Combined genomarkers approach to *Salmonella* characterization reveals that nucleotide sequence differences in the phase 1 flagellin gene *fliC* are markers for variation within serotypes

Chloe Bishop,1 Christiane Honisch,2Thomas Goldman,2Mikeal Mosko,2 Sereyrathana Keng,2 Catherine Arnold1 and Saheer Gharbia1

1Centre for Infections, Health Protection Agency, Colindale, London, UK
2Sequenom Inc., San Diego, California, USA

INTRODUCTION

Salmonellosis is one of the most common and widely distributed diseases of humans and animals and constitutes a major public health burden. There are millions of human cases worldwide every year, resulting in thousands of deaths (WHO, 2005). *Salmonella* serotyping identifies variation of flagellar and somatic antigens by reaction to specific antisera to separate isolates into over 2600 serotypes. Most serotypes exhibit diphagic flagellar antigen expression by alternately expressing two genes, *fliC* (phase 1) and *fljB* (phase 2), which encode flagellins of different antigenicity. *Salmonella* is a very clonal genus and it is often necessary to type beyond the serotype level to distinguish between outbreak strains (Liebana et al., 2001), requiring additional laboratory methods such as PFGE (Murase et al., 1995; Powell et al., 1994), phage typing (Ward et al., 1987) plasmid profiling (Threlfall et al., 1989) and ribotyping (Liebana et al., 2001).

Characterization of isolates that relies on gene expression can be affected by storage and growth conditions. Consequently, these tests can vary within and between laboratories. There is a clear demand for clinically relevant schemes that are capable of typing and subtyping, ideally with a single technological approach. DNA sequence-based typing methods offer faster, more portable results that allow a more informative analysis than phenotypic tests; none has yet emerged clearly as the ideal procedure (Wattiau et al., 2011).

Although the majority of *Salmonella* flagellin sequences are now known, to date there are no molecular approaches that can resolve all of the serotypes. Typically, DNA-based schemes can resolve only the major disease-causing serotypes (McQuiston et al., 2004a; Mortimer et al., 2004; Sonne-Hansen & Jenabian, 2005) and the power of these sequences to discriminate beyond the serotype level is not known. Four O-antigens have been typed by length heterogeneity PCR (Luk et al., 1993). Since then, PCR assays targeting individual serotypes have been developed (Fitzgerald et al., 2003, 2006). PCR combined with probe-based identification of amplicons has been used to detect six O-groups (Fitzgerald et al., 2007).
Other approaches have assayed genomic differences to deduce serotypes, for example fluorescent amplified fragment length polymorphism (Scott et al., 2001) and multi-locus enzyme electrophoresis (Boyd et al., 1997). These methods provide a view of genetic diversity between strains and partially group some serotypes but on the whole do not group or identify serotypes. Multi-locus sequence typing has been used to discriminate between Salmonella strains by sampling variation in a set of housekeeping genes (Kotetishvili et al., 2002). Sequencing schemes not restricted to housekeeping genes have been used for serotype prediction for selected serotypes (Morales et al., 2006; Sukhnanand et al., 2005), but it is not known whether these schemes would be applicable for a broader range of serotypes. A few previous works considered a broader range of serotypes, such as the development of a probe hybridization-based assay, which targeted 10–45 described polymorphisms (Lauri et al., 2011), although the assay’s ability to characterize unknown isolates was not tested. Selection of informative gene targets for typing and subtyping of Salmonella is critical; the acquisition of virulence factors and antibiotic resistance genes leads to the emergence of new strains and outbreaks.

In this study, DNA sequence variations in 14 selected genes were detected using base-specific cleavage of target sequences and matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry (MALDI-TOF MS) analysis, as described previously (Bayliss et al., 2010; Stanssens et al., 2004). The level of speed, automation and standardized scalable sample processing possible using this technology (Honisch et al., 2007) could enable efficient outbreak monitoring. We investigated whether this approach could meet the need for a rapid method that provides clinically relevant characterization of isolates and whether it could reveal the relationship between strains and serotypes.

