Whole genome sequence analysis of *Salmonella* Typhi isolated in Thailand before and after the introduction of a national immunization program

Zoe A. Dyson 1,2*, Duy Pham Thanh 3, Ladaporn Bodhidatta 4, Carl Jeffries Mason 4, Apichai Srijan 4, Maia A. Rabaa 3,5, Phat Voong Vinh 3, Tuyen Ha Thanh 3, Guy E. Thwaites 3,5, Stephen Baker 3,5,6† and Kathryn E. Holt 1,2†*

1 Centre for Systems Genomics, University of Melbourne, Parkville, Victoria 3052, Australia
2 Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria 3010, Australia
3 The Hospital for Tropical Diseases, Wellcome Trust Major Overseas Programme, Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam
4 Department of Enteric Diseases, Armed Forces Research Institute of Medical Sciences, Bangkok, Thailand
5 Centre for Tropical Medicine and Global Health, Oxford University, Oxford, United Kingdom
6 The London School of Hygiene and Tropical Medicine, London, United Kingdom

* Corresponding Author
† Joint senior authors

*S. Typhi in Thailand before and after immunization*
Abstract

Vaccines against Salmonella Typhi, the causative agent of typhoid fever, are commonly used by travellers, however, there are few examples of national immunization programs in endemic areas. There is therefore a paucity of data on the impact of typhoid immunization programs on localised populations of S. Typhi. Here we have used whole genome sequencing (WGS) to characterise 44 historical bacterial isolates collected before and after a national typhoid immunization program that was implemented in Thailand in 1977 in response to a large outbreak; the program was highly effective in reducing typhoid case numbers. Thai isolates were highly diverse, including 10 distinct phylogenetic lineages or genotypes. Novel prophage and plasmids were also detected, including examples that were previously only reported in Shigella sonnei and Escherichia coli. The majority of S. Typhi genotypes observed prior to the immunization program were not observed following it. Post-vaccine era isolates were more closely related to S. Typhi isolated from neighbouring countries than to earlier Thai isolates, providing no evidence for the local persistence of endemic S. Typhi following the national immunization program. Rather, later cases of typhoid appeared to be caused by the occasional importation of common genotypes from neighbouring Vietnam, Laos, and Cambodia. These data show the value of WGS in understanding the impacts of vaccination on pathogen populations and provide support for the proposal that large-scale typhoid immunization programs in endemic areas could result in lasting local disease elimination, although larger prospective studies are needed to test this directly.
Author Summary

Typhoid fever is a systemic infection caused by the bacterium *Salmonella* Typhi. Typhoid fever is associated with inadequate hygiene in low-income settings and a lack of sanitation infrastructure. A sustained outbreak of typhoid fever occurred in Thailand in the 1970s, which peaked in 1975-1976. In response to this typhoid fever outbreak the government of Thailand initiated an immunization program, which resulted in a dramatic reduction in the number of typhoid cases in Thailand. To better understand the population of *S. Typhi* circulating in Thailand at this time, as well as the impact of the immunization program on the pathogen population, we sequenced the genomes of 44 *S. Typhi* obtained from hospitals in Thailand before and after the immunization program. The genome sequences showed that isolates of *S. Typhi* bacteria isolated from post-immunization era typhoid cases were likely imported from neighbouring countries, rather than strains that have persisted in Thailand throughout the immunization period. Our work provides the first historical insights into *S. Typhi* in Thailand during the 1970s, and provides a model for the impact of immunization on *S. Typhi* populations.
Introduction

Salmonella enterica subspecies enterica serovar Typhi (S. Typhi) is a human restricted bacterial pathogen and the etiological agent of typhoid fever. S. Typhi is transmitted faeco-orally and can establish asymptomatic carriage in a small subset of an exposed population (1). Recent estimates (2-4) place the global burden of typhoid fever at 25-30 million cases annually, of which 200,000 are associated with deaths. Typhoid fever occurs most commonly in industrialising countries, specifically in locations with limited sanitation and related infrastructure (5); children and young adults are among the most vulnerable populations in these settings (6-8). Immunization and antimicrobial therapy are the major mechanisms by which typhoid fever is controlled (9-12). However, neither of these approaches are optimal and resistance against antimicrobials has become increasingly common in S. Typhi since the 1970s (13-15). Additionally, while a number of typhoid vaccines are licenced for use (9, 16-19), they are not widely used as a public health tools in endemic areas, with the exception of controlling severe outbreaks such as those following natural disasters (20-23).

