Translational Genomics to Develop a *Salmonella enterica* Serovar Paratyphi A Multiplex Polymerase Chain Reaction Assay

Hong-Yu Ou,* Cindy Teh Shuan Ju,[†]
Kwai-Lin Thong,[†] Norazah Ahmad,[‡] Zixin Deng,*
Michael R. Barer,^{§¶} and Kumar Rajakumar^{§¶}

From the Laboratory of Microbial Metabolism and School of Life Science and Biotechnology,* Shanghai Jiaotong University, Shanghai, China; the Institute of Biological Sciences,[†] University of Malaya, Kuala Lumpur, Malaysia; the Institute for Medical Research,[‡] Kuala Lumpur, Malaysia; the Department of Infection, Immunity, and Inflammation,[§] Leicester Medical School, University of Leicester, Leicester, United Kingdom; and the Department of Clinical Microbiology,[¶] University Hospitals of Leicester NHS Trust, Leicester, United Kingdom

The use of pathogen genome sequence data for the control and management of infections remains an ongoing challenge. We describe a broadly applicable, web-enabled approach that can be used to develop bacteria-specific polymerase chain reaction (PCR) assays. *Salmonella enterica* Paratyphi A has emerged as a major cause of enteric fever in Asia. Culture-based diagnosis is slow and frequently negative in patients with suspected typhoid and paratyphoid fever, potentially compromising patient management and public health. We used the MobilomeFINDER web-server to perform *in silico* subtractive hybridization, thus identifying 43 protein-coding sequences (CDSs) that were present in two Paratyphi A strains but not in other sequenced *Salmonella* genomes. After exclusion of 29 CDSs found to be variably present in Paratyphi A strains by microarray hybridization and grouping of remaining CDSs by genomic location, four dispersed targets (*stkF*, *spa2473*, *spa2539*, *hsdM*) were used to develop a highly discriminatory multiplex PCR assay. All 52 Paratyphi A strains within the diverse panel investigated produced one of two pathognomonic four-band signatures. Given rapid and ongoing expansion of DNA and comparative genomics databases, our universally accessible web-server-supported do-it-yourself approach offers the potential to contribute significantly to the rapid development of species-, serovar-, or pathotype-specific PCR assays targeting pre-existing and emerging bacterial pathogens. (J Mol Diagn 2007, 9:624–630; DOI: 10.2353/jmoldx.2007.070064)

Typhoid and paratyphoid fevers caused by *Salmonella enterica* serovars Typhi, Paratyphi A, Paratyphi B, and Paratyphi C continue to pose significant public health problems in many parts of the world, especially in developing countries.¹ Given similar presentations, these conditions cannot be reliably distinguished on clinical grounds alone.^{2,3} In addition, major increases in the incidence of Paratyphi A infection have been noted recently, particularly in China, India, and Pakistan.^{4,5} Remarkably, studies in some geographic localities have shown that up to 64% of patients with culture-proven enteric fever are infected with *S. Paratyphi A* rather than *S. Typhi* strains.^{4,6} From a clinical, epidemiological, and infection control perspective, it is essential to gather accurate data on the precise etiological agent of enteric fevers because present day vaccines are effective against *S. Typhi* only and antibiotic susceptibility patterns between serovars potentially differ markedly from one locality to the next. Ideally, these data would guide the formulation of empirical and definitive treatment regimes, thus helping to minimize antimicrobial abuse and preserve the utility of these agents. This information is also central to considerations regarding prioritization of Paratyphi A vaccine development and would be invaluable in any subsequent implementation of targeted and/or large-scale immunization programs directed at prevention of enteric fevers.

Early diagnosis and prompt treatment with appropriate antimicrobial agents are crucial to reducing the morbidity and mortality associated with enteric fever. However, conventional diagnosis that relies on isolation of the pathogen from blood, bone marrow aspirate, and/or stool cultures and subsequent biochemical and serological testing of suspect isolates requires a minimum of 4 to 8 days for final identification. Furthermore, the likelihood of a positive blood culture declines significantly with delayed presentation or prior

Supported by the Leicestershire Medical Research Foundation (*mediSearch* grant to K.R. and M.R.B.); the Ministry of Science and Technology, China (863 program grant 2006AA02Z328 and a Shanghai Rising-Star Program grant to H.Y.O.); and the University of Malaya (grant Vot F079/2005b to K.L.T.).

H.Y.O. and C.T.S.J. contributed equally to this study.

Accepted for publication August 23, 2007.

Supplementary material for this article can be found on <http://jmd.amjpathol.org>.