METHODS

Bacterial strains. In this study, 224 Salmonella strains were characterized (Table S1, available in JMM Online), comprising 85 reference strains from the Salmonella Genetic Stock Centre (University of Calgary, Canada) and panels of serotyped and antibiotic resistance tested strains gratefully received from the Salmonella Reference Unit, Health Protection Agency, UK. The collection comprised 51 serotypes, including multiple representatives of some medically common serotypes, plus 18 non-typable isolates. DNA was extracted from cells using a MagNA Pure instrument and a Total Nucleic Acid Extraction kit 1 (Roche) according to the manufacturer’s instructions.

The 14 gene targets selected for comparative sequence analysis were: phase 1 flagellin (flfC), flagellar cap protein (flfD), flagellar hook-associated protein (flgK), fimbrin (fimA), virulence-associated secretory protein (spaM), antigen presentation protein (spaN), phosphomannomutase structural protein (manB), GDP-mannose synthesis (manC), glucose-1-phosphate thymidylyltransferase (rmlA), isocitrate dehydrogenase (icdA), glyceraldehyde-3-phosphate dehydrogenase (gapA), glutamine synthetase (glnA), 6-phosphogluconate dehydrogenase (gnd) and DNA gyrase subunit A (gyrA). The sequence of the gnd gene was targeted with two amplicons because of its length. Primer design was supported by sequence conservation in the central region of the gene.

Primer design and PCR. Salmonella sequences in GenBank, comprising full genome sequences, full gene sequences and partial gene sequences, were used to identify regions of interest and design target-specific PCR primers (Table S2). The majority of primers were synthesized twice by Eurofins (Germany), once tagged with a T7 promoter and once with a 10-mer tag (Table S2). Two 5 µl PCRs were performed per amplicon. Using tagged primers, one reaction incorporated a T7 promoter into the forward strand and a 10-mer tag into the reverse strand; the second reaction incorporated a 10-mer tag into the forward strand and a T7 promoter into the reverse strand. In the course of development, some primers were synthesized with a T7 tag at the 5’ end of the forward primer and an SP6 tag at the 5’ end of the reverse primer. This streamlined the production to one 10 µl PCR per amplicon.

Five nanograms of DNA was amplified using HotStarTaq DNA polymerase (Qiagen). PCRs contained 1 x PCR buffer (containing 1.5 mM MgCl2; pH 8.7; Qiagen), 200 µM each dNTP, 0.1 U HotStar Taq polymerase and 1 pmol each primer. PCRs were performed in a 384-well plate format. Cycling parameters were one cycle of 94 °C for 15 min, followed by 45 cycles of 94 °C for 20 s, 62 °C for 30 s and 72 °C for 1 min. A final extension step at 72 °C for 3 min completed the programme.

Comparative sequence analysis. Comparative sequence analysis of the amplified target regions was performed using a base-specific cleavage approach (MassCLEAVE; Sequenom) followed by MALDI-TOF MS, as described previously (Honisch et al., 2007). Briefly, 0.04 U shrimp alkaline phosphatase was added per microlitre of PCR product and incubated for 20 min at 37 °C with the PCRs to dephosphorylate unincorporated dNTPs, followed by a 5 min denaturation at 85 °C. PCR/shrimp alkaline phosphatase reactions were split into four reaction wells at 2 μl each. Each well was subjected to a different in vitro transcription/cleavage reaction by adding 2.5 μl transcription/cleavage cocktail (Sequenom) and incubating for 3 h at 37 °C. This cleaved the products of the forward and reverse strands at C or U. Reactions were conditioned by dilution with 22 μl water and mixing with 6 mg Clean Resin (Sequenom) for 10 min. Reactions were dispensed onto 384-element SpectroCHIPs using a nanodispenser (Sequenom). Nucleotide fragments in the four base-specific cleavage reactions were separated and measured on a MassARRAY Compact Analyser MALDI-TOF mass spectrophotometer (Sequenom). Sample sequences were deduced using iSEQ software (Sequenom) to match sample and reference sequences and to identify peak variations and resulting new sequences. Standard Sanger dyelexy sequencing was performed for concordance testing and to generate reference sequences of highly variable target regions.