A sustained typhoid fever outbreak occurred in Thailand in the 1970s. A sharp increase in cases was observed in 1973-1974, which finally peaked in 1975-1976. In response, the government of Thailand established a national typhoid immunization program, which represented the first programmatic use of a typhoid vaccine in the country (24). The immunization program targeted over 5 million school aged children (7-12 years) in Bangkok between 1977 and 1987 (80% of the eligible population). These children received a single locally produced heat/phenol-inactivated subcutaneous dose of $2.5 \times 10^8$ S. Typhi organisms (9, 24). Data from four teaching hospitals in Bangkok showed a 93% reduction in blood culture confirmed infections with S. Typhi between 1976 (n=2,000) and 1985 (n=132) (9, 24). Notably, no significant decline was observed in isolation rates of Salmonella Paratyphi A (S. Paratyphi A), a Salmonella serovar distinct from S. Typhi that causes a clinical indistinguishable disease to typhoid fever, but for which S. Typhi vaccines provide little or no cross-protection (9). This observation suggests that the reduction in S. Typhi infections was
not attributable to improvements in infrastructure and hygiene practices only (5, 9, 21, 24). While the inactivated S. Typhi vaccine was found to be highly efficacious (23, 24), it is no longer used as a consequence of being overly reactogenic (9, 17, 24, 25). A Vi capsular polysaccharide vaccine (16) and live-attenuated oral vaccine of strain Ty21a (17) have since replaced this vaccine for travellers to endemic locations (5, 22, 25).

The typhoid immunization program in Thailand provided a unique opportunity to investigate the impact of immunization on S. Typhi populations circulating within an endemic area. Here we present an analysis of a historical collection of 44 S. Typhi isolates obtained from patients in Thailand between 1973 and 1992 (before and during the immunization program). As S. Typhi populations demonstrate little genetic diversity, we used whole genome sequencing (WGS) to characterise these isolates, and core genome phylogenetic approaches to compare the historic isolates from Thailand to a recently published global S. Typhi genomic framework (4).

Materials and methods

Ethics statement

Salmonella Typhi isolates were collected during febrile disease surveillance studies in Thailand. IRB approval was granted for these studies from the Research Ethics Board of Health (REBH) and the Walter Reed Army Institute of Research (WRAIR) Institutional Review Board, USA. Oral consent was obtained from a parent or guardian at the time of enrolment into the study.

Bacterial isolation and antimicrobial susceptibility testing

Forty-four S. Typhi isolated from patients with suspected typhoid fever attending hospitals in Bangkok, Nonthaburi, Loi, and Srakaew, in Thailand between 1973 and 1992 were available for genome sequencing in this study (Fig 1 and Table S1). At the time of original isolation, bacterial cultures were transferred on nutrient agar slants to the department of Enteric
Diseases, Armed Forces Research Institute of Medical Sciences (AFRIMS), Bangkok, Thailand for identification and antimicrobial susceptibility testing. At AFRIMS, bacterial isolates were subcultured on Hektoen Enteric agar (HE) and identification was performed by biochemical testing on Kligler iron agar slants, tryptone broth for indole, lysine decarboxylase medium, ornithine decarboxylase medium, urease test, mannitol and motility media (Becker Dickenson, Thailand). Serological agglutination was performed using Salmonella O antisera and Salmonella Vi antiserum (Difco, USA). Bacterial strains were stored frozen at -70°C in 10% skimmed milk or lyophilised in 10% skimmed milk; lyophilized ampoules were stored at 2-8°C. Prior to DNA extraction for sequencing, lyophilized bacteria was rehydrated with trypticase soy broth, inoculated on McConkey agar and incubated at 37°C for 18-24 hours. If bacteria was stored frozen in skimmed milk, organisms were inoculated directly onto McConkey agar after thawing and then incubated at 37°C for 18-24 hours.

Antimicrobial susceptibility testing against ampicillin, chloramphenicol, cephalothin, gentamicin, kanamycin, neomycin, sulfisoxazole, trimethoprim/sulfamethoxazole, and tetracycline was performed by disk diffusion according to Clinical and Laboratory Standards Institute (CLSI) (26-29).