Address reprint requests to Dr. Kumar Rajakumar, Department of Infection, Immunity, and Inflammation, Leicester Medical School, University of Leicester, Maurice Shock Building, University Rd., P.O. Box 138, Leicester LE1 9HN, UK. E-mail: kr46@le.ac.uk.

antibiotic use. Significantly, blood cultures have been shown to be negative in 40 to 80% of cases of PCR-confirmed typhoid.⁷⁻⁹ However, there are no comparable data for Paratyphi A-associated infections. Prompt case detection and accurate epidemiological data are central to optimum public health control measures. Ominously, increasing incidence of multiantibiotic-resistant Typhi and Paratyphi A strains and/or those exhibiting reduced susceptibility to fluoroquinolones threaten to undermine the efficacy of empirical therapy regimens.^{1,10,11} Early identification of the specific etiological agent and knowledge of local resistance patterns would be invaluable in guiding antimicrobial choice. Because fluoroquinolones are only indicated for the treatment of infections caused by susceptible strains, real-time *gyrA* mutation testing that identifies resistance-associated polymorphisms within the fluoroquinolone-target gyrase gene may also be feasible in well-resourced settings.^{12,13}

We have used the MobilomeFINDER web server that we recently developed to exploit available *Salmonella* genome sequence information and microarray-derived comparative genomics data¹⁴ to identify potential serovar-specific genes as targets in a novel Paratyphi A multiplex PCR (mPCR) assay. Our simple step-by-step approach is readily accessible to the broader community of medical technologists and clinicians and would facilitate rapid translation of genomics data into pathogen-specific PCR assays that yield highly predictive and/or diagnostic signatures for a wide range of pathogens.

Materials and Methods

Databases

Complete genome sequences of *S. Paratyphi A* strain ATCC 9150 (GenBank accession no. CP000026)¹⁴ and four non-Paratyphi A strains, *S. Typhi* Ty2 (GenBank accession no. AE014613),¹⁵ *S. Typhi* CT18 (GenBank accession no. AL513382),¹⁶ *S. Typhimurium* LT2 (AE006468),¹⁷ and *S. Choleraesuis* SC-B67 (GenBank accession no. AE017220),¹⁸ were downloaded from GenBank. ATCC 9150 protein-coding sequence (CDS) annotation data were also extracted from GenBank. Publicly available contigs for 11 partially sequenced genomes were accessed online (Supplemental Table S1,¹⁴⁻¹⁸ available at <http://jmd.amjpathol.org>).¹⁹

Determination of *Salmonella Paratyphi A*-Specific CDS

CDSs that were present in the ATCC 9150 genome but that were not represented within the available *Salmonella* genome sequence data derived from non-Paratyphi A strains (Supplemental Table S1, see <http://jmd.amjpathol.org>) were selected using the GenomeSubtractor utility²⁰ found on our MobilomeFINDER²¹ web server: <http://mml.sjtu.edu.cn/MobilomeFINDER/>. GenomeSubtractor was used to perform a high-throughput BLASTN search, parse the resulting report, and generate a list of strain-specific CDSs. In brief,

each annotated CDS was used as a query in a similarity search against each non-Paratyphi A genome sequence (reference genome) using BLASTN²² and default NCBI BLASTN parameters, with the exception that F was set to F (no filter for repeated sequences). The stretch of sequence from the reference genome with the highest BLAST bit score for each query sequence was retrieved and a homology score (*H* value) calculated for each annotated CDS in turn. Bit scores are normalized values automatically produced by the BLAST algorithm that allow different alignments to be compared even when different scoring matrices have been used. In essence, the higher the bit score the more statistically significant the match (for details, see <http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=handbook.section.614>). The homology score we used had been proposed by Fukiya and colleagues²³ and reflected the degree of similarity between the matching reference genome sequence and the CDS itself in terms of the length of match and the degree of identity at a DNA level. For each query, the *H* value was calculated as $H = i \times l_m / l_q$, where *i* was the level of identity of the region with the highest bit score expressed as a frequency of between 0 and 1, *l_m* the length of the highest scoring matching sequence (including gaps), and *l_q* the query length. If the BLASTN *E* value of the best-hit was ≥ 0.01 , its *H* value was defined as zero.²³ A threshold value of *H₀* = 0.1 was chosen to select CDSs that had little or no similarity to sequences in the other genomes. All 4093 annotated ATCC 9150 CDSs were analyzed against the 4 completely sequenced and 10 partially sequenced non-Paratyphi A genome sequences in turn (Supplemental Table S1, see <http://jmd.amjpathol.org> for details of genome sequences). If all resulting *H* values for a given ATCC 9150 CDS were ≤ 0.1 , the CDSs were considered to be strain-specific with respect to the non-Paratyphi A genomes listed in Supplemental Table S1 (see <http://jmd.amjpathol.org>).

Oligonucleotide Primer Design

Oligonucleotide primers specific for each target ATCC 9150 CDS were selected using PrimerSelect (DNASTAR; Lasergene, Madison, WI). The program AutoDimer²⁴ was used to screen all four primer pairs for potential hairpin and primer-dimer interactions. Candidate primers were then checked by BLASTN against the available *Salmonella* genome sequences to minimize the likelihood of nonspecific amplification from nontarget loci. Selected primer pairs were then tested using *in silico* uniplex- and mPCR assays with the PCR amplification and multiplex PCR programs, respectively, available at <http://insilico.ehu.es/>.²⁵ If *in silico* mPCR resulted in amplicons of the expected size only, the primer pair/set was considered to be specific. Details of primers used in this study are listed in Table 1.