Data analysis. Sample sequences were identified based on identification of the best matching reference sequences and interpretation of peak variations using the iSEQ software. Cluster analysis was performed using BioNumerics (Applied Maths) and a UPGMA algorithm to reveal the relatedness of sequences.

RESULTS

MALDI-TOF MS data were obtained for 14 loci from 224 strains and 2935 sequences (summarized in Fig. 1). All loci were polymorphic, ranging from 14 unique sequence types for gyrA up to 66 for flfC. The success rate for amplification ranged from 85 % for flfD to 94 % for flfC.
Clustering and sequence variation

Cluster analysis identified relationships between the sequences. The combined length of branch pairs represents the distance between pairs of sequences. Diversity in the sequences of each locus was calculated using Simpson’s index of diversity, which produced a single numerical value to compare loci or combined loci (Table 1).

Analysis of serotype-derived sequences

*fliC* amplicons of 1147 bp in length were shown to have the highest discriminating power of all individual loci. Sequences were clustered by the antigen they encoded and sequence diversity was detected in genetic determinants of single antigens (Fig. 2), supporting findings of previous studies (McQuiston et al., 2004b; Mortimer et al., 2004). Eleven flagellin types (b; c; d; e,h; i; k; l,v; r; z; z16; and z4,z23) occurred in more than one serotype in the study. Sequences differed between serotypes in all but four cases: *fliC-e,h* was homologous in *Salmonella enterica* subsp. *enterica* Larochelle (*S. Larochelle*) and *S. Saintpaul*, as well as in *S. Anatum* and *S. Newington*; *fliC-d* was homologous in *S. Muenchen* and *S. Stanley*. In these three cases, strains differed by at least two other loci, including *glnA* and *manC*. The fourth case of a *fliC* sequence shared between serotypes was *S. Enteritidis* and non-flagellate *S. Gallinarum*, serotypes known to be closely related; however, the sequences of *flgK* differed between these two serotypes.

There were 11 serotypes that demonstrated variability in their *fliC* sequence: *S. Derby*, *S. Dublin*, *S. Enteritidis*, *S. Heidelberg*, *S. Muenchen*, *S. Newport*, *S. Panama*, *S. Saintpaul*, *S. Senftenberg*, *S. Stanley* and *S. Typhimurium*. For six of these serotypes, a correlation was observed between the *fliC* sequences and variations at other loci.

Table 1. Simpson’s index of diversity for individual amplicons and amplicon combinations

Simpson’s index of diversity shows the abilities of single or combined amplicons to discriminate between isolates. The amplicon size shown relates to the longest sequence in the reference database.

<table>
<thead>
<tr>
<th>Amplicon (combination)</th>
<th>Simpson’s index of diversity</th>
<th>Amplicon size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>fliC</em></td>
<td>0.960</td>
<td>1147</td>
</tr>
<tr>
<td><em>fliD</em></td>
<td>0.890</td>
<td>625</td>
</tr>
<tr>
<td><em>flgK</em></td>
<td>0.760</td>
<td>551</td>
</tr>
<tr>
<td><em>manB</em></td>
<td>0.932</td>
<td>415</td>
</tr>
<tr>
<td><em>rmlA</em></td>
<td>0.909</td>
<td>282</td>
</tr>
<tr>
<td><em>gyrA</em></td>
<td>0.803</td>
<td>458</td>
</tr>
<tr>
<td><em>fliC-manB</em></td>
<td>0.713</td>
<td>300</td>
</tr>
<tr>
<td><em>fliC-fliD-flgK</em></td>
<td>0.969</td>
<td>1562</td>
</tr>
<tr>
<td><em>icdA-glnA-gnd</em></td>
<td>0.969</td>
<td>2323</td>
</tr>
<tr>
<td><em>fliC-fliD-flgK-manB-manC-rmlA</em></td>
<td>0.969</td>
<td>2643</td>
</tr>
<tr>
<td><em>fmlA-spaM-spaN</em></td>
<td>0.965</td>
<td>3505</td>
</tr>
<tr>
<td><em>manB-manC-rmlA</em></td>
<td>0.961</td>
<td>1208</td>
</tr>
<tr>
<td>z4,z23</td>
<td></td>
<td>1182</td>
</tr>
</tbody>
</table>
Fig. 2. A phylogenetic tree of 200 isolates based on the fliC amplicon sequences, constructed using the UPGMA method. Bootstrap values are shown. The serotypes are colour coded by the antigen encoded by fliC. Branches are labelled by serotype, the number of representatives and the antigen encoded by fliC.

