High-Throughput Genotyping of *Salmonella enterica* Serovar Typhi Allowing Geographical Assignment of Haplotypes and Pathotypes within an Urban District of Jakarta, Indonesia

Stephen Baker,1,2* Kathryn Holt,1 Esther van de Vosse,3 Philippe Roumagnac,4 Sally Whitehead,1 Emma King,1 Philip Ewels,1 Andrew Keniry,1 François-Xavier Weill,5 Diane Lightfoot,6 Jaap T. van Dissel,3 Kenneth E. Sanderson,7 Jeremy Farrar,2 Mark Achtman,8 Panagiotis Deloukas,1 and Gordon Dougan1

The Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United Kingdom1; Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam2; Department of Infectious Diseases, Leiden University Medical Centre, Leiden, The Netherlands3; Centre National de la Recherche Scientifique, Saint Paul Lez Durance, France4; Centre National de Référence des Salmonella, Institut Pasteur, Paris, France5; Microbiological Diagnostic Unit-Public Health Laboratory, University of Melbourne, Melbourne, Victoria, Australia6; Department of Biological Sciences, University of Calgary, Calgary, Alberta, Canada7; and Environmental Research Institute, University College Cork, Cork, Ireland8

Received 20 November 2007/Returned for modification 8 February 2008/Accepted 17 February 2008

High-throughput epidemiological typing systems that provide phylogenetic and genotypic information are beneficial for tracking bacterial pathogens in the field. The incidence of *Salmonella enterica* serovar Typhi infection in Indonesia is high and is associated with atypical phenotypic traits such as expression of the *j* and the *z66* flagellum antigens. Utilizing a high-throughput genotyping platform to investigate known nucleotide polymorphisms dispersed around the genome, we determined the haplotypes of 140 serovar Typhi isolates associated with Indonesia. We identified nine distinct serovar Typhi haplotypes circulating in Indonesia for more than 30 years, with eight of these present in a single Jakarta suburb within a 2-year period. One dominant haplotype, H59, is associated with *j* and *z66* flagellum expression, representing a potential pathotype unique to Indonesia. Phylogenetic analysis suggests that H59 *z66* *j* isolates emerged relatively recently in terms of the origin of serovar Typhi and are geographically restricted. These data demonstrate the potential of high-throughput genotyping platforms for analyzing serovar Typhi populations in the field. The study also provides insight into the evolution of serovar Typhi and demonstrates the value of a molecular epidemiological technique that is exchangeable, that is internet friendly, and that has global utility.

*Salmonella enterica* serovar Typhi remains a major public health problem in many developing countries with approximately 22 million cases of typhoid fever reported annually (3, 22). After primary culture, serovar Typhi is identified using serotyping, and strains are subsequently differentiated using phage typing, pulsed-field gel electrophoresis (30), or assays based on PCR including IS200 fingerprinting (31), RAPD (randomly) amplified polymorphic DNA-PCR (26), variable number tandem repeat scanning (18), and amplified fragment length polymorphisms (20). All of these techniques have the potential to distinguish between different isolates, but they have limited or no ability to identify phylogenetic relationships, evolutionary trends, or genotypic characteristics. Techniques such as RAPD-PCR, for example, have considerable inherent experimental variability, and data are not readily comparable between laboratories.

Approaches involving direct interrogation of DNA sequence

---

1 Corresponding author. Mailing address: Oxford University Clinical Research Unit, Hospital for Tropical Diseases, 190 Ben Ham Tu, Quan 5, Ho Chi Minh City, Vietnam. Phone: (84-8) 9 241 761. Fax: (84-8) 9 238 904. E-mail: sbaker@oucru.org.