**Genome sequencing and SNP analysis**

Genomic DNA from the 44 S. Typhi from Thailand was extracted using the Wizard Genomic DNA Extraction Kit (Promega, Wisconsin, USA). Two μg of genomic DNA was subjected to indexed WGS on an Illumina Hiseq 2000 platform at the Wellcome Trust Sanger Institute, to generate 100 bp paired-end reads. For analysis of SNPs, paired end Illumina reads were mapped to the reference sequence of S. Typhi CT18 (accession no: AL513382) (30) using the RedDog (v1.4) mapping pipeline, available at https://github.com/katholt/reddog. RedDog uses Bowtie (v2.2.3) (31) to map reads to the reference sequence, then high quality SNPs called with quality scores above 30 are extracted from the alignments using SAMtools (v0.1.19) (32). SNPs were filtered to exclude those with less than 5 reads mapped or with
greater than 2.5 times the average read depth (representing putative repeated sequences), or
with ambiguous base calls. For each SNP that passed these criteria in any one isolate,
consensus base calls for the SNP locus were extracted from all genomes (ambiguous base
calls and those with phred quality scores less than 20 were treated as unknowns and
represented with a gap character). SNPs with confident homozygous allele calls (i.e. phred
score >20) in >95% of the S. Typhi genomes (representing a ‘soft’ core genome of common
S. Typhi sequences) were concatenated to produce an alignment of alleles at 45,893 variant
sites. The resultant allele calls for 68 of these SNPs were used to assign isolates to previously
declared lineages according to an extended S. Typhi genotyping framework (33)(code
available at https://github.com/katholt/genotyphi). SNPs called in phage regions, repetitive
sequences (354 kb; ~7.4% of bases in the CT18 reference chromosome, as defined previously
(34)), or recombinant regions (~180kb; <4% of the CT18 reference chromosome, identified
using Gubbins (v1.4.4) (35)) were excluded, resulting in a final set of 1,850 SNPs identified
in an alignment length of 4,275,037 bp for the 44 isolates. For global context, raw read data
(4) were also subjected to genotyping analysis and those isolates sharing the genotypes that
were observed in the Thai collection (n=340) were subjected to the same SNP analyses,
resulting in a final set of 9,700 SNPs for a total of 386 isolates. For each alignment, SNP
alleles from Paratyphi A strain 12601 (36) were also included as an outgroup.

Phylogenetic and SNP analysis

Maximum likelihood (ML) phylogenetic trees (Figs 1-2) were constructed using the 1,850
and 9,700 bp SNP alignments, respectively, using RAxML (v 8.1.23) (37) with a generalized
time-reversible model and a gamma distribution to model site specific recombination
(GTR+Γ substitution model; GTRGAMMA in RAxML), with Felsenstein correction for
ascertainment bias. Support for ML phylogenies was assessed via 100 bootstrap
pseudoanalyses of the alignments. For the larger tree containing global isolates, clades
containing only isolates from only a single country were collapsed manually in R using the
drop.tip() function in the ape package (38). Pairwise SNP distances between isolates were
calculated from the SNP alignments using the dist.gene() function in the ape package for R (38).

Accessory genome analysis

Acquired antimicrobial resistance (AMR) genes were detected, and their precise alleles determined, by mapping to the ARG-Annot database (39) of known AMR genes using SRST2 v0.1.5 (40). Plasmid replicon sequences were identified using SRST2 to screen reads for replicons in the PlasmidFinder database (41). Raw read data was assembled de novo with SPAdes (v 3.5.0) (42) and circular contigs were identified visually and extracted using the assembly graph viewer Bandage (v0.7.0) (43). These putative plasmid sequences were annotated using Prokka (v1.10) (44) followed by manual curation. Where IncHI1 plasmid replicons were identified using SRST2, and their presence confirmed by visual inspection of the assembly graphs, IncHI1 plasmid MLST (pMLST) sequence types were determined using SRST2 (15, 45, 46). Where resistance genes were detected from short read data, Bandage was used to inspect their location in the corresponding de novo assembly graph in order to determine whether they were encoded in the bacterial chromosome or on a plasmid.

Assembled contigs were concatenated and putative prophage genomes were identified with the PHAge Search Tool (PHAST) (47), and their novelty determined by BLASTN analysis against the GenBank database. Pairwise alignments between novel and known prophage sequences were visualised using the genoPlotR package for R (48).