C. Bishop and others

S. Butantan (n=1) b
S. Ohio (n=2) b
S. Hull (n=1) b
S. Java (n=3) b
S. Muenchen (n=11) d
S. Stanley (n=10) d
S. Stanley (n=2) d
S. Schwarzengrund (n=2) d
S. Champaign (n=1) k
S. Hadar (n=1) z10
S. Blockley (n=1) k
S. Ilb_50:k:z (n=1) k
S. Indiana (n=2) z
S. Worthington (n=1) z
S. Give (n=1) [d],I,v
S. Panama (n=1) I,v
S. Brandenburg (n=1) I,v
S. Bredeney (n=1) I,v
Unnamed (n=2) e
S. Salford (n=1) I,v
Unnamed (n=3) I,v
S. Panama (n=2) I,v
S. Miami (n=1) a
S. Hadar (n=12) z10
S. Haifa (n=1) z10
S. Mbandaka (n=1) z10
S. Bareilly (n=2) y
S. Gold Coast (n=1) r
S. Rubislaw (n=1) r
S. Virchow (n=11) r
S. Heidelberg (n=11) r
S. Infantis (n=9) r
S. Typhimurium (n=1) i
S. Typhimurium (n=5) i
S. Typhimurium (n=3) i
S. Typhimurium (n=24) i
S. Kentucky (n=4) i
S. Choleraesuis (n=1) [c]
S. Decatur (n=1) c
S. Anatum (n=1) e,h
S. Newton (n=1) e,h
S. Saintpaul (n=2) e,h
S. Newport (n=1) e,h
S. Newport (n=2) e,h
S. Larochelle (n=1) e,h
S. Saintpaul (n=8) e,h
S. Newport (n=13) e,h
S. Reading (n=1) e,h
S. Saintpaul (n=1) e,h
S. Enteritidis (n=1) g,m
S. Enteritidis (n=3) g,m
S. Gallinarum (n=1) -
S. Dublin (n=1) g,p
S. Duisbergar (n=1) d
S. Emek (n=1) g,m,s
S. Montevideo (n=3) g,m,[p],s
S. Derby (n=1) f,g
Unnamed (n=1)
S. Senftenberg (n=2) g,[s],t
S. Senftenberg (n=6) g,[s],t
S. Senftenberg (n=4) g,[s],t
S. Oranienberg (n=1) m,t
S. VII_1,40:g,z_{s1}:z_{s1} (n=1) g_{z_{s1}}
S. Illa_62_{z_{s1},z_{s23}}:z_{s1} (n=1) z_{s1},z_{s23}
S. Stanleyville (n=1) z_{s1},z_{s23}
Where more than two representative strains were available for testing, the correlation between \textit{fliC} sequence variation and variation at the other loci was evident. For example, strains of \textit{S. Newport} (\(n=15\)) had two \textit{fliC} sequence types (3 and 34). The division of \textit{S. Newport} strains into two groups was supported by differences in eight other loci, with similar patterns for \textit{S. Panama}, \textit{S. Saintpaul} and \textit{S. Senftenberg}, \textit{S. Heidelberg} and \textit{S. Muenchen} were variable at \textit{fliC} and other loci but no clear correlation with \textit{fliC} sequence was observed for either serotype.

The gene combination \textit{fliC} and \textit{manB} gave the highest Simpson’s index value of any individual gene region or combination of gene regions tested and separated isolates into the highest number of groups. The \textit{fliC-manB} combination largely clustered strains by serotype. However, it was the gene combination \textit{fliC-flgK-manC} that separated strains into 105 profiles and, crucially, separated all serotypes.