Bacterial Strains, Culture Media, and Serotyping

Fifty-two Paratyphi A strains from sporadic cases of enteric fever, 75 *S. enterica* subspecies *enterica* non-Paratyphi A strains, and 14 non-*Salmonella* bacterial strains were retrieved from glycerol stocks held at the University of Malaya (Supplemental Tables S2 to S4, see <http://>

Table 1. Oligonucleotides Used for the *Salmonella* Paratyphi A Multiplex PCR Assay

Target CDS	Forward primer sequence	Reverse primer sequence	Predicted product (bp)*
<i>spa0180</i>	5'-CCCGGTCTTGTGTTGCTTATG-3'	5'-CGTGACGTTGGGAAGGTAGA-3'	159
<i>spa2473</i>	5'-GGGTGAAAGCTCCGGTAATAAA-3'	5'-CTGACGGGACGAACCAAACTA-3'	324
<i>spa2539</i>	5'-TTATGGCCCAATATCGTCTCC-3'	5'-GAGTGTGCAAGTCAGCTATTTGTT-3'	484
<i>spa4289</i>	5'-ATGCCCGACACAGCCTAA-3'	5'-GAATCGTGCTGGATGACCTTTA-3'	627

*Predicted product lengths were obtained by using the *in silico* multiplex PCR program available at http://insilico.ehu.es/multiplex_PCR/²⁵ with the *Salmonella* Paratyphi A strain ATCC 9150 genome as a template.

jmd.amjpathol.org). The *Salmonella* bacteria used in this study had been identified to genus, species, and serovar level by phenotypic, biochemical, and serological testing at source institutions and reference centers. In the event of ambiguous or incomplete data, serovar identity was verified and/or determined at the *Salmonella* Reference Centre, IMR, using standard methods as previously described.²⁶ Strains were routinely cultured on Luria-Bertani medium.

Uniplex and Multiplex PCR Conditions and Amplicon Sequencing

PCR amplification was performed in a total reaction volume of 25 μ l containing 1 $\times$ Promega buffer (Promega, Madison, WI), 2.5 mmol/L MgCl₂, 200 μ mol/L of each dNTP, 2.5 U of TaqDNA polymerase (Promega), 0.4 μ mol/L of each primer and 5 μ l of DNA template. Genomic DNA template was prepared as follows. A single bacterial colony was picked, boiled in 50 μ l of water for 10 minutes, cooled on ice, and spun. Five μ l of the supernatant was used in the PCR assay. The PCR cycle comprised a 5-minute 95°C initial denaturation, followed by 30 cycles of 94°C for 30 seconds (denaturation), 57.6°C for 30 seconds (annealing), and 72.0°C for 1 minute (extension), and a final extension at 72.0°C for 10 minutes. Five- μ l reaction aliquots were resolved through a 1.5% agarose gel and DNA bands visualized after ethidium bromide staining.

Results

Selection of PCR Targets

All 4093 annotated CDSs within the *S. Paratyphi* A strain ATCC 9150 genome were downloaded from GenBank and screened for nucleotide sequence similarity against complete and partial *Salmonella enterica* subspecies *enterica* genome sequences (Figure 1). Data for the following serovars were available: Typhi, Typhimurium, Choleraesuis, Paratyphi B, Dublin, Enteritidis, Gallinarum, and Pullorum. The genomes of a *S. enterica* subspecies *S. diarizonae* and a *S. bongori* strain were also used in this comparative screen (Supplemental Table S1, see <http://jmd.amjpathol.org>). Forty-three ATCC 9150 CDSs had a maximum *H* value of ≤ 0.1 when compared with the non-Paratyphi A genomes. Because the *H* value = 0 indicated no significant match and *H* value = 1 reflected 100% DNA sequence identity throughout the full-length of the CDS (see Materials and Methods for details), these 43 CDSs were clearly unique to ATCC 9150. Furthermore, all 43 CDSs

exhibited near identity (*H* value >0.98) to corresponding orthologues in the genome of a second Paratyphi A strain (Sanger Centre). We grouped the 43 CDSs into 11 sets of immediately contiguous genes using the ArrayOme program²⁷ and further rationalized these into six clusters by merging sets of contiguous CDSs separated by less than 10 kb on the ATCC 9150 genome (Supplemental Table S5,^{14,20,22} see <http://jmd.amjpathol.org>).