Nucleotide sequence and sequence read data accession numbers

Raw sequence data have been submitted to the European Nucleotide Archive (ENA) under project PRJEB5281; individual sample accession numbers are listed in Table S1. Assembled phage and protein sequences were deposited in GenBank, accession numbers are listed in Table 1.
Results

The population structure of S. Typhi in Thailand

All 44 S. Typhi isolates collected between 1973 and 1992 were subjected to WGS and SNP analysis. Genome-wide SNPs were used to construct a ML phylogeny and isolates were assigned to previously defined genotypes (33) using a subset of SNPs (see Methods). These analyses subdivided the population into ten distinct genotypes, each corresponding to a specific lineage in the ML phylogeny (Fig 1). Genotype 3.2.1 (which includes the reference genome CT18, isolated from Vietnam in 1993 (30)) was the most common (n=14, 32%), followed by genotype 2.1.7 (n=10, 23%). Genotypes 2.0 (n=1, 2%) and 4.1 (n=3, 7%) were observed only in 1973 (pre-vaccine period). Genotypes 2.1.7 (n=10, 23%), 2.3.4 (n=1, 2%), 3.4.0 (n=2, 5%), 3.0.0 (n=3, 7%), 3.1.2 (n=2, 5%), were observed only after 1981 (post-vaccine period). Each of these post-immunization genotypes was from a single location and time period (Fig 1), consistent with short-term localised transmission. The only exceptions were the two S. Typhi 3.1.2 isolates, that were from Srakaew in 1989 and Bangkok in 1992 and separated by just 4 SNPs. Genotypes 3.2.1 and 2.4.0 were observed amongst both pre- and post-vaccine isolates.

Thai S. Typhi in the context of a global genomic framework

Based on the Thai S. Typhi genotyping results we hypothesised that the post-immunization typhoid infections in Thailand resulted from occasional re-introduction of S. Typhi from outside the country, as opposed to long-term persistence of S. Typhi lineages within Thailand.

To explore this possibility, and to provide a global context for our analysis, we examined 1,832 S. Typhi genomes from a recently published global collection that included isolates from 63 countries (4). Genome-wide SNP-based ML trees for each of these genotypes, showing the relationships between Thai and global isolates, are shown in Fig 2. In general, all Thai isolates were closely related to recent isolates sourced from neighbouring countries including Vietnam, Laos and Cambodia (Fig 2), consistent with regional endemic circulation. The S. Typhi genomes in the global collection were mainly isolated 2-3 decades after the Thai
isolates as we did not have access to contemporaneous isolates from these countries that could
identify specific transfer events. However, all but three of the post-vaccine Thai isolates
shared shorter SNP distances with isolates from neighbouring countries than they did with
pre-vaccination Thai isolates (see Fig 3), consistent with these cases being caused by
occasional re-introduction of genotypes circulating in the region. Notably, Thai S. Typhi 3.2.1
that were isolated in 1986-7 clustered separately from the 1973 pre-vaccine isolates (≥60
SNPs apart), but closely with isolates from Vietnam and Cambodia (differing by as few as 7
SNPs; Fig 2H). Post-vaccine Thai S. Typhi 2.4 formed two distinct groups that were not
consistent with direct descendance from earlier isolates (Fig 2E). These data are therefore
consistent with transfer of S. Typhi into Thailand from neighbouring countries during the
post-immunization program era, although the long-term circulation of ancestral populations in
Thailand remains an unlikely alternative explanation.