The \textit{fliD} gene encodes the flagellar cap protein. Expressed at the cell surface, \textit{fliD} could reasonably be expected to vary between strains but sequences were homologous within serotypes \textit{S. Hadar}, \textit{S. Heidelberg}, \textit{S. Infantis}, \textit{S. Java}, \textit{S. Ohio}, \textit{S. Stanley}, \textit{S. Typhimurium} and \textit{S. Virchow}. Other serotypes that showed variation in their \textit{fliD} sequence are largely polyphyletic serotypes \textit{S. Muenchen}, \textit{S. Newport}, \textit{S. Saintpaul}, \textit{S. Miami} and \textit{S. Choleraesuis}. The \textit{fliD} amplicon did show evidence of serotype-specific sequences for a subset of strains. For example, all \textit{S. Typhimurium} strains exhibited \textit{fliD} sequence type 1 but shared this sequence with four other serotypes.

The \textit{flgK} gene encodes the flagellar hook assembly protein and interacts with the subunit proteins encoded by \textit{fliC}. The \textit{flgK} amplicon was the most conserved among strains. More than one-third of strains shared one sequence type and no correlation was observed between serotype and \textit{flgK} sequence. Interestingly, this amplicon did vary between the related strains \textit{S. Enteritidis} and \textit{S. Gallinarum}.

Clustering by concatenated O-antigen gene sequences (\textit{manB-manC-rmlA}) generally grouped strains by serotype. Sequences fell into seven clusters, each containing two or more serotypes and between one and five O-groups, but serotypes clustered at branch ends. Among all tested amplicons associated with O-antigen biosynthesis in this study, \textit{manB} had the highest discriminating power as reflected in the highest Simpson’s index value.

\textbf{Analysis of metabolic pathway representative genes}

This group of genes (\textit{gapA}, \textit{icdA} and \textit{gnd}) was included as potential markers for overall genetic relatedness of strains. Individual housekeeping genes and concatenated sequences broadly clustered strains by serotype.

\textbf{Analysis of virulence-associated sequences}

Three genes involved in virulence were analysed, \textit{fimA}, \textit{spaM} and \textit{spaN}. The \textit{fimA} gene encodes a surface-expressed protein, the fimbrial subunit. Of the three virulence-associated gene regions, \textit{fimA} was the most variable. Genes \textit{spaM} and \textit{spaN} encode proteins involved in cell invasion. Strains of \textit{S. Dublin} and \textit{S. Choleraesuis} are more likely to cause systemic infection than other non-typhoidal serotypes (Blaser & Feldman, 1981). Except for the \textit{spaN} amplicon from \textit{S. Choleraesuis} strain B06, strains of \textit{S. Dublin} and \textit{S. Choleraesuis} shared \textit{spaM} and \textit{spaN} sequence similarity with other strains; nothing in the clustering of these two serotypes suggested that \textit{spa} gene sequences could be used as virulence markers. The amplicon generated from \textit{spaN} was found to be more variable than that of \textit{spaM}, supporting the findings of a previous study (Boyd \textit{et al.}, 1997).

\textbf{Analysis of genes associated with antibiotic resistance}

Mutations in the \textit{gyrA} gene that confer amino acid changes in DNA gyrase subunit A can cause resistance to quinolones (Piddock, 2002). In this study, \textit{gyrA} had the lowest value by Simpson’s index of diversity, as only 99 of the 224 strains tested were isolates associated with outbreaks of antibiotic resistant strains. The majority of the strains had \textit{gyrA} sequence variation levels in line with housekeeping genes. Nucleotide polymorphisms in \textit{gyrA} were found not to correlate with serotype; variation related only to antibiotic resistance. Strains formed nine clusters; strains in clusters 2–9 had various antibiotic resistance profiles but all were known to be quinolone resistant, with the exception of seven strains whose antibiotic resistance profile was not determined. The strains in cluster 1 were wild-type strains. Four strains in cluster one were found to be quinolone resistant and may have mutations in genes other than \textit{gyrA} that have been linked to quinolone resistance (Breines \textit{et al.}, 1997; Khodursky \textit{et al.}, 1995).