Next, we examined the 43 ATCC 9150 CDSs against a previously reported comparative genomic hybridization dataset for ATCC 9150 and 12 other Paratyphi A strains that had been derived using a *Salmonella* multistrain (Typhimurium, Typhi, and Paratyphi A) nonredundant microarray (Figure 1; Supplemental Table S5, see <http://jmd.amjpathol.org>).¹⁴ Two clusters of variably present CDS, labeled as C5 and C6 in Supplemental Table S5 (<http://jmd.amjpathol.org>), were identified: a prophage, SPA-3-P2, that encoded 21 CDSs and resulted in free phage on induction¹⁴ and a smaller cluster containing eight CDSs that bore likely DNA replication-related genes. The remaining 14 CDSs were found to be present in all 13 Paratyphi A strains examined in the earlier study. These 14 CDSs mapped to four clusters of 7.0, 9.1, 0.6, and 1.8 kb in size that were located at genomic coordinates 0.2, 2.5, 2.6, and 4.4 Mb, respectively (clusters C1 to C4, Supplemental Table S5, see <http://jmd.amjpathol.org>). Primers were designed to amplify an empirically selected CDS representative of each cluster (*spa0180*, *spa2473*, *spa2539*, and *spa4289*). *spa0180* (*stkF*) and *spa4289* (*hsdM*) had previously been reported to code for a putative fimbrial protein and DNA methyltransferase, respectively;¹⁴ the remaining two CDSs coded for hypothetical proteins. Details of the primers used in this study are shown in Table 1.

Multiplex PCR Analysis of Strains

Multiplex PCR conditions were selected after temperature gradient experiments and validation with ATCC 9150 and seven other well-defined Paratyphi A strains. As expected all eight Paratyphi A strains produced four bands. However, the size of the smallest band, presumed to be specific to *spa0180*, varied among strains. ATCC 9150 and three other strains yielded the expected 159-bp product, whereas the remaining four Paratyphi A strains produced a fragment of ~ 180 bp instead (Figure 2B; lanes 3, 4, 7, and 8). This larger fragment was also amplified using uniplex-PCR targeting *spa0180* alone (data not shown).

We then tested this mPCR assay against 52 Paratyphi A and 75 *S. enterica* subspecies *enterica* non-Paratyphi A strains that had been previously characterized by API identification and *Salmonella* serotyping. The assay was defined

Figure 1. Schematic diagram highlighting the comparative genomics approach used to develop the *Salmonella* Paratyphi A-specific multiplex PCR assay. GenomeSubtractor identified 43 ATCC 9150 strain-specific CDS, 14 of which had been shown by microarray-CGH to be conserved in 12 other Paratyphi A strains.¹⁴ These 14 CDSs were mapped to the ATCC 9150 genes using ArrayOme²⁷ and rationalized into four clusters. Paratyphi A-specific CDS label colors denote the coding strand. The *in silico* multiplex PCR output was generated with primers designed to target four disparately located CDS: http://insilico.ebu.es/multiplex_PCR/.²⁵

as diagnostic of Paratyphi A when the three larger specific fragments and either the 159 bp or ~180 bp were seen. Using these criteria, the Paratyphi A mPCR assay correctly identified all 52 Paratyphi A strains and demonstrated a specificity of 100%. None of the non-Paratyphi A strains produced more than two bands, with the vast majority resulting in the amplification of either 0 or 1 band only. The 159-bp band, corresponding to the *spa0180* target, was the most frequently observed product, detectable in 30 of 75 non-Paratyphi A *Salmonella* strains tested. In the 11 instances in which two or more representatives of a non-Paratyphi A serovar were tested, results for all four targets were concordant across the serovar. None of the 14 non-

Salmonella bacteria tested produced detectable amplicons. Full details are available in Supplemental Tables S2 to S4 (<http://jmd.amjpathol.org>).

Paratyphi A-Associated *spa0180* Polymorphism

The *spa0180* CDS that codes for a 199-amino acid protein had been annotated as the putative fimbrial protein gene *stkF*.¹⁴ Because the mPCR assay resulted in an ~180-bp band instead of the expected 159 bp in 11 of 52 Paratyphi A strains, we partially sequenced the ~180-bp amplicon from two such Paratyphi A strains isolated at

Figure 2. Multiplex PCR assay targeting *spa0180*, *spa2473*, *spa2539*, and *spa4289* that produces amplicons of 159 or ~180 bp, 324 bp, 484 bp, and 627 bp, respectively. **A:** Multiplex and uniplex PCR analysis of *S. Paratyphi* A ATCC 9150: **lane 1**, 100-bp DNA ladder (Promega); **lane 2**, four-target multiplex PCR; **lanes 3 to 6** uniplex PCR targeting *spa0180*, *spa2473*, *spa2539*, and *spa4289*, respectively. **B:** *Salmonella* Paratyphi A strains: **lane 1**, 100-bp DNA ladder (Promega); **lane 2**, TKL034; **lane 3**, TKL008; **lane 4**, TKL009; **lane 5**, TKL010; **lane 6**, TKL011; **lane 7**, TKL012; and **lane 8**, TKL013. **C:** Selected non-Paratyphi A strains: **lane 1**, 100-bp DNA ladder (Promega); **lane 2**, Paratyphi A ATCC 9150 (control); **lane 3**, Albany TKL102; **lane 4**, Bareilly TKL107; **lane 5**, Chingola TKL113; **lane 6**, Kentucky TKL132; **lane 7**, Lomita TKL137; **lane 8**, Newport TKL141; **lane 9**, Okerara TKL142; **lane 10**, Paratyphi C ATCC 9068; **lane 11**, Tshion-gwe TKL155; **lane 12**, Virchow TKL172; **lane 13**, Waycross TKL174; **lane 14**, Weltevreden TKL175; and **lane 15**, Paratyphi B ATCC 8759.

