. C227-11), highlighting the potential role of RM systems in plasmid acquisition.

DISCUSSION

In the present study, our MLST data collection for *S.* Typhi, to date, is the largest in numbers and were from the most diverse representatives available (geographical regions, sources) over a period of 108 years. We successfully demonstrated the practicality of scanning 1000s of bacterial genomes. With this scale of data, we also demonstrated unequivocally that ST1 and ST2 are the two most predominant STs globally and are highly successful in dissemination since their emergence. Our results concurred with previous analyses, which relied on limited numbers of *S.* Typhi strains (Kidgell et al., 2002; Zhang et al., 2011; Dahiya et al., 2013). The populations of ST1 and ST2 have been expanding drastically since their emergence from the MRCA and are still co-circulating in the population for at least a century.

Salmonella Typhi, a human-restricted pathogen has long been associated with human migration as well as co-evolution with humans (Kidgell et al., 2002). In this study, no evidence of bias in the spatial and temporal distribution of STs was identified from the MLST data, suggesting that the rates of human migrations across continents have been extremely high to disseminate equal genotypes across the human populations. The *S.* Typhi population may have suffered a recent bottleneck during its course of evolution, evidenced by the limited STs and loci recombination. However, the evolutionary signals from the MLST alone are also rather weak, probably owing to its short evolutionary distance which supports the previous age estimates (∼50,000 years-old; Kidgell et al., 2002) of this relatively "young pathogen." Similar observations have been observed in other monomorphic pathogens such as *Mycobacterium tuberculosis*, the etiologic agent of tuberculosis (Achtman, 2008; Gagneux, 2012).

It is interesting to note that the "primitive" ST3 was completely lost in the population (at least from our data) since the first report of its emergence in Africa in Kidgell et al. (2002). The enigmatic disappearance of this ST, perhaps due to the transient Darwinian selection that rendered the bacteria less fit as a result of harmful mutations accumulated over evolutionary time. However, we cannot rule out the possibility of genetic drift or/and the effects of environmental changes such as those observed in its close relative, *S.* Paratyphi (Zhou et al., 2014). Indeed, the understanding of such evolutionary dynamics is lacking without constructing a genealogy of *S.* Typhi with the inclusion of all populations over a long period of time.

Whilst the identification of uncommon STs were reported intermittently, such as ST890 and ST892 in China (Zhang et al., 2011) and ST1856 in Indonesia recently, these events are extremely rare, and possibly reflects the variants from the existing clones. A similar phenomenon was also observed for ST2233 and ST2259 in this study. On the contrary, the population of ST8 is likely to have poor dissemination and adaptation in humans, in view of its low prevalence globally, as well as being geographically-restricted. It is reasonable to speculate that,

the population may still maintain the ability to cause human carriage, perhaps poorly, given the fact that the population has been around for at least 37 years. This was further supported by the WGS phylogenomic analyses, which revealed that the ST8 population is likely to diverge from the shared common ancestors of ST1- and ST2-typed population, before substantially diluted by the expansion of ST1 and ST2-typed *S.* Typhi out of Africa.