\textbf{DISCUSSION}

This approach to \textit{Salmonella} typing meets the need for a rapid method for clinically relevant characterization of isolates. The \textit{fliC} sequence was found to be a marker of genetic variation within serotypes in addition to between serotypes. To our knowledge, this is the first time that this has been reported. We found that, for \textit{S. Newport}, \textit{S. Panama}, \textit{S. Saintpaul} and \textit{S. Senftenberg}, and possibly \textit{S. Dublin} and \textit{S. Miami}, \textit{fliC} sequence variation reflected genomic diversity within these serotypes. Apart from \textit{S. Senftenberg}, these serotypes have been acknowledged previously as polyphyletic. This study has shown for the first time that \textit{S. Senftenberg} may also be polyphyletic.

The data arising from this set of strains suggested that, in most cases, there is enough variability within \textit{fliC} to separate serotypes with the same phase 1 flagellin antigen. Furthermore, the aforementioned correlation with variation at other loci indicated that \textit{fliC} variation reflects genuine subtypes within serotypes. The \textit{fliC} amplicon is a good marker for these purposes.
predictor of serotype for this set of strains. For the vast majority of strains in the study, the \(fliC\) sequence alone was sufficient for the prediction of serotypes and with the addition of only \(flgK\) and \(manC\), all serotypes could be separated. Markers of the O-antigen added confidence to the serotype prediction. The results of this study suggest that the three-gene scheme would be an efficient and valuable approach for \textit{Salmonella} typing by reference laboratories.

Limitations to the application of the technology were that some targets did not amplify for all of the isolates, possibly due to sequence differences within the primer-binding regions, whilst the comparative success of \(fliC\) amplification can be attributed to the comprehensive dataset available. Other sequences could not be resolved by the comparative sequencing algorithm due to lack of reference sequences in the database, in which case the corresponding samples were subjected to dideoxy sequencing. Resulting new sequences were added to the reference database, supporting subsequent analyses. Duplication of the \(manB\) gene in some S. Montevideo isolates has been hypothesized.