2 Supplemental material for this article may be found at http://jcm.asm.org/.

3 Published ahead of print on 5 March 2008.
Serovar Typhi populations circulating in a particular geographical region (36). Typhoid fever has a particularly high incidence (810/100,000) and mortality rate in Indonesia (27), where cases are sometimes associated with unusual clinical manifestations, including neurological complications (11). In addition, serovar Typhi isolated in Indonesia can express uncommon flagellum antigens that are not routinely found elsewhere, such as Hj (j) and H:z66 (z66). The j antigen, like the d antigen, is encoded by the chromosomal gene fliC, but the j fliC allele harbors a 251-bp deletion in the variable region of the gene, thus changing the dominant antigenic epitope on the flagellum (6). Expression of j antigen has been associated with a milder clinical presentation of typhoid fever (8). The z66 antigen was first described in 1981, and the genetic determinants for expression were recently shown to be located on a novel linear plasmid pBSSB1 (1, 9, 13). Thus, z66+-serovar Typhi harbor two different flagellin genes, although z66 expression is dominant due to the action of a repressor protein encoded on pBSSB1 that targets the chromosomal fliC gene (2).

It has been hypothesized that serovar Typhi strains from Indonesia have a higher level of genetic diversity than isolates from many other typhoid endemic regions (19). Furthermore, because z66+-strains harbor two flagellin genes they were postulated to be ancestral to global serovar Typhi (7). However, this conclusion was not supported by SNP typing of 105 strains, according to which, all seven z66+-strains were restricted to a local branch (24). In the present study we have used a high-throughput platform to assign 140 Indonesian serovar Typhi isolates to particular haplotypes. Furthermore, we have mapped the origin of some of these haplotyped isolates within a particular geographical area and have associated haplotypes with a particular flagellum antigen type.

MATERIALS AND METHODS

Bacterial strains. The strains used in the present study were isolated either in Indonesia, in the vicinity of Indonesia, or from tourists returning to their native countries who had contracted typhoid in Indonesia; they are from five different sources: (i) Institut Pasteur, Paris, France, 17 strains including 11 controls strains that were described by Roumagnac et al. (24); (ii) The Sanger Institute, Cambridge, United Kingdom, 11 strains, including two z66+-strains from Baker et al. (1) and two strains whose genome has been sequenced (4, 21); (iii) The Sanger Institute, Cambridge, United Kingdom, 27 strains exhibiting an E/H11001 that were described by Roumagnac et al. (24); (iv) Leiden University Medical Centre, Leiden, The Netherlands, 84 strains from Vollaard et al. (35); and (v) Microbiological Diagnostic Unit-Public Health Laboratory, The University of Melbourne, Melbourne, Victoria, Australia, 22 strains isolated from travelers returning from Indonesia. Detailed descriptions of the strains used and the data produced are provided in Table S1 in the supplemental material.

Serovar Typhi genotyping assay. Primers for the study were specifically designed to target the previously described single nucleotide base changes (24) using MassARRAY software and were supplied by Sigma Genosys (Haverhill, United Kingdom). Genotyping was performed by using Sequenom homogeneous mass extend (Sequenom, Inc., San Diego, CA). Multiplexed assays of 1 to 10 SNPs per multiplex were designed by using Sequenom Assay Design v3.0.2.0. Genomic DNA was prepared predominantly using the Wizard Genomic DNA purification kit (Promega) and was diluted to a final concentration of 4 ng/µl. The genomic DNA underwent a locus-specific PCR before being treated with shrimp alkaline phosphatase to remove unincorporated nucleotides. The Sequenom homogeneous mass extend reaction was then performed according to the manufacturer’s specifications. Oligonucleotide primers anneal adjacent to the SNP of interest, and a mixture of terminator nucleotides allow extension through the SNP site, thus creating allele specific extension products each with unique mass. Extension products were then desalted through the addition of an anion-exchange resin, before their weights were analyzed by matrix-assisted laser desorption ionization-time of flight mass spectrometry. In total, 84 DNA fragments were analyzed for calculation of the genotype for each individual bacterial isolate. Genotypes were assigned and viewed by using the SpectroCALLER and SpectroACQUIRE softwares (Sequenom), respectively. The specific haplotypes for all strains were assigned on the basis of the previously described haplotypes (24). A control strain E00-7866 (superscript c) originating from Morocco was originally designated H45 but was redesignated H46 on the basis of our data. The specific polymorphisms in E00-7866 were verified by DNA sequencing, and this strain was found to be correctly assigned to H46. Furthermore, two Indonesian serovar Typhi controls (403Ty and 404Ty) were assigned to H59 by our analysis, rather than to H5 and H2, respectively, as previously described (24). It was subsequently shown that the SNPs differentiating 403Ty H5 and 404Ty H2 from H59 were introduced during laboratory manipulations and were therefore not investigated.