The findings from the comparative analyses of the ST8-, ST2230-, and ST2359-typed genomes relative to ST1- and ST2-typed genomes could be the key factors underpinning the successful traits of predominant populations. As we attempted to interrogate the genomic signatures of these genomes, we found very limited evidence of marked variations in regards to gene gains/loss. The SPI-7, which is implicated as the main virulence repertoires of *S.* Typhi, was completely intact in all *S.* Typhi with uncommon STs, in addition to other SPIs. This finding is consistent with an earlier study that *S.* Typhi bearing SPI-7 was less invasive than the one with the entire SPI-7 missing (Bueno et al., 2004), suggesting that the presence of SPI-7 may be relevant for the fitness of the bacterium. Besides, the genomes harbor very limited numbers of strain-specific genes, which are mainly hypothetical or phages-related. A similar observation was also noted in phage regions, whereby the sequences are highly conserved with little to no variations, except for the few variations punctuated in between the intact phage regions, suggesting the possible hotspots for genome variations. Similarly, the conservation of CRISPRs regions in *S.* Typhi of different STs detected may reflect common defense mechanism acquired by the isolates to fight off incoming foreign DNA. The abovementioned findings are of little surprise as this clonality is well documented in previous studies (Roumagnac et al., 2006; Holt et al., 2008). To probe further, we performed RM system analyses, to detect for any genomic peculiarity of this population. Although, our data indicated no lesion detected in the RM systems; however it is worth mentioning that both the plasmid bearing genomes carried an additional RM systems in comparison with other studied genomes. The presence of these RM systems may facilitate the intake of plasmids/acquired resistance. Such phenomenon has been observed in various bacteria (Oliveira et al., 2014), but its function in *S.* Typhi is so far, remains poorly understood, particularly its role in moderating virulence of the pathogen in the human host. The presence of plasmids in the 76-1292 and 05-8683, confer these strains with MDR phenotype, which may enhance the virulence of these strains. Nonetheless, these plasmids may grant little fitness advantage since the ST8-typed population with the MDR traits were missing in the subsequent isolations. It is likely that the presence of plasmids for maintaining the MDR traits may incur additional fitness burden on the already-less-fit bacterium, thus impacting the rate of successful infection and transmission. Such notion is also implicated in some *S.* Typhi population in which the plasmids were chromosomally integrated (Wong et al., 2015), possibly in the similar efforts to reduce fitness burden.

Intriguingly, our analyses indicated that the salient differences between the ST8- and ST1/ST2-typed populations are the presence of several non-synonymous mutations in few key virulence genes, such as the *flhB*, *sipC*, and *tviD*, which are uniquely present within this ST8-typed population. The affected gene, *sipC*, which encodes for a type III effector protein of the SPI-1 involved in translocation and active modulation, which has been implicated in host invasion. Previous studies have demonstrated that *Salmonella* mutant strains lacking the *sipC* were less invasive (Myeni and Zhou, 2010). An earlier study of *flhB* in *S.* Typhimurium showed that the mutation in the gene (structural gene for hook-associated protein 1; HAP1) resulted in polyhooks and altered flagellar hook length, which may have a severe impact on motility and adhesion of the bacterium (Hirano et al., 1994). The virulence of *S.* Typhi often correlates well with the presence of the Vi antigen (Arricau et al., 1998; Wetter et al., 2012). The mutation found in *tviD*, which is required for biogenesis of the Vi polysaccharide, may potentially affect the virulence of the pathogen.

The genomes of ST8 also harbor a numbers of non-synonymous mutations in several virulence genes (*clpB*, *Sfm/fimF*, *sseJ*, *hilD*, *tviE*) which are involved in invasion, colonization, virulence and regulation in various stages of pathogenesis. Although, these genes are not uniquely acquired, but may be related to the transitionary mutations occur along the evolutionary timeline in gaining fitness advantages. Notwithstanding, it is important to note that although ST1 and ST2 are seemingly successful, but the extent of its fitness remains unclear. The mutations may be accumulative rather than occurring at a single event since some of these mutations were strain-specific, but others are commonly shared unique mutations. Further, loss of function by premature stop codon or truncation from deletion/insertion in the virulence genes of *S.* Typhi is quite rare but the deletion in *fimI* and *misL* may suggest a selective advantage gained for such aberration. The deletion in *fimI*, a gene in type I fimbrial operon, may affect the biofilm formation ability of *S.* Typhi, as reported in some studies of the closest model, the *S.* Typhimurium (Teplitski et al., 2006). The *misL* (located at SPI-3) gene, which is an important colonization factor in the intestinal persistence during the infection of *S.* Typhimurium was disrupted by deletions. Although some essential genes such as *misL*, may be annotated as a pseudogene, the functionality of these pseudogenes in *S.* Typhi remains unclear. For example, *shdA* which was annotated as a pseudogene, was recently found to be functioning in *S.* Typhi (Urrutia et al., 2014), thus the effect of pseudogenization remains speculative. Interestingly, a very limited number of mutations from ST8-typed *S.* Typhi were shared with np45 and ST821/98, indicated that these populations were not closely related, and apparently, they evolved under distinct evolutionary events. These findings are in line with the phylogenomic analyses and MLST data, in which these strains were distantly related to ST8-population, suggestive of variants from the existing clones. However, we have very limited evidence to imply for its poor dissemination and whether or not this population is newly emerged, re-emerging or on the verge of "extinction," explained largely on the basis of intermittent and short encounters, limited numbers of samples available and phylogeographical incongruence.

