Characterization of new *Salmonella* serovars by whole-genome sequencing and traditional typing techniques

Janet Bale,¹ Daniele Meunier,¹ François-Xavier Weill,² Elizabeth dePinna,¹ Tansy Peters¹ and Satheesh Nair¹

²InstitutPasteur, Unité des Bactéries Pathogènes Entéropathogènes, WHO Collaborating Center for Reference and Research on Salmonella, Paris, France

Serotyping forms the basis of all national and international surveillance networks for *Salmonella*. Public health microbiology is currently being transformed by high-throughput DNA sequencing, which opens the door to serovar determination using this powerful technique. Twenty-nine *Salmonella* isolates referred to the Public Health England between 1994 and 2004 for serovar identification were selected for this study, and they all presented with novel antigenic formulae. Results from a combination of traditional phenotypic and molecular assays were compared. Twenty-two isolates (76%) were subsequently independently confirmed as new types; of these, 18 (82%) were grouped as *Salmonella enterica* subspecies I, and four (18%) were *S. enterica* subspecies II. In general, it is shown that there is concordance between the DNA sequence type and traditional phenotypic serotype, but it would be necessary to analyse a larger data set to confirm this. Traditional multilocus sequence typing (MLST) by Sanger sequencing also correlates to *in silico* whole-genome sequencing MLST. This permits the continuation of traditional serovar nomenclature alongside sequence type methods and enhances the ability to infer true phylogenetic relationships between isolates.

**INTRODUCTION**

Salmonellosis continues to be a major public health concern worldwide. In the European Union, over 90 000 salmonellosis cases are reported every year, with around 8000 of these in the UK alone (Public Health England, 2014). Classification and identification of *Salmonella* are complex and are mainly based on serological detection of antigens.

Serotyping has traditionally been considered as the gold standard method and is performed according to the White–Kauffmann–Le Minor (WKL) scheme (Grimont & Weill, 2007). This scheme identifies the somatic (O) and flagellar (H) antigens based on the agglutination of bacteria with specific antisera. This method has stood the test of time because it is applicable across all the *Salmonella* spp. and is widely understood. It also provides a user-friendly designation for taxonomic differentiation and pathogenic grouping. Despite its widespread use, traditional serotypic identification has a number of drawbacks limiting its full use to a small number of reference laboratories worldwide. Testing often takes ≥3 days, is expensive and laborious and requires specialist resources and technical expertise for the maintenance of >250 typing polyclonal rabbit antisera and reference strains. Moreover, the identification process by serotyping can be complex due to isolates that do not fully express their antigens or show new genetic combinations (Nair et al., 2014).

*Salmonella* strains can be found in a variety of host species and can persist in the environment and also actively exchange genetic material (Andino & Hanning, 2015). As such, we recognize around 10 *Salmonella* with new or novel antigen combinations each year in our laboratory alone. Additionally, we regularly identify *Salmonella* which are rough (autoagglutinating) or monophasic or express only a partial antigenic formula; these isolates cannot be assigned with a definitive serovar name and have to be given an indeterminate identification of *Salmonella* sp. together with their partial antigenic formula. It is possible that, accordingly, we are misinterpreting valuable phenotypic information and missing the identification of new *Salmonella* serovars, i.e. new types not published in the WKL scheme.

Abbreviations: MLST, multilocus sequence typing; PHE, Public Health England; ST, sequence type; WGS, whole-genome sequencing; WKL, White–Kauffmann–Le Minor.
Bacterial isolates. Twenty-nine Salmonella isolates referred to the Public Health England (PHE) between 1994 and 2004 for serovar identification were selected for this study. Each potential new type isolate was stored on microbeads at −80 °C and recovered onto MacConkey agar before subsequent testing. The isolates were received from UK private and public diagnostic laboratories and were from both clinical and food sources.

Serotyping. Isolates were plated on semi-selective agar, and serotyping was carried out using specific rabbit antisera produced in-house. A modified in-house serotyping technique was used (Shipp & Rowe, 1980), and initial results were confirmed by traditional methods (Kauffmann, 1966). The antigenic structure was identified and confirmed according to the WKL scheme.

Biotyping/real-time PCR. All isolates were initially characterized by an in-house biochemical test array based on the methods described by Barrow & Feltham (1993) and organic acid utilization tests by the method described by Kauffmann (1966). The isolates were latterly re-screened after the introduction of two in-house TaqMan® real-time PCR assays for identification of the most common subspecies of Salmonella enterica (subsp. I and III) (Hopkins et al., 2009, 2011).

RESULTS

Twenty-nine isolates were chosen on the basis that they did not conform to a known antigenic structure as published in the WKL scheme at the time of isolation and that they were epidemiologically unrelated. However, because of the longevity of the study and problems with ambiguous serology results, seven of the serovars were subsequently shown to already exist in the recently updated WKL scheme (Table 1).