### Table

<table>
<thead>
<tr>
<th>Serotype</th>
<th>Strain</th>
<th>Gene target</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. Dublin</td>
<td>B14</td>
<td>41 ND 1 ND 1 ND 1 ND 1 ND 1 ND 1 ND 1 ND 1 ND 1 ND 1 ND 1 ND 1 ND 1 ND 1 ND</td>
</tr>
<tr>
<td>S. Dublin</td>
<td>B12</td>
<td>42 9 3 17 26 17 1 ND 2 2 1 14 7 6 3</td>
</tr>
<tr>
<td>S. Miami</td>
<td>B28</td>
<td>36 34 2 23 3 18 1 9 23 11 31 11 1 27 31</td>
</tr>
<tr>
<td>S. Miami</td>
<td>B29</td>
<td>48 23 27 44 4 3 1 11 8 1 11 4 22 5 16</td>
</tr>
<tr>
<td>S. Newport</td>
<td>B36</td>
<td>34 17 5 20 32 2 1 2 14 1 22 7 9 2 14</td>
</tr>
<tr>
<td>S. Newport</td>
<td>B38</td>
<td>34 17 5 45 32 2 1 2 14 1 22 7 9 2 14</td>
</tr>
<tr>
<td>S. Newport (n=2)</td>
<td>03_01 and 03_10</td>
<td>3 7 1 2 2 2 2 6 7 3 2 1 4 2 6</td>
</tr>
<tr>
<td>S. Newport</td>
<td>02_32</td>
<td>3 7 2 2 2 2 3 7 3 2 1 3 2 6</td>
</tr>
<tr>
<td>S. Newport</td>
<td>03_11</td>
<td>3 7 2 2 2 2 4 6 2 1 2 1 4 2 ND</td>
</tr>
<tr>
<td>S. Newport</td>
<td>03_14</td>
<td>3 7 1 2 2 2 4 6 2 1 2 1 4 2 ND</td>
</tr>
<tr>
<td>S. Newport</td>
<td>03_17</td>
<td>3 7 2 2 2 2 4 7 2 1 2 1 4 2 6</td>
</tr>
<tr>
<td>S. Newport</td>
<td>03_23</td>
<td>3 7 3 2 2 2 4 2 1 2 1 4 2 ND</td>
</tr>
<tr>
<td>S. Newport</td>
<td>03_24</td>
<td>3 4 2 2 2 2 6 7 3 2 1 1 4 2 6</td>
</tr>
<tr>
<td>S. Newport</td>
<td>02_38</td>
<td>3 7 1 2 2 2 4 7 2 1 2 1 4 2 ND</td>
</tr>
<tr>
<td>S. Newport</td>
<td>B37</td>
<td>3 8 5 12 2 19 1 13 14 5 3 19 8 2 17</td>
</tr>
<tr>
<td>S. Newport</td>
<td>B38</td>
<td>3 8 5 2 2 2 1 9 7 3 32 17 1 2 2</td>
</tr>
<tr>
<td>S. Newport</td>
<td>BS06</td>
<td>3 8 5 ND 1 2 1 13 14 5 3 3 4 2 17</td>
</tr>
<tr>
<td>S. Newport</td>
<td>BS47</td>
<td>3 8 2 12 2 2 1 9 29 5 3 19 6 2 ND</td>
</tr>
<tr>
<td>S. Panama</td>
<td>BS31</td>
<td>55 14 1 18 4 3 1 11 11 1 11 4 18 5 18</td>
</tr>
<tr>
<td>S. Panama</td>
<td>B39</td>
<td>65 14 1 18 4 3 1 11 11 1 11 4 18 5 18</td>
</tr>
<tr>
<td>S. Panama (n=2)</td>
<td>B40 and B41</td>
<td>55 14 1 18 4 3 1 11 11 1 11 4 18 5 18</td>
</tr>
<tr>
<td>S. Saintpaul (n=3)</td>
<td>A25, A26 and A38</td>
<td>7 6 4 6 11 1 1 47 1 2 1 3 2 1 1</td>
</tr>
<tr>
<td>S. Saintpaul</td>
<td>A22</td>
<td>38 28 1 8 3 1 1 47 1 2 1 11 2 1 1</td>
</tr>
<tr>
<td>S. Saintpaul</td>
<td>A23</td>
<td>7 6 10 6 11 1 1 47 1 2 1 3 2 1 1</td>
</tr>
<tr>
<td>S. Saintpaul</td>
<td>A24</td>
<td>7 6 1 6 11 1 1 47 1 2 1 1 2 1 1</td>
</tr>
<tr>
<td>S. Saintpaul</td>
<td>A28</td>
<td>7 6 4 6 ND 1 1 47 1 2 1 3 2 1 1</td>
</tr>
<tr>
<td>S. Saintpaul</td>
<td>B55</td>
<td>7 6 1 6 11 1 1 47 1 2 1 3 2 1 1</td>
</tr>
<tr>
<td>S. Saintpaul</td>
<td>BS05</td>
<td>7 6 10 6 8 1 1 47 1 2 1 18 11 1 1</td>
</tr>
<tr>
<td>S. Saintpaul</td>
<td>A29</td>
<td>18 4 3 25 9 9 1 16 28 1 14 10 15 3 22</td>
</tr>
<tr>
<td>S. Saintpaul</td>
<td>B56</td>
<td>18 4 3 25 9 9 1 16 28 1 14 10 11 3 22</td>
</tr>
<tr>
<td>S. Senftenberg (n=2)</td>
<td>02_21 and 02_25</td>
<td>10 ND 2 3 4 5 8 5 5 1 7 1 24 4 5</td>
</tr>
<tr>
<td>S. Senftenberg (n=2)</td>
<td>02_27 and 02_29</td>
<td>10 ND 2 3 4 5 8 6 5 1 7 1 24 4 5</td>
</tr>
<tr>
<td>S. Senftenberg</td>
<td>02_18</td>
<td>10 3 2 3 4 5 8 5 5 1 7 1 24 4 5</td>
</tr>
<tr>
<td>S. Senftenberg</td>
<td>03_08</td>
<td>10 4 2 3 4 5 8 6 5 1 7 1 24 4 5</td>
</tr>
<tr>
<td>S. Senftenberg</td>
<td>B59</td>
<td>31 ND 2 3 4 5 6 5 5 1 7 3 24 4 5</td>
</tr>
<tr>
<td>S. Senftenberg</td>
<td>BS37</td>
<td>31 14 2 ND 4 5 6 5 5 1 7 1 24 4 5</td>
</tr>
<tr>
<td>S. Senftenberg</td>
<td>02_19</td>
<td>14 ND 14 3 8 5 5 5 1 12 11 18 4 5</td>
</tr>
<tr>
<td>S. Senftenberg</td>
<td>02_20</td>
<td>14 ND 13 3 8 5 5 5 1 12 11 18 4 5</td>
</tr>
<tr>
<td>S. Senftenberg</td>
<td>02_24</td>
<td>14 ND 13 3 8 5 5 5 1 12 11 18 4 5</td>
</tr>
<tr>
<td>S. Senftenberg</td>
<td>BS42</td>
<td>14 ND 20 12 8 3 1 9 ND ND 39 20 ND 35 40</td>
</tr>
</tbody>
</table>