distinct geographic localities (TKL008 and TKL025). The data confirmed these amplicons as *spa0180*-specific and demonstrated the presence of a 26-bp insertion toward the start of the structural gene that disrupted the reading frame and resulted in a premature stop at codon position 47 (Figure 3). The insertion was identical in the two strains and lay immediately adjacent to an identical 26-bp stretch of sequence in the native Paratyphi A ATCC 9150 *spa0180* gene, giving rise to a perfect tandem repeat within the mutant form (*spa0180v*) of this gene (Figure 3).

Discussion

An *in silico* subtractive hybridization strategy was used to identify CDSs that were unique to two Paratyphi A strains but not to the other available *Salmonella* genomes. A similar BLASTN-based subtraction approach has been used to

identify signature genes and/or nucleic acid sequences specific for particular bacteria.^{28,29} However, to the best of our knowledge this is only the third time that both *in silico* subtraction and microarray-based genome profiling have been used to develop a highly specific PCR assay for any bacterium.^{30,31} It is important to note that the streamlined, multistep, generic approach that we report is supported by the publicly available MobilomeFINDER web server²¹ that facilitates among other analyses PCR target identification, physical mapping of targets to the source genome, primer design and *in silico* PCR validation. An online tutorial that we aim to develop further in response to user comments is also provided via this site.

Our highly discriminatory four-target mPCR assay accurately identified all 52 Paratyphi A strains among 127 *S. enterica* strains (32 serovars) and 14 non-*Salmonella* bacteria with there being no false-positive identifications. An innovative feature of our assay is the targeting of multiple independent CDSs to detect and identify a single type of bacterium. In an era of genomics and bioinformatics, the use of multiple targets offers three major advantages. First, in the case of *Salmonella*, with more than 2000 distinct serovars, it would not be practical to experimentally validate an assay against the vast majority of serovars; hence, on purely probabilistic grounds, a test based on only one target would be much more likely to lead to misidentification than one based on a complex multiband signature. Indeed, when a larger panel of *Salmonella* bacteria were investigated, no single target was 100% specific for Paratyphi A (Supplemental Tables S2 and S3, see <http://jmd.amjpathol.org>), fully justifying this point. Second, given that

Figure 3. The *spa0180* polymorphism and the hypothesized model for low-frequency fimbrial phase switching. The *spa0180* DNA sequence is based on two sequenced Paratyphi A genomes, whereas the *spa0180v* sequence shown was derived by sequencing the ~180-bp *spa0180* amplicons generated using template DNA from Paratyphi A strains TKL008 and TKL025. The 10-bp closely spaced direct repeats are as indicated. The **light gray arrows** indicate the native 26-bp sequence in *spa0180* and the corresponding 26-bp tandem repeat in *spa0180v*. Predicted amino acid sequences for the regions of the genes represented are shown. Amino acid changes and the premature stop codon in *spa0180v* are shown in bold. The predicted numbers of amino acid residues for each gene are shown in parentheses.

diagnostic DNA sequences are selected to permit differentiation between closely related bacteria, it is likely that these constitute part of the potentially unstable accessory genome. The accessory genome comprises a strain's complement of horizontally acquired DNA as opposed to core DNA that is common to all members of the species. If so, a uniplex-PCR assay validated against a panel of contemporary strains may prove to be less robust when field tested throughout a decade or more, whereas an mPCR assay could be more resilient in the face of genome degradation and/or remodeling provided less stringent signatures were still accepted as proof of identity. Third, the use of multiple dispersed targets that were known on the basis of available genome data to be restricted to a single type of bacterium negates the need to assay a large panel of nontarget organisms to ascertain specificity as the probability of a pathognomonic multiband signature being produced by nontarget bacteria would approach zero.

The MobilomeFINDER-facilitated approach that we describe could also be used to identify and map multiple noncoding sequence targets. However, in this study we chose to target CDS specifically, and not noncoding sequences, because we hypothesized that coding genetic material that had previously been shown to be conserved across 13 Paratyphi A strains¹⁴ would be far more likely to be conserved than seemingly nonsense noncoding sequences that may have exhibited significant single nucleotide polymorphisms and/or other minor mutations across the broader population of strains.

Hirose and colleagues³² had previously used a traditional phenotype-gene association strategy to develop an mPCR assay that identified Typhi and Paratyphi A strains by selective amplification of five O, H, and Vi antigen genes. In their assay, selective amplification of the *O*-antigen-related *tyv* (*rfbE*) gene of Typhi strains, but not that of Paratyphi A strains, resulted from minor sequence polymorphism rather than the absence of this gene in strains of the latter serovar (for further details, see Supplemental Table S6, available at <http://jmd.amjpathol.org>).