It remains elusive as to why strains of ST1 and ST2 are evolutionarily selected, albeit *S.* Typhi is thought to be shaped by weak diversifying selection forces along its evolutionary timeline (Roumagnac et al., 2006; Holt et al., 2008). It is reasonable to speculate that these successful STs may have predisposed to advantageous genetic variations that drive evolutions, allowing the pathogens to be more adapted in the host as observed hitherto in several genomic studies of *S.* Typhi (Parkhill et al., 2001; Deng et al., 2004; Baddam et al., 2012a,b; Ong et al., 2012; Yap et al., 2014; Hendriksen et al., 2015). Thus, extensive genomic studies of the *S.* Typhi strains from the endemic regions, particularly with the inclusion of rare subtypes of *S.* Typhi are required to construct full-scale evolutionary history to understand the mutational events that have occurred along the evolutionary timeline.

In summary, although the *S.* Typhi strains were obtained from diverse geographical locations spanning over a century, extremely low divergence was observed with two predominant STs (ST1 and ST2) being identified, a finding that was unequivocally supported by both analyses (experimentally and genome-derived STs). From the large scale scanning, we have identified three highly uncommon STs, the ST8, ST2233, and ST2359. Notably, the ST3 together with two recently reported novel STs (ST890 and ST892) identified in the previous studies are completely absent. The present findings provided strong evidence that *S.* Typhi strains possess a high level of temporal stability and phylogeographical structuring, supported largely by the phylogeographical signals observed in the phylogenomic tree. While many of the genome features have been studied earlier, this work highlighted the genomic signatures of the as-yet-uncharacterized ST8-typed population, which exhibit "unsuccessful trait" that may play vital roles in their poor dissemination. The understanding of these traits may have an important impact on disease control and vaccine development as this population may reflect models of attenuation.

AUTHOR CONTRIBUTIONS

K-PY designed the study, performed the experiment and analyses, and wrote the manuscript. K-PY and WH interpreted data. K-PY, WH, HG, LC, and KT assisted in data interpretation, drafting, and critically reviewed the manuscript and contributed important intellectual input. KT provided funding for the project and supplied equipment and research consumables. KT and LC supervised the project. All authors contributed to the editing of the manuscript and all authors read and approved the final manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2016.00270

REFERENCES

Achtman, M. (2008). Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu. Rev. Microbiol.* 62, 53–70. doi: 10.1146/annurev.micro.62.081307.162832

Achtman, M., Wain, J., Weill, F. X., Nair, S., Zhou, Z., Sangal, V., et al. (2012). Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*. *PLoS Pathog.* 8:e1002776. doi: 10.1371/journal.ppat.1002776

Arricau, N., Hermant, D., Waxin, H., Ecobichon, C., Duffey, P. S., and Popoff, M. Y. (1998). The RcsB–RcsC regulatory system of *Salmonella* Typhi differentially modulates the expression of invasion proteins, flagellin and Vi antigen in response to osmolarity. *Mol. Microbiol.* 29, 835–850. doi: 10.1046/j.1365-2958.1998.00976.x