**Fig. 3.** Sequence types for six serotypes that demonstrated \(fliC\) variation, showing that \(fliC\) sequence types correlated with sequence variation at other loci. The sequences for each amplicon are labelled arbitrarily and sequence similarity is not inferred. Coloured blocks indicate a sequence type that is unique to isolates in this study having a particular \(fliC\) sequence or set of sequences. White blocks indicate that the sequence was shared between isolates with a different \(fliC\) sequence or set of sequences, or where a result was not determined (ND).
(Sukhnanand et al., 2005). This study did not identify such instances; however, mixed sequences would be detected by the technology (Honisch et al., 2007).

For these four pairs of isolates (S. Larchelle and S. Saintpaul, S. Anatum and S. Newington, S. Muenchen and S. Stanley, and S. Enteritidis and S. Gallinarum), the same phase 1 flagellin type was exhibited and there were no DNA sequence differences in the fliC amplicon within the pairs. We were able to differentiate between the isolates using other loci but the sharing of fliC sequences may suggest relationships and would be an interesting path for further research. A larger study including further serotypes may reveal further instances of shared fliC sequence types and, although the study shows the strength of a single amplicon in predicting serotypes and even subtyping isolates, it should be used in combination with other loci in a robust scheme. Sequence variation within the gyrA gene is well documented to correlate with quinolone resistance. Largely, clustering of strains by gyrA amplicon sequence separated quinolone-resistant strains from quinolone-sensitive strains. The results of this study have proved the concept that antibiotic resistance markers could be identified rapidly and concurrently with serotyping and subtyping data.

An advantage of the method developed here is that it does not depend on gene expression. Traditional laboratory techniques for obtaining the full serotype involve forcing the cell to switch between the expression of phase 1 and phase 2 flagellins. If this fails, the isolate is declared not typable. A set of non-typable strains was included in the study. A proportion of these strains had unique fliC sequence types but, as fliC sequences cluster by antigen, the serotype could still be predicted. Furthermore, matching the sequence types of other loci could be used to increase the confidence of this prediction.

Comparative sequence analysis by MALDI-TOF MS is a high-throughput method, suitable for obtaining sequence data from multiple targets for each sample in the same run. The technology provides sufficient time to result, automation and speed to track an ongoing outbreak and is applicable to a range of micro-organisms. Next-generation sequencing is becoming the genetic analysis approach of choice in more and more centres. However, particularly for amplicon sequencing, very large numbers of samples would need to be multiplexed to realize the cost benefits and this may not suit laboratories that need a fast turnaround time. Furthermore, sequencing using the MassARRAY approach is able to obtain read lengths of over 1140 bp, comparable with Sanger sequencing but with the advantage of offering higher throughput and rapid results.

Historically, combinations of typing methods have been used to detect serological and genomic variation. Notwithstanding all its advantages, serology misses important genetic differences that affect pathogenicity, such as the acquisition of virulence factors and antibiotic resistance genes. The method developed here includes combinations of clinically significant genetic markers with molecular serotype for cross-referencing to previous strains. By detecting genetic sequence-based differences that cannot be detected at the phenotypic level, the presented multi-genomarker approach on the MassARRAY system is more discriminatory than traditional serotyping methods. The method has the potential to be used in place of the multiple application approach commonly required to characterize and subtype Salmonella isolates.

**REFERENCES**