Acquired antimicrobial resistance
We identified acquired AMR genes in the genomes of four S. Typhi genotype 3.2.1 that were
isolated in Srakaew in 1986 (Fig 1, Table 1). These isolates shared the same four AMR
genes: sul1 (sulphonamides), catA1 (chloramphenicol), tet(B) (tetracyclines), and aadA1
(aminoglycosides) which were carried on near-identical plasmids of IncHI1 plasmid sequence
type 2 (PST2). Although the presence of insertion sequences (IS) in these plasmids prevented
the complete sequences from being assembled, the regions of these plasmids encoding the
AMR genes were identical in all assemblies. This commonality suggests they are a single
plasmid (referred to as pTy036_01 in Fig 1 and Table 1) that was likely acquired in a
common ancestor of this clade. The chromosomal and IncHI1 plasmid sequences for these
four isolates were very closely related to those of a 1993 Vietnamese isolate (Viety1-60_1993) in the global S. Typhi collection (45), consistent with regional transfer.
We identified three non-AMR related plasmids amongst the Thai isolates (Fig 1, Table 1). Ty004 (genotype 2.2) carried two novel plasmids that assembled into circular sequences, pTy004_01 and pTy004_02. The largest, pTy004_01, was a novel variant of the cryptic plasmid pHCM2 (30, 49) (Fig 4). Ty004 was isolated in Bangkok in 1973, making pTy004_01 the earliest example of a pHCM2-like plasmid reported to date. pTy004_01 was distant from other pHCM2-like plasmids in the global S. Typhi genome collection, sharing 92% coverage and 99% nucleotide identity with the reference sequence pHCM2 of S. Typhi CT18 (genotype 3.2.1) which was isolated approximately 20 years later in Vietnam (30). The pTy004_01 sequence (Fig 4) appears to be ~2 kbp larger than pHCM2, and encodes an additional tRNA-Lys as well as an insertion of a hypothetical protein (orf17) into a putative DNA polymerase gene (HCM2.0015c in pHCM2, divided into orf16 and orf18 in pTy004_01). Plasmid pTy004_02 was ~38 kbp in size and similar to E. coli plasmid pEQ2 (65% coverage, 98% nucleotide identity), encoding genes for conjugation, chromosomal partitioning, addiction systems and an abortive infection protein (orf44). Three isolates (Ty031, Ty042, and Ty049) all of genotype 3.0.0 and obtained from Srakaew in 1986, carried a ~40 kbp cryptic plasmid that we named pTy031_01. This plasmid was similar to that carried by Enterobacter hormaechei strain CAV1176 (83% coverage, 96% identity) and encoded genes for chromosomal partitioning, addiction systems, and a putative restriction modification system (orf33-orf34).

PHAST analysis revealed the presence of novel intact prophages in three Thai S. Typhi isolates (Fig 1, Table 1). Two S. Typhi 3.1.2, isolated from Srakaew in 1989 and Bangkok in 1992, shared a novel phage STYP1 that was similar to fiAA91-ss infective for Shigella sonnei (Fig 5A). However, the S. Typhi phage lacked the cytolethal distending toxin cdt genes and the IS21 element found in phage fiAA91-ss (50). This prophage sequence had a mosaic architecture, incorporating a number of putative insertions of phage tail fiber genes that were not present in the fiAA91-ss reference genome (Fig 5A). Additionally, a single isolate of...
genotype 4.1 obtained from Bangkok in 1973 contained a novel SfIV-like phage, here named STYP2, that lacked the serotype conversion gene Gtr cluster and IS1 element of phage SfIV (51). Again, the novel Thai phage variant also encoded novel tail fiber genes not in the SfIV reference genome, as well as a Dam methylase gene (orf37) (Fig 5B)

Discussion

These data provide a historical insight into the population structure of S. Typhi in Thailand in 1973 (pre-immunization program, n=11) and 1981-1992 (post-immunization program, n=33). It has been reported that the national S. Typhi immunization program in Thailand, which commenced in 1977, was highly effective in reducing the burden of typhoid fever (9). Our data are consistent with the hypothesis that the vaccine program successfully depleted the endemic S. Typhi population to the extent that most subsequent typhoid cases resulted from sporadic introduction of non-indigenous S. Typhi, rather than long-term persistence of the pre-vaccine era population. It is apparent that these introductions were sometimes accompanied by limited local transmissions, resulting in small, localized outbreaks, but we found no evidence to suggest that these result in the establishment of stable local source populations. Notably, the post-immunization S. Typhi isolates from Loi (in the north of Thailand near the border with Laos, from which it is separated by the Mekong river) were most closely related to Laos isolates, whilst those from the capital Bangkok and nearby Nonthaburi and Srakaew districts were closely related to other isolates from across Southeast Asia (Fig 2), suggesting there may have been multiple routes of import into Thailand.

Our study is limited by the sample of isolates available for analysis, which was small and reflects opportunistic sampling of sporadic local cases in the four sites and historical storage. However, it is notable that the Thai isolates cluster according to site, consistent with limited local transmission rather than dissemination of lineages between locations. The only exception to this was two genotype 3.1.2 isolates, which were collected from Srakaew in 1989 and Bangkok in 1992 and differed by only 4 SNPs. This is consistent with either
transfer between these cities in Thailand following an initial introduction into the country, or
two independent transfers into Thailand from a common source. The phylogenetic structure is
most suggestive of the latter, but denser samples from Thailand and/or potential source
populations would be required to resolve this with confidence. While our sample is small, this
study is nevertheless the largest to date exploring genetic diversity amongst *S. Typhi* from
Thailand. An earlier global haplotyping study that included seven Thai isolates (52) identified
five distinct haplotypes in Thailand (H3, 1989; H42, 1990; H50, 2002; Vi- H52, 1990; H79,
2002), three of which are related to genotypes that we identified amongst Thai strains in this
study (H79, 2.3.4; H52, 3.4; H42, 3.1.2) (33). Therefore, our genomic snapshot of the Thai
*Typhi* population is consistent with previous insights and is likely reasonably representative
for the study period.