Baddam, R., Kumar, N., Thong, K. L., Ngoi, S. T., Teh, C. S. J., Yap, K. P., et al. (2012a). Genetic fine structure of a *Salmonella enterica* serovar Typhi strain associated with the 2005 outbreak of typhoid fever in Kelantan, Malaysia. *J. Bacteriol.* 194, 3565–3566. doi: 10.1128/JB.00581-12

Baddam, R., Thong, K. L., Avasthi, T. S., Shaik, S., Yap, K. P., Teh, C. S. J., et al. (2012b). Whole-genome sequences and comparative genomics of *Salmonella enterica* serovar Typhi isolates from patients with fatal and nonfatal typhoid fever in Papua New Guinea. *J. Bacteriol.* 194, 5122–5123. doi: 10.1128/JB.01051-12

Bueno, S. M., Santiviago, C. A., Murillo, A. A., Fuentes, J. A., Trombert, A. N., Rodas, P. I., et al. (2004). Precise excision of the large pathogenicity island, SPI7, in *Salmonella enterica* serovar Typhi. *J. Bacteriol.* 186, 3202–3213. doi: 10.1128/JB.186.10.3202-3213.2004

Carattoli, A., Zankari, E., Garcìa-Fernandez, A., Larsen, M. V., Lund, O., Villa, L., et al. (2014). PlasmidFinder and pMLST: *in silico* detection and typing of plasmids. *Antimicrob. Agents Chemother.* 58, 3895–3903. doi: 10.1128/AAC.02412-14

Carver, T., Berriman, M., Tivey, A., Patel, C., Böhme, U., Barrell, B. G., et al. (2008). Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* 24, 2672–2676. doi: 10.1093/bioinformatics/btn529

Chelvam, K. K., Yap, K. P., Chai, L. C., and Thong, K. L. (2015). Variable responses to carbon utilization between planktonic and biofilm cells of a human carrier strain of *Salmonella enterica* serovar Typhi. *PLoS ONE* 10:0126207. doi: 10.1371/journal.pone.0126207

Chen, C., Zhang, W., Zheng, H., Lan, R., Wang, H., and Du, P. (2013). Minimum core genome sequence typing of bacterial pathogens: a unified approach for clinical and public health microbiology. *J. Clin. Microbiol.* 51, 2582–2591. doi: 10.1128/JCM.00535-13

Coburn, B., Grassl, G. A., and Finlay, B. B. (2007). *Salmonella*, the host and disease: a brief review. *Immunol. Cell. Biol.* 85, 112–118. doi: 10.1038/sj.icb.7100007

Crump, J. A., and Mintz, E. D. (2010). Global trends in typhoid and paratyphoid fever. *Clin. Infect. Dis.* 50, 241–246. doi: 10.1086/649541

Dahiya, S., Kapil, A., Kumar, R., Das, B. K., Sood, S., Chaudhry, R., et al. (2013). Multiple locus sequence typing of *Salmonella* Typhi, isolated in north India-a preliminary study. *Indian J. Med. Res.* 137, 957–962.

Deng, W., Liou, S. R., Plunkett, I. I. I. G., Mayhew, G. F., Rose, D. J., and Burland, V. (2004). Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2 and CT18. *J. Bacteriol.* 185, 2330–2337. doi: 10.1128/JB.185.7.2330-2337. 2003

Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E., and Crook, D. W. (2012). Transforming clinical microbiology with bacterial genome sequencing. *Nat. Rev. Genet.* 13, 601–612. doi: 10.1038/nrg3226

Dougan, G., and Baker, S. (2014). *Salmonella enterica* serovar Typhi and the pathogenesis of typhoid fever. *Annu. Rev. Microbiol.* 68, 317–336. doi: 10.1146/annurev-micro-091313-103739

Fabre, L., Zhang, J., Guigon, G., Le Hello, S., Guibert, V., Accou-Demartin, M., et al. (2012). CRISPR typing and subtyping for improved laboratory surveillance of *Salmonella* infections. *PLoS ONE* 7:e36995. doi: 10.1371/journal.pone. 0036995