PCR for flagellin gene variation. The fliC gene and the fliB(z66) gene were amplified with primers specific for each locus. Amplification of the fliC gene was performed with the primers fliC_F (TTAACGCAATATAAGGAGAG) and fliC_R (ATGGCACAAGTCTTAAAATAC) and produced a 1,521-bp product for the d allele and a 1,273-bp product for the j allele (6). Amplification of the fliB(z66) gene was performed as previously described (1) using z66Flag_R (ATGGGCGACAAAGTCTAATACT) and z66Flag_R (TATACACCCGACAGACAGTAC). Control PCR amplicons from the aroC gene were produced for all experimental isolates using the primers aroCfor (CCTGGCACCTCGCGCTATAC) and aroCrev (CCACACACGATCGTGGCC) (10). PCR was performed in a 25-µl volume using PCR Supermix Taq polymerase (Invitrogen) and cycled on an MJ Research thermal cycler; the products were analyzed on a 1% agarose gel, and the sizes were estimated by comparison to the migration of Hyperladder I (Bioline).

RESULTS

Utility of a high-throughput genotyping method for the analysis of serovar Typhi. We examined 143 serovar Typhi strains isolated between 1975 and 2005 from typhoid cases originating in or around Indonesia or from travelers returning from Indonesia. These isolates were subjected to conventional H and O typing and many were found to express the antigenic flagella variants j and z66. We have previously described a

FIG. 1. Minimal spanning tree based upon SNPs, demonstrating the relationship of the isolates used in the present study. This illustration is modified from Fig. 1 of Roumagnac et al. (24) and demonstrates the relationship of the 161 strains used here. The 18 serovar Typhi control strains with known haplotypes are labeled green and include CT18 (superscript a), Ty2 (superscript b), E00-7866 (superscript c), and AG3 (superscript d). The 140 experimental serovar Typhi strains are distinguished by red clusters, and the three serovar Paratyphi A isolates are colored yellow. The specific haplotype for each cluster is labeled (e.g., H45), and undetected haplotypes are gray. The size of each cluster corresponds to the number of strains identified in each group; this is demonstrated by the key in the bottom right corner.
scheme based on denaturing high-performance liquid chroma-
tography analysis that facilitates the assignment of serovar Typhi isolates to individual haplotypes that can be placed on a phylogenetic tree illustrative of the evolution of this serovar (24). We included 18 strains that had been previously haplo-
typed (24) as controls; therefore, the total number of isolates analyzed was 161. SNPs are frequently used in human geno-
typing as markers for regions of genome variation and as refer-
ence points to map genetic associations. Consequently, we
selected the Sequenom high-throughput genotyping platform
that is routinely used for analyzing mammalian DNA for SNP
typing of serovar Typhi. The design of the Sequenom serovar
Typhi SNP assay permitted the simultaneous detection of 84
SNPs in the serovar Typhi isolates, and the full data set is
provided in Table S1 in the supplemental material.

The minimal spanning tree assembled from the raw data is
shown in Fig. 1 and is adapted from Roumagnac et al. (24).
Sequenom analysis of the 161 serovar Typhi isolates correctly
assigned all control strains to the predicted haplotype (green
circles; Fig. 1). The novel Indonesian serovar Typhi isolates
were assigned to nine different haplotypes (red circles in Fig.
1), five of which also contained control isolates (green circles
within red circles in Fig. 1). Significantly, the remaining 41
haplotypes reported previously in a global serovar Typhi col-
lection, including three haplotypes described in Indonesia—
H14, H58, and H70 (24), were undetected (gray circles in Fig.
1). Therefore, only a limited selection of the known haplotypes
in the serovar Typhi population are circulating in this region.