In this study, using an entirely distinct approach to that of Hirose and colleagues,³² we developed an mPCR assay that produced one of two distinct but highly characteristic four-band signatures, each of which was diagnostic of *S. Paratyphi A*. Concordant mPCR results for all Paratyphi A strains tested supported their clonal nature, although the observed *spa0180* polymorphism suggested the possibility of two distinct lineages. However, given the closely spaced 10-bp direct repeats in *spa0180* (Figure 3), slipped-strand mispairing would seem to be a likely mechanism, suggesting independent mutation events instead.³³ Because this type of mutation is frequently reversible, an alternative hypothesis would be that the observed polymorphism reflected low-frequency switching between the two *spa0180* alleles (Figure 3). The *spa0180* (*stkF*) gene is part of a Paratyphi A-specific fimbrial gene cluster (*stkABCDEFG*) that exhibits sequence and organizational similarity to a poorly characterized cluster in several *Escherichia coli* and *Salmonella* strains¹⁴ (K.R. and H.-Y.O., unpublished data). StkA is presently annotated as a major fimbrial subunit protein,¹⁴ whereas the *E. coli* MG1655 orthologue of StkF (YadK) has been labeled as a predicted adhesin-like pro-

tein (GenBank accession no. AAC73247), although neither of these assignments has been experimentally verified. The observed 26-bp intragenic duplication event may be mediating a form of phase switching or immune evasion³⁴ by turning off a key fimbrial structural gene. In *E. coli* and several other *Enterobacteriaceae*, fimbrial phase switching is primarily governed by an invertible *fimA* promoter-bearing fragment.³⁵ However, this is not the case with *Salmonella* in which complex, multigene, transcriptional and translational mechanisms that vary from fimbrial operon to fimbrial operon seem to function.^{35–37} Significantly, there are no reports of programmed sequence duplication-based gene disruption in *Salmonella* or other *Enterobacteriaceae*. Further investigation of the Paratyphi A *stk* gene cluster and the proposed fimbrial switching model is warranted, particularly given the demonstrated role in virulence of several *Salmonella* fimbriae.^{38,39} The fortuitous inclusion of *stkF* in our assay highlighted a further potential dividend of a mPCR approach; had this target been excluded, a common *S. Paratyphi A* genetic polymorphism would have gone undetected. Given the observed wider distribution of *stkF* (Supplemental Table S3, see <http://jmd.amjpathol.org>), the phase variation model that we propose could equally apply to Stk appendages in many other *S. enterica* serovars.

In conclusion, using a bioinformatics-led translational genomics approach, we have fast-tracked the development of a Paratyphi A-specific mPCR assay that would permit same-day identification of suspect bacterial isolates. Our approach could also be used to develop broad, highly discriminatory mPCR tests targeting many distinct pathogens in a single assay. Diagnostic and/or prognostic assays that detect 20 or more targets could conceivably be tailored to take account of local epidemiology, identify emerging pathogens or noxious genes, and/or address particular disease syndromes. Luminex xMAP⁴⁰ or equivalent high-plex technologies promise to make such assays viable once current logistical challenges are addressed.^{13,41} We are currently developing a combined Typhi/Paratyphi A/pan-*Salmonella* mPCR assay for direct testing of blood, fecal, food, and environmental specimens to expedite detection of *Salmonella* and permit rapid clinically relevant serotype differentiation. Such an assay would significantly increase the number of laboratory-confirmed cases of nontyphoidal *Salmonella*, Typhi, and Paratyphi A infection, the last of which represents a major emergent disease in several parts of Asia. Given the rapid and ongoing expansion in DNA, genome sequence, and microarray-derived comparative genomics data, the approach we have taken offers the potential to radically transform the development of species-, serovar-, or pathotype-specific PCR assays targeting pre-existing and emerging pathogens.^{28,30,42}

Acknowledgments

We thank the cited genome sequencing centers for their public data access policy and acknowledge the use of unpublished genome sequence data (Supplemental Table S1, see <http://jmd.amjpathol.org>); and Koh Yin Tee at IMR for serotyping *Salmonella* strains.