Gagneux, S. (2012). Host–pathogen coevolution in human tuberculosis. *Philos. Trans. R. Soc. B. Biol. Sci.* 367, 850–859. doi: 10.1098/rstb.2011.0316

Gonzalez-Escobedo, G., Marshall, J. M., and Gunn, J. (2010). Chronic, and acute infection of the gallbladder by *Salmonella* Typhi: understanding the carrier state. *Nat. Rev. Microbiol.* 9, 9–14. doi: 10.1038/nrmicro2490

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 35, W52–W57. doi: 10.1093/nar/gkm360

Haubold, B., Klötzl, F., and Pfaffelhuber, P. (2015). andi: fast and accurate estimation of evolutionary distances between closely related genomes. *Bioinformatics* 31, 1169–1175. doi: 10.1093/bioinformatics/btu815

Hendriksen, R. S., Leekitcharoenphon, P., Mikoleit, M., Jensen, J. D., Kaas, R. S., Roer, L., et al. (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? *J. Clin. Microbiol.* 53, 677–680. doi: 10.1128/JCM.03104-14

Hirano, T., Yamaguchi, S., Oosawa, K., and Aizawa, S. L. (1994). Roles of FliK and FlhB in determination of flagellar hook length in *Salmonella typhimurium*. *J. Bacteriol.* 176, 5439–5449.

Ho, W. S., Gan, H. M., Yap, K. P., Balan, G., Yeo, C. C., and Thong, K. L. (2012). Genome sequence of multidrug-resistant *Escherichia coli* EC302/04, isolated from a human tracheal aspirate. *J. Bacteriol.* 194, 6691–6692. doi: 10.1128/JB.01804-12

Holt, K. E., Parkhill, J., Mazzoni, C. J., Roumagnac, P., Weill, F. X., 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

Holt, K. E., Thomson, N. R., Wain, J., Langridge, G. C., Hasan, R., Bhutta, Z. A., et al. (2009). Pseudogene accumulation in the evolutionary histories of *Salmonella enterica* serovars paratyphi A and Typhi. *BMC Genomics* 10:36. doi: 10.1186/1471-2164-10-36

Huson, D. H., and Bryant, D. (2005). *Estimating Phylogenetic Trees and Networks Using Splitstree 4*. Available at: www.splitstree.org

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kidgell, C., Reichard, U., Wain, J., Linz, B., Torpdahl, M., and Dougan, G. (2002). *Salmonella* Typhi, the causative agent of typhoid fever, is approximately 50,000 years old. *Infect. Genet. Evol.* 2, 39–45. doi: 10.1016/S1567-1348(02)00089-8

Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H. H., Rognes, T., and Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35, 3100–3108. doi: 10.1093/nar/gkm160

Larsen, M. V., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R. L., et al. (2012). Multilocus sequence typing of total genome sequenced bacteria. *J. Clin. Microbiol.* 50, 1355–1361. doi: 10.1128/JCM.06094-11

Laslett, D., and Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 32, 11–16. doi: 10.1093/nar/gkh152

Leekitcharoenphon, P., Lukjancenko, O., Friis, C., Aarestrup, F. M., and Ussery, D. W. (2012). Genomic variation in *Salmonella enterica* core genes for epidemiological typing. *BMC Genomics* 13:88. doi: 10.1186/1471-2164-13-88

Li, H., Li, P., Xie, J., Yi, S., Yang, C., Wang, J., et al. (2014). New clustered regularly interspaced short palindromic repeat locus spacer pair typing method based on the newly incorporated spacer for *Salmonella enterica*. *J. Clin. Microbiol.* 52, 2955–2962. doi: 10.1128/JCM.00696-14

Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., et al. (2010). Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the global burden of disease study. *Lancet* 380, 2095–2128. doi: 10.1016/S0140-6736(12) 61728-0