The assay was additionally validated by assigning three ex-
perimental isolates to the ancestral haplotype, H45 (yellow
circle in Fig. 1). These isolates were originally designated ser-
ovar Typhi, but when reserotyped they were found to be sero-
var Paratyphi A (588, 618, and 2585; see Table S1 in the
supplemental material). All of the SNPs in the assay are spe-
cific for serovar Typhi; therefore, analysis of any other Salmon-
ella serovars with this SNP assay would assign them to the
ancestral haplotype H45.

Haplotypes H59 and H8 accounted for 53% (73) and 24%
(34), respectively, of the serovar Typhi isolates recovered, and
all haplotypes, except H84, included more than one isolate.
Serovar Typhi haplotypes H42, H45, H50, H52, and H85 have
been described previously as originating from geographically
dispersed locations including Asia, Africa, and South America,
providing evidence for the global dissemination of some strains
(24). In contrast, H59 appears to be restricted to Indonesia and
may be a recently evolved variant. H59 strains were identified
in all five collections (see Materials and Methods) of experi-
mental strains and were consistently isolated over a 30-year
period (see Table S1 in the supplemental material). In addi-
tion, H42, H50, and H85 were also identified in different col-
lections, with H85 representatives being isolated as far apart as
1987 and 2003 (see Table S1 in the supplemental material).
This supports our previous observations that serovar Typhi
strains can persist in a particular geographical region for pe-
riods extending several decades (24).

Serovar Typhi strains of haplotype H58 are frequently iso-
lated in many parts of Southeast Asia and are associated with
resistance to nalidixic acid, which may be a direct consequence
of the common use of fluoroquinolones (17). We were unable
to identify any H58 strains within these experimental Indone-
sian strains (superscript d in Fig. 1), although one fluoroquin-
olone-resistant strain belonging to H58 has been detected from
a French traveler returning from Indonesia in 2003 (24). Im-
portantly, our assay also interrogated the DNA for mutations
in DNA gyrase, which leads to fluoroquinolone resistance (34).
No such mutations in DNA gyrase were detected in this ex-
perimental population of serovar Typhi. This result indicates
that there has been no recent clonal expansion of H58 in
Indonesia, regardless of the fact that fluoroquinolone-resistant
H58 strains may have been introduced in this country from
near neighbors, such as Vietnam, where such strains are com-
mon. These data concur with those of other studies demon-
strating the antimicrobial susceptibility of serovar Typhi and
other enteric bacteria in Indonesia (14, 32) and may indicate
a different epidemiological dynamic of the disease in this
country.

Multiple serovar Typhi haplotypes are circulating in a local
region. Serovar Typhi described in collection 4 (see Table S1
in the supplemental material) were isolated between June 2001
and February 2003 in the Jatinegara district in Eastern Jakarta
in a study by Vollaard et al. (35). To understand the circulation
of serovar Typhi strains in this local region, we attempted to
retrospectively correlate data, locating the residence of the
individual typhoid patients with a specific serovar Typhi hap-
lotype. A map demonstrating the distribution of individual
serovar Typhi strains isolated within the ~10.5-km² Jatinegara
district is shown in Fig. 2. We were able to pinpoint the serovar
Typhi haplotype from all of the residence of the patients for
which we had residential mapping data (54 of 84 cases in total).
We identified eight circulating haplotypes in the Jatinegara
district, providing unequivocally evidence that there are inde-
pendent clusters of typhoid transmission involving distinct
haplotypes in this small area. Four serovar Typhi isolates of
identical haplotype (H59) were isolated in the same household
(red circle within a black circle), suggesting direct horizontal
transmission. The methodology presented here indicates that
haplotyping data may prove to be a powerful tool to assist the
tracking of serovar Typhi within a community by stratifying the bacterial population in a reliable and reproducible manner. A prospective study combining mapping information and haplotyping data would give insight into the mechanisms of transmission and long-term carriage within a population where typhoid is endemic.