The presence of novel plasmids and prophages in the Thai isolates is also noteworthy. While
small plasmids of unknown function have been observed in *S. Typhi* previously (53), they are
infrequent compared to the IncHI1 MDR plasmid and the cryptic plasmid pHCM2 (54).
Presumably, such plasmids are ephemeral; possibly because their maintenance imposes a
fitness burden on the host cells so a strong selective advantage is required for retention (55,
56). It is also possible that the lack of previous reports regarding the diversity of small
plasmids in *S. Typhi* reflects a technological complexity, however, this is bypassed with high-
throughput WGS and we detected negligible small plasmid content in the global collection of
1,832 genomes using the same screening approach (57). Notably, few of the Thai plasmids
share nucleotide sequence homology with those previously described in *S. Typhi*, but were
closely related to those found in other *Enterobacteriaceae*. The novel pHCM2-like plasmid
(pTy004_01) and two additional plasmids (pTy004_02 and pTy031_01) harbored genes
associated with phage resistance, which could provide protection against phage predation (58-
61). We also observed two novel prophages integrated into Thai genomes, which both
showed variation in their phage tail structural regions compared to close neighbors found in
*Shigella/E. coli*. These regions are typically responsible for binding of phage to host receptors
thus the variation in these regions may be associated with recent adaptations to the S. Typhi host. While genomic data from more recent S. Typhi collections shows limited evidence for genetic exchange with other organisms (4), the detection amongst older Thai isolates of both phage and plasmids that have been previously associated with E. coli/Shigella suggests that genetic exchange may have been more common in the past or in certain localized populations.

Overall, these data provide valuable historical insights into the S. Typhi populations circulating in Thailand during the 1970s and 1980s, and early examples of the two most common S. Typhi plasmids, as well as other mobile elements identified within the S. Typhi population.

Acknowledgements

This project was funded by the Wellcome Trust of Great Britain (106158/Z/14/Z); SB is a Sir Henry Dale Fellow, jointly funded by the Wellcome Trust and the Royal Society (100087/Z/12/Z) and ZAD is funded by strategic award #106158. KEH is supported by fellowship #1061409 from the NHMRC of Australia. DTP is a leadership fellow funded through the Oak Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The view expressed in this article are those of the author(s) and do not reflect the official policy of the Department of the Army, Department of Defense, or the US government.

References


482
484 Characterisation and distribution of a cryptic Salmonella Typhi plasmid pHCM2.
486
distending toxin type V-encoding bacteriophage from Escherichia coli O157:H7 to
489
genomics of bacteriophage SfIV: a novel serotype converting phage from Shigella
492
495
496 53. Holt KE, Perkins TT, Dougan G, Kingsley RA. Genomics and pathogenesis of
497 Salmonella enterica serovars Typhi and Paratyphi A. Salmonella: From genome to
499
500 54. Holt KE, Parkhill J, Mazzoni CJ, Roumagnac P, Weill F-X, Goodhead I. High-
501 throughput sequencing provides insights into genome variation and evolution in
503
504 55. Rychlik I, Gregorova D, Hradecka H. Distribution and function of plasmids in
506
508 Survey of plasmids in Salmonella Typhi from Chile and Thailand. J Infect Dis.