Martínez-Gamboa, A., Silva, C., Fernández-Mora, M., Wiesner, M., Ponce de León, A., and Catva, E. (2015). IS200 and multilocus sequence typing for the identification of *Salmonella enterica* serovar Typhi strains from Indonesia. *Int. Microbiol* 18, 99–104. doi: 10.2436/20.1501.01.239

Murray, C. J. L., Vos, T., Lozano, R., Naghavi, M., Flaxman, A. D., Michaud, C., et al. (2010). Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the global burden of disease study. *Lancet* 380, 2197–2223. doi: 10.1016/S0140-6736(12)61689-4

Myeni, S. K., and Zhou, D. (2010). The C terminus of SipC binds and bundles F-actin to promote *Salmonella* invasion. *J. Biol. Chem.* 285, 13357–13363. doi: 10.1074/jbc.M109.094045

Oliveira, P. H., Touchon, M., and Rocha, E. P. (2014). The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res.* 42, 10618–10631. doi: 10.1093/nar/gku734

Ong, S. Y., Pratap, C. B., Wan, X., Hou, S., Rahman, A. Y. A., Saito, J. A., et al. (2012). Complete genome sequence of *Salmonella enterica* subsp. enterica serovar Typhi P-stx-12. *J. Bacteriol.* 194, 2115–2116. doi: 10.1128/JB.00121-12

Osama, A., Gan, H. M., Teh, C. S. J., Yap, K. P., and Thong, K. L. (2012). Genome sequence and comparative genomics analysis of a *Vibrio cholerae* O1 strain isolated from a cholera patient in Malaysia. *J. Bacteriol.* 194, 6933–6933. doi: 10.1128/JB.01832-12

Parkhill, J., Dougan, G., James, K. D., Thomson, N. R., Pickard, D., Wain, J., et al. (2001). Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* 413, 848–852. doi: 10.1038/35101607

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5:e9490. doi: 10.1371/journal.pone.0009490

Roumagnac, P., Weill, F. X., Dolecek, C., Baker, S., Brisse, S., Chinh, N. T., et al. (2006). Evolutionary history of *Salmonella* Typhi. *Science* 314, 1301–1304. doi: 10.1126/science.1134933

Sherry, N. L., Porter, J. L., Seemann, T., Watkins, A., Stinear, T. P., and Howden, B. P. (2013). Outbreak investigation using high-throughput genome sequencing within a diagnostic microbiology laboratory. *J. Clin. Microbiol.* 51, 1396–1401. doi: 10.1128/JCM.03332-12

Suhaimi, N. S. M., Yap, K. P., Ajam, N., and Thong, K. L. (2014). Genome sequence of kosakonia radicincitans UMEnt01/12, a bacterium associated with bacterial wilt diseased banana plant. *FEMS Microbiol. Lett.* 358, 11–13. doi: 10.1111/1574-6968.12537

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

Teplitski, M., Al-Agely, A., and Ahmer, B. M. (2006). Contribution of the SirA regulon to biofilm formation in *Salmonella enterica* serovar typhimurium. *Microbiology* 152, 3411–3424. doi: 10.1099/mic.0.29118-0

Thong, K. L., Cordano, A. M., Yassin, R. M., and Pang, T. (1996a). Molecular analysis of environmental and human isolates of *Salmonella* Typhi. *Appl. Environ. Microbiol.* 62, 271–274.

Thong, K. L., Goh, Y. L., Yasin, R. M., Lau, M. G., Passey, M., Winston, G., et al. (2002). Increasing genetic diversity of *Salmonella enterica* serovar Typhi isolates from papua new guinea over the period from 1992 to 1999. *J. Clin. Microbiol.* 40, 4156–4160. doi: 10.1128/JCM.40.11.4156-4160.2002

Thong, K. L., Passey, M., Clegg, A., Combs, B. G., Yassin, R. M., and Pang, T. (1996b). Molecular analysis of isolates of *Salmonella* Typhi obtained from patients with fatal and nonfatal typhoid fever. *J. Clin. Microbiol.* 34, 1029–1033.