**Flagellin gene variation.** Serovar Typhi Indonesian isolates can express h, j, or z66 flagellum antigens. Consequently, we used PCR analysis of the fliC gene and the fljB(z66) gene using DNA extracted from all 140 experimental strains to determine which flagellin type they encoded. The primers were designed so that the fliC d flagellin allele would generate a DNA fragment of 1,521 bp compared to one of 1,273 bp for the j antigen (data not shown). The globally ubiquitous d fliC allele amplicon was detected in 85 of 140 (61%) of serovar Typhi and was dominant in all haplotypes except for the Indonesian specific H59 group (gray in Fig. 3A). The smaller fliC j allele, which was amplified in the remaining 55 isolates, was dominant in group H59 and was additionally identified in two other haplotypes (white in Fig. 3A).

Similarly, presence of the fljB(z66) gene on the pBSSB1 linear plasmid was confirmed by amplification of a 1,479-bp product specific to this region (data not shown). A 1,479-bp amplicon particular for the fljB(z66) gene was generated in 59 of 140 isolates, all of which fall into the H59 haplotype (gray shading in Fig. 3B). In addition, when the fljB(z66) data are combined with the fliC data the majority of strains (43 of 59) that can express the z66 antigen harbored a fliC j allele on the chromosome. Haplotyp e H59 contains four different sub-

**DISCUSSION**

Serovar Typhi is a young, genetically monomorphic bacterial pathogen that has not been in existence long enough to generate extensive sequence polymorphisms (15, 23–25). Here, we have investigated the population structure of serovar Typhi associated with Indonesia by using SNP typing and flagellin-based PCR analysis. In the collections we identified nine haplotypes composed of two dominant and seven less common haplotypes. Our data show that in a relatively small area (~10.5 km²) there are several different haplotypes of serovar Typhi circulating within the local population, suggesting the existence of independent transmission clusters. These data are in agreement with other studies, which report multiple strain types circulating within a specific location (28–30). Furthermore, we identified bacteria of the same haplotype isolated over a 30-year period, providing further support for a vital role of persistent carriage and/or prolonged dissemination of serovar Typhi in the environment.

Typhoid fever remains a significant public health problem in many parts of Southeast Asia. Dissemination of serovar Typhi is of particular importance considering the wave of multiple drug-resistant strains that is currently spreading across Asia and the movement of people associated with current rapid economic growth. There is a pressing need for more effective epidemiological scrutiny of this organism, enabling better understanding of the nature of spread, which will ultimately facilitate control policies. Molecular methods appear to be the most robust, but such methods require standardization for the communication and analysis of the resulting data. The SNP-based approach is reproducible, results are applicable in any setting, and the data can be integrated into datasets produced in multiple laboratories using a variety of SNP detection technologies. This is in contrast to other schemes that detect strain differences but produce variable results in different laboratory settings. In addition, such schemes typically provide little or no information on phylogeny or phenotype and are unable to identify clonal expansion in a region.

The high-throughput system utilized here has been used to perform MLST on a number of Neisseria meningitidis reference strains (12). We have demonstrated the effectiveness of the same platform for detecting haplotypes in a genetically monomorphic gram-negative bacterial pathogen. Further flexibility can be anticipated when additional sensitivity is incorporated after more SNP variation is discovered.

Although the system has benefits, high-throughput SNP typing is currently not suitable for routine use in laboratories without specific infrastructure or financial support to run such assays. However, SNP detection technology has been driven in recent years by the need for massive throughput (currently >1 million SNPs for human genotyping studies). Microbial genotyping is much less demanding, with studies such as this providing epidemiologically informative data by assaying fewer than 100 SNP loci. With current high-throughput SNP detec-
tion technology, running costs for bacterial genotyping assays should be feasible for clinical laboratories, although setup costs may be prohibitive in some settings. The development of novel medium-throughput SNP detection technologies will be an important step toward the adoption of SNP-based microbial genotyping in clinical laboratories worldwide.