References

- Bhan MK, Bahl R, Bhatnagar S: Typhoid and paratyphoid fever. *Lancet* 2005, 366:749–762
- Basnyat B: Typhoid and paratyphoid fever. *Lancet* 2005, 366:1603
- Vollaard AM, Ali S, Widjaja S, Asthen HA, Visser LG, Surjadi C, van Dissel JT: Identification of typhoid fever and paratyphoid fever cases at presentation in outpatient clinics in Jakarta. *Indonesia Trans R Soc Trop Med Hyg* 2005, 99:440–450
- Ochiai RL, Wang X, von Seidlein L, Yang J, Bhutta ZA, Bhattacharya SK, Agtini M, Deen JL, Wain J, Kim DR, Ali M, Acosta CJ, Jodar L, Clemens JD: *Salmonella* Paratyphi A rates. *Asia Emerg Infect Dis* 2005, 11:1764–1766
- Sood S, Kapil A, Dash N, Das BK, Goel V, Seth P: Paratyphoid fever in India: an emerging problem. *Emerg Infect Dis* 1999, 5:483–484
- Shlim DR, Schwartz E, Eaton M: Clinical importance of *Salmonella* Paratyphi A infection to enteric fever in Nepal. *J Travel Med* 1995, 2:165–168
- Kumar A, Arora V, Bashamboo A, Ali S: Detection of *Salmonella* typhi by polymerase chain reaction: implications in diagnosis of typhoid fever. *Infect Genet Evol* 2002, 2:107–110
- Massi MN, Shirakawa T, Gotoh A, Bishnu A, Hatta M, Kawabata M: Quantitative detection of *Salmonella enterica* serovar Typhi from blood of suspected typhoid fever patients by real-time PCR. *Int J Med Microbiol* 2005, 295:117–120
- Prakash P, Mishra OP, Singh AK, Gulati AK, Nath G: Evaluation of nested PCR in diagnosis of typhoid fever. *J Clin Microbiol* 2005, 43:431–432
- Adachi T, Sagara H, Hirose K, Watanabe H: Fluoroquinolone-resistant *Salmonella* Paratyphi A. *Emerg Infect Dis* 2005, 11:172–174
- Chandel DS, Chaudhry R, Dhawan B, Pandey A, Dey AB: Drug-resistant *Salmonella enterica* serotype Paratyphi A in India. *Emerg Infect Dis* 2000, 6:420–421
- Randall LP, Coldham NG, Woodward MJ: Detection of mutations in *Salmonella enterica* gyrA, gyrB, parC and parE genes by denaturing high performance liquid chromatography (DHPLC) using standard HPLC instrumentation. *J Antimicrob Chemother* 2005, 56:619–623
- Tenover FC: Rapid detection and identification of bacterial pathogens using novel molecular technologies: infection control and beyond. *Clin Infect Dis* 2007, 44:418–423
- McClelland M, Sanderson KE, Clifton SW, Latreille P, Porwollik S, Sabo A, Meyer R, Bieri T, Ozersky P, McLellan M, Harkins CR, Wang C, Nguyen C, Berghoff A, Elliott G, Kohlberg S, Strong C, Du F, Carter J, Kremizki C, Layman D, Leonard S, Sun H, Fulton L, Nash W, Miner T, Minx P, Delehaunty K, Fronick C, Magrini V, Nhan M, Warren W, Florea L, Spieth J, Wilson RK: Comparison of genome degradation in Paratyphi A and Typhi, human-restricted serovars of *Salmonella enterica* that cause typhoid. *Nat Genet* 2004, 36:1268–1274
- Deng W, Liou SR, Plunkett III G, Mayhew GF, Rose DJ, Burland V, Kodoyianni V, Schwartz DC, Blattner FR: Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2 and CT18. *J Bacteriol* 2003, 185:2330–2337
- Parkhill J, Dougan G, James KD, Thomson NR, Pickard D, Wain J, Churcher C, Mungall KL, Bentley SD, Holden MT, Sebahia M, Baker S, Basham D, Brooks K, Chillingworth T, Connerton P, Cronin A, Davis P, Davies RM, Dowd L, White N, Farrar J, Feltwell T, Hamlin N, Haque A, Hien TT, Holroyd S, Jagels K, Krogh A, Larsen TS, Leather S, Moule S, O'Gaora P, Parry C, Quail M, Rutherford K, Simmonds M, Skelton J, Stevens K, Whitehead S, Barrell BG: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* 2001, 413:848–852
- McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P, Courtney L, Porwollik S, Ali J, Dante M, Du F, Hou S, Layman D, Leonard S, Nguyen C, Scott K, Holmes A, Grewal N, Mulvaney E, Ryan E, Sun H, Florea L, Miller W, Stoneking T, Nhan M, Waterston R, Wilson RK: Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature* 2001, 413:852–856
- Chiu CH, Tang P, Chu C, Hu S, Bao Q, Yu J, Chou YY, Wang HS, Lee YS: The genome sequence of *Salmonella enterica* serovar Choleraesuis, a highly invasive and resistant zoonotic pathogen. *Nucleic Acids Res* 2005, 33:1690–1698
- Lioliou K, Tavernarakis N, Hugenholtz P, Kyrpides NC: The Genomes On Line Database (GOLD) v. 2: a monitor of genome projects worldwide. *Nucleic Acids Res* 2006, 34(Database issue):D332–D334