Thong, K. L., Puthucheary, S., Yassin, R. M., Sudarmono, P., Padmidewi, M., Soewandojo, E., et al. (1995). Analysis of *Salmonella* Typhi isolates from

Southeast Asia by pulsed-field gel electrophoresis. *J. Clin. Microbiol.* 33, 1938–1941.

Urrutia, I. M., Fuentes, J. A., Valenzuela, L. M., Ortega, A. P., Hidalgo, A. A., and Mora, G. C. (2014). *Salmonella* Typhi shdA: pseudogene or allelic variant? *Infect. Genet. Evol.* 26, 146–152. doi: 10.1016/j.meegid.2014.05.013

Urwin, R., and Maiden, M. C. (2003). Multi-locus sequence typing: a tool for global epidemiology. *Trends Microbiol.* 11, 479–487. doi: 10.1016/j.tim.2003.08.006

Wang, H. Q., Pan, J. C., Zhang, W., Zheng, W., Yu, H., and Meng, D. M. (2009). Multilocus sequence typing of *Salmonella* Typhi isolates in Hangzhou. *Chin. J. Health Lab. Tech.* 8, 18.

Wetter, M., Goulding, D., Pickard, D., Kowarik, M., Waechter, C. J., Dougan, G., et al. (2012). Molecular characterization of the viaB locus encoding the biosynthetic machinery for Vi capsule formation in *Salmonella* Typhi. *PLoS ONE* 7:e45609. doi: 10.1371/journal.pone.0045609

Wong, V. K., Baker, S., Pickard, D. J., Parkhill, J., Page, A. J., Feasey, N. A., 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

Yap, K. P., Gan, H. M., Teh, C. S., Chai, L. C., and Thong, K. L. (2014). Comparative genomics of closely related *Salmonella enterica* serovar Typhi strains reveals genome dynamics and the acquisition of novel pathogenic elements. *BMC Genomics* 15:1007. doi: 10.1186/1471-2164-15-1007

Yap, K. P., Gan, H. M., Teh, C. S. J., Baddam, R., Chai, L. C., Kumar, N., et al. (2012a). Genome sequence and comparative pathogenomics analysis of a *Salmonella enterica* serovar Typhi strain associated with a typhoid carrier in Malaysia. *J. Bacteriol.* 194, 5970–5971. doi: 10.1128/JB.01416-12

Yap, K. P., Teh, C. S. J., Baddam, R., Chai, L. C., Kumar, N., Avasthi, T. S., et al. (2012b). Insights from the genome sequence of a *Salmonella enterica* serovar Typhi strain associated with a sporadic case of typhoid fever in Malaysia. *J. Bacteriol.* 194, 5124–5125. doi: 10.1128/JB.01062-12

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., et al. (2012). Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* 67, 2640–2644. doi: 10.1093/jac/dks261

Zhang, H., Zhang, X., Yan, M., Pang, B., Kan, B., Xu, H., et al. (2011). Genotyping of *Salmonella enterica* serovar Typhi strains isolated from 1959 to 2006 in China and analysis of genetic diversity by genomic microarray. *Croat. Med. J.* 52, 688–693. doi: 10.3325/cmj.2011.52.688

Zhao, Y., Wu, J., Yang, J., Sun, S., Xiao, J., and Yu, J. (2012). PGAP: pan-genomes analysis pipeline. *Bioinformatics* 28, 416–441. doi: 10.1093/bioinformatics/btr655

Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., and Wishart, D. S. (2011). PHAST: a fast phage search tool. *Nucleic Acids Res.* 39, W347–W352. doi: 10.1093/nar/gkr485

Zhou, Z., McCann, A., Weill, F. X., Blin, C., Nair, S., Wain, J., et al. (2014). Transient Darwinian selection in *Salmonella enterica* serovar paratyphi A during 450 years of global spread of enteric fever. *Proc. Natl. Acad. Sci. U.S.A.* 111, 12199–12204. doi: 10.1073/pnas.1411012111