<table>
<thead>
<tr>
<th>Isolate</th>
<th>Genotype</th>
<th>Name</th>
<th>Replicons detected and/or attachment sites</th>
<th>Size (no. putative genes)</th>
<th>Accession number</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>004</td>
<td>2.0.0</td>
<td>pTy004_01</td>
<td>FIB (pHCM2)</td>
<td>108,998 bp (133)</td>
<td>KX833209</td>
<td>Cryptic, Phage defence (Rha protein)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pTy004_02</td>
<td>X1</td>
<td>38,266 bp (49)</td>
<td>KX833212</td>
<td>Phage defence (Abortive Infection)</td>
</tr>
<tr>
<td>031</td>
<td>3.0.0</td>
<td>pTy031_01</td>
<td>N/A</td>
<td>40,835 bp (53)</td>
<td>KX833210</td>
<td>Phage defence (Restriction Modification)</td>
</tr>
<tr>
<td>036</td>
<td>3.2.1</td>
<td>pTy036_01</td>
<td>HH1</td>
<td>~215 kbp</td>
<td>N/A.</td>
<td>AMR (sul1, catA1, tetB, aadA1)</td>
</tr>
<tr>
<td>054</td>
<td>3.1.2</td>
<td>Prophage STYP1</td>
<td>attL CAAGCTGGTCAG</td>
<td>28,946 bp (39)</td>
<td>KX833211</td>
<td>Cryptic</td>
</tr>
<tr>
<td>055</td>
<td></td>
<td></td>
<td>attR CAAGCTGGTCAG</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Summary of mobile genetic elements observed in S. Typhi isolates from Thailand
<table>
<thead>
<tr>
<th>Isolate</th>
<th>Genotype Name</th>
<th>Replicons detected and/or attachment sites</th>
<th>Size (no. putative genes)</th>
<th>Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>013</td>
<td>4.1.0 Prophage STYP2</td>
<td>attL: ATTCGTAATGCGAAGGTAGGTTCGACTCCTATTATCGGCACCAT; attR: ATTCGTAATGCGAAGGTAGGTTCGACTCCT</td>
<td>34,700 bp (50)</td>
<td>KX833213</td>
</tr>
</tbody>
</table>

Cryptic

528
Figure 1. Genomic analysis of Thai S. Typhi.

(A) Maximum likelihood phylogenetic tree (outgroup rooted). Strains are labelled with their three digit name code, year of isolation (pink shading indicates post-vaccine isolates); source location (shaded by city, as indicated in panel B); and plasmid content (any antibiotic resistance genes are indicated in italics). Branch lengths are indicative of the number of SNPs.

(B) Locations from which S. Typhi were isolated in Thailand. (C) Total number of positive blood cultures of S. Typhi (black) and Paratyphi A (grey) between 1970 and 1985; immunization period is indicated in pink; reproduced using data from reference (9).

Figure 2. Zoomed in phylogenies showing relationships of Thai S. Typhi to global isolates. Midpoint rooted ML trees including S. Typhi isolates from the Thai and global collections are shown, for each genotype that was observed amongst the Thai isolates. Colored branches and nodes indicate country of origin, according to the inset legend. Year of isolation is shown to the left; pink and red, Thai isolates obtained before and after the introduction of the immunization program; grey and black, non-Thai isolates obtained before and after the introduction of the immunization program. Thai isolates are also labelled to indicate their city of origin: L, Loi; B, Bangkok; S, Srakaew; N, Nonthaburi. SNP distances between isolates as well as AMR plasmids are labelled, with any resistance genes indicated in italics. Branch lengths are indicative of the number of SNPs.

Figure 3. SNP distances for Thai and global collection isolates.

SNP distance between post-vaccine Thai isolates and their closest pre-vaccine Thai and post-vaccine global collection relatives, colored points indicate country of origin.

Figure 4. Blast comparison of novel plasmid pTy004_01 with pHCM2 (AL513383).

Shaded regions indicate areas of sequence homology, intensity of shading indicates relative nucleotide similarity. Arrows represent protein coding genes, direction indicates coding strand.
Figure 5. Blast comparison of novel phages observed in Thai S. Typhi isolates to nearest known phage sequences.

(A) Novel phage STYP1 compared to Shigella sonnei phage fiAA91-ss (NC_022750). (B) Novel phage STYP2 compared to Shigella flexneri phage SfIV (NC_022749). Shaded regions indicate areas of sequence homology, intensity of shading indicates relative nucleotide similarity. Arrows represent protein coding genes (direction indicates coding strand), colored by encoded protein functions: red, DNA packaging module; orange, virion morphogenesis module; yellow, cargo genes; blue, DNA replication and lysogenic cycle maintenance; green, lysis module.
A) Phylogenetic tree showing the evolutionary relationships of typhoid fever cases. The tree includes the dates and locations of the cases, with branches colored to represent different types of phage.

B) World map highlighting the location of the cases in Thailand.

C) Bar chart showing the number of blood culture confirmed cases of typhoid and paratyphoid per year from 1970 to 1985.