- Ou HY, Chen LL, Lonnen J, Chaudhuri RR, Thani AB, Smith R, Garton NJ, Hinton J, Pallen M, Barer MR, Rajakumar K: A novel strategy for the identification of genomic islands by comparative analysis of the contents and contexts of tRNA sites in closely related bacteria. *Nucleic Acids Res* 2006, 34:e3
- Ou HY, He X, Harrison EM, Kulasekara BR, Thani AB, Kadioglu A, Lory S, Hinton JC, Barer MR, Deng Z, Rajakumar K: MobilomeFINDER: web-based tools for in silico and experimental discovery of bacterial genomic islands. *Nucleic Acids Res* 2007, 35:W97–W104
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997, 25:3389–3402
- Fukui S, Mizoguchi H, Tobe T, Mori H: Extensive genomic diversity in pathogenic *Escherichia coli* and *Shigella* strains revealed by comparative genomic hybridization microarray. *J Bacteriol* 2004, 186:3911–3921
- Vallone PM, Butler JM: AutoDimer: a screening tool for primer-dimer and hairpin structures. *Biotechniques* 2004, 37:226–231
- Bikandi J, San Millan R, Rementeira A, Garaizar J: In silico analysis of complete bacterial genomes: PCR, AFLP-PCR and endonuclease restriction. *Bioinformatics* 2004, 20:798–799
- Popoff MY: *Antigenic Formulas of Salmonella Serovars*, ed 8. Paris, WHO Collaborating Centre for Reference and Research on *Salmonella*, Institute Pasteur; 2001
- Ou HY, Smith R, Lucchini S, Hinton J, Chaudhuri RR, Pallen M, Barer MR, Rajakumar K: ArrayOme: a program for estimating the sizes of microarray-visualized bacterial genomes. *Nucleic Acids Res* 2005, 33:e3
- Griffiths E, Ventresca MS, Gupta RS: BLAST screening of chlamydial genomes to identify signature proteins that are unique for the Chlamydiales. *Chlamydiaceae, Chlamydia and Chlamydia groups of species. BMC Genomics* 2006, 7:14
- Ratushna VG, Sturgill DM, Ramamoorthy S, Reichow SA, He Y, Lathigra R, Sriranganathan N, Halling SM, Boyle SM, Gibas CJ: Molecular targets for rapid identification of *Brucella* spp. *BMC Microbiol* 2006, 6:13
- Kim S, Frye JG, Hu J, Fedorka-Cray PJ, Gautam R, Boyle DS: Multiplex PCR-based method for identification of common clinical serotypes of *Salmonella enterica* subsp. *Enterica*. *J Clin Microbiol* 2006, 44:3608–3615
- Semret M, Alexander DC, Turenne CY, Haas de P, Overduin P, Soolingen van D, Cousins D, Behr MA: Genomic Polymorphisms for *Mycobacterium avium* subsp. *Paratuberculosis* diagnosis. *J Clin Microbiol* 2005, 43:3704–3712
- Hirose K, Itoh K, Nakajima H, Kurazono T, Yamaguchi M, Moriya K, Ezaki T, Kawamura Y, Tamura K, Watanabe H: Selective amplification of *tyv* (*rfaE*), *prt* (*rfaB*), *viaB*, and *fliC* genes by multiplex PCR for identification of *Salmonella enterica* serovars Typhi and Paratyphi A. *J Clin Microbiol* 2002, 40:633–636
- Torres-Cruz J, van der Woude MW: Slipped-strand mispairing can function as a phase variation mechanism in *Escherichia coli*. *J Bacteriol* 2003, 185:6990–6994
- Norris TL, Baumler AJ: Phase variation of the *lpl* operon is a mechanism to evade cross-immunity between *Salmonella* serotypes. *Proc Natl Acad Sci USA* 1999, 96:13393–13398
- Nicholson B, Low D: DNA methylation-dependent regulation of *Pef* expression in *Salmonella typhimurium*. *Mol Microbiol* 2000, 35:728–742
- Clegg S, Hancox LS, Yeh KS: *Salmonella typhimurium* fimbrial phase variation and *FimA* expression. *J Bacteriol* 1996, 178:542–545
- Tinker JK, Clegg S: Control of *FimY* translation and type 1 fimbrial production by the arginine tRNA encoded by *fimU* in *Salmonella enterica* serovar *Typhimurium*. *Mol Microbiol* 2001, 40:757–768
- Althouse C, Patterson S, Fedorka-Cray P, Isaacson RE: Type 1 fimbriae of *Salmonella enterica* serovar *Typhimurium* bind to enterocytes and contribute to colonization of swine in vivo. *Infect Immun* 2003, 71:6446–6452
- Edwards RA, Matlock BC, Heffernan BJ, Maloy SR: Genomic analysis and growth-phase-dependent regulation of the SEF14 fimbriae of *Salmonella enterica* serovar *Enteritidis*. *Microbiology* 2001, 147:2705–2715
- Brunstein J, Thomas E: Direct screening of clinical specimens for multiple respiratory pathogens using the Genaco respiratory panels 1 and 2. *Diagn Mol Pathol* 2006, 15:169–173
- Baron EJ: Implications of new technology for infectious diseases practice. *Clin Infect Dis* 2006, 43:1318–1323
- Panagea S, Winstanley C, Parsons YN, Walshaw MJ, Ledson MJ, Hart CA: PCR-based detection of a cystic fibrosis epidemic strain of *Pseudomonas aeruginosa*. *Mol Diagn* 2003, 7:195–200