A practical solution for short-term development would be to house SNP typing facilities in regional reference laboratories linked via the internet through a global database. Simple “in the field” assays can also be established to detect a subset of discriminatory SNPs using a targeted approach. For example, MLST is performed routinely in many laboratories, while PCR analysis of nine gene fragments would have detected all of the haplotypes circulating in the present study. The resulting data would provide information about population structure in a particular region. Indeed, we are currently developing such a method based on PCR that can distinguish many of the major serovar Typhi haplotypes, which could potentially be universally adopted in clinical and research laboratories alike. However, it is important to note that the validity of this kind of simplified genotyping scheme depends on prior knowledge of the underlying phylogeny provided by comprehensive SNP typing studies.

It was originally hypothesized that z66-serovar Typhi isolates from Indonesia were the precursor to global serovar Typhi, and it was assumed that the possession of a second phase of the flagella antigen was an ancestral state which has subsequently been lost (19). Conversely, an “out of Africa” hypothesis was proposed as the global source of Typhi (25), which would predict that the z66 gene was acquired later, possibly by horizontal gene transfer. We now know that z66 is present in only a single haplotype, and that the z66 flagellin gene \([fljB(z66)]\) and \(flIC\) repressor \([fljA(z66)]\) are located on a plasmid, indicating a relatively recent origin (1). The present study demonstrates that the acquisition of the pBSSB1 linear plasmid permitting the expression of z66 antigen most likely occurred on several independent occasions or that homologous recombination has taken place between different haplotypes. Furthermore, we suggest that pBSSB1 was horizontally transferred into an serovar Typhi strain carrying the \(d\) flagellin allele, and over time the gene became truncated in some strains to form the \(j\) flagellin epitope. Expression of the z66 antigen is the default in serovar Typhi strains harboring the \(fljB(z66)\) locus, and \(flIC\) is effectively silenced (2). The truncated \(j\) allele can be found in three haplotypes, i.e., H50, H59, and H85, suggesting either that deletion has occurred spontaneously on several independent occasions or that homologous recombination has taken place between different haplotypes. However, it seems that strains harboring \(fljB(z66)\) preferentially possess the \(j\) \(flIC\) chromosomal allele.

In conclusion, our results demonstrate the adaptation of a serovar Typhi genotyping scheme for a single country and ultimately a localized typhoid endemic region. The methodology offers a high level of sensitivity, thus allowing interrogation of phenotypes or pinpointing the locality of specific strains. The technique can in principle be used to define serovar Typhi circulating globally and is also potentially applicable to other bacterial pathogens. Our data prove that a number of distinct haplotypes of serovar Typhi can circulate in relatively small geographical areas. However, somewhat paradoxically, we found that H58 strains that are currently circulating in other parts of South Asia and are associated with multiple drug resistance were not detected in Indonesia. Furthermore, strains harboring the z66 antigen are associated with the Indonesian archipelago, there is no evidence for the spread of these organisms to other countries where typhoid is endemic. It is likely that understanding the “Indonesian exception” would aid our understanding of the global epidemiology of typhoid fever. By combining haplotyping with phenotyping we were able to gain considerable insight into the population structure of serovar Typhi circulating in this region. Surveillance of strains using these methods combined with assessment of social and medical practices over a prolonged period will add vital information on how serovar Typhi is evolving and spreading in the human population.

ACKNOWLEDGMENTS

We thank Soegianto Ali and Jonathan Hardy for supplying strains and information that have assisted with this study.

This study was funded by The Wellcome Trust, London, United Kingdom. K. E. S. was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada. M. A. is a Principal Investigator of the Scientific Foundation of Ireland (grant 05/FE1/B882) and the Max Planck Society for the Advancement of Science.

REFERENCES