Serologically, the remaining 22 novel isolates were shown to belong to a diverse range of 14 WKL groups. The most represented was group C, with five isolates, and groups D, E, F, G and O each with two isolates. H antigens were equally heterogeneous.

The real-time PCR, targeting the Salmonella pathogenicity island 1-encoded hilA gene (which encodes an invasion gene transcriptional activator), verified that 18 of the strains were hilA+ confirming their identity as S. enterica subsp. enterica (previously known as subspecies I). Originally, in-house biotyping was used to screen all isolates, and latterly, Biolog allowed us to confirm the early identities and re-evaluate any ambiguous results. Of the remaining four isolates (hilA− and trt+), Biolog confirmed them to be S. enterica subsp. salamae (previously subspecies II). Using a combination of Taqman PCR and Biolog, it was possible to confirm all previously equivocal identifications.

These 22 strains have now been confirmed as new types by the Pasteur Institute and have been assigned names in the WKL scheme (Issenhuth-Jeanjean et al., 2015).

Furthermore, the sequencing data produced in this study have enabled MLST profiles to be assigned to 28 of the original isolates. These new profiles have been included in the
Table 1. Combined results from novel *Salmonella* isolates identified by PHE

<table>
<thead>
<tr>
<th>PHE</th>
<th>Pasteur ID</th>
<th>Phenotype</th>
<th>PHE ST*</th>
<th>Name</th>
<th>MLST profile</th>
<th>TaqMan ID</th>
<th>Biolog ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>New</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S 149374</td>
<td>9996/09</td>
<td>II 35g236xn,m,x</td>
<td>1812</td>
<td>II 35g236xn,m,x</td>
<td>95, 333, 27, 53, 55, 414, 375</td>
<td>SS II or IV</td>
<td><em>S. enterica</em> (gp2) subsp. <em>salamae</em></td>
</tr>
<tr>
<td>S 150132</td>
<td>9997/09</td>
<td>I 17g.m.-</td>
<td>1761</td>
<td>Wembleypark</td>
<td>150, 423, 153, 531, 435, 146, 166</td>
<td>SS I</td>
<td><em>S. enterica</em> (gp2) subsp. <em>salamae</em></td>
</tr>
<tr>
<td>S 161070</td>
<td>9999/09</td>
<td>I 1323:k:1.6</td>
<td>1762</td>
<td>Noho</td>
<td>127, 90, 40, 361, 76, 100, 126</td>
<td>SS I</td>
<td><em>S. enterica</em> (gp1) subsp. <em>enterica</em></td>
</tr>
<tr>
<td>S 161097</td>
<td>10000/09</td>
<td>I 6,72k:25</td>
<td>1760</td>
<td>Ashford</td>
<td>312, 90, 328, 530, 76, 337, 291</td>
<td>SS I</td>
<td><em>S. enterica</em> (gp2) subsp. <em>salamae</em></td>
</tr>
<tr>
<td>S 161496</td>
<td>10002/09</td>
<td>I 51:z:1,2</td>
<td>1763</td>
<td>Elmdon</td>
<td>75, 2, 18, 110, 325, 310, 459</td>
<td>SS I</td>
<td><em>S. enterica</em> (gp1) subsp. <em>enterica</em></td>
</tr>
<tr>
<td>S 166692</td>
<td>10003/09</td>
<td>I 3,10:k:2,g:1,5</td>
<td>1764</td>
<td>Torbay</td>
<td>13, 140, 17, 16, 456, 322, 460</td>
<td>SS I</td>
<td><em>S. enterica</em> (gp2) subsp. <em>salamae</em></td>
</tr>
<tr>
<td>S 168288</td>
<td>10005/09</td>
<td>I 8,20:x:1.6</td>
<td>1765</td>
<td>Colchester</td>
<td>443, 86, 40, 332, 61, 58, 461</td>
<td>SS I</td>
<td><em>S. enterica</em> (gp1) subsp. <em>enterica</em></td>
</tr>
<tr>
<td>S 169259</td>
<td>10006/09</td>
<td>I 6,8:k:1,7</td>
<td>1766</td>
<td>Sandbanks</td>
<td>444, 110, 35, 533, 140, 105</td>
<td>SS I</td>
<td><em>S. enterica</em> (gp1) subsp. <em>enterica</em></td>
</tr>
<tr>
<td>S 170312</td>
<td>10007/09</td>
<td>II 35m:te,n,x</td>
<td>1642</td>
<td>II 35m:te,n,x</td>
<td>260, 176, 27, 514, 18, 231, 174</td>
<td>SS II or IV</td>
<td><em>S. enterica</em> (gp2) subsp. <em>salamae</em></td>
</tr>
<tr>
<td>S 176223</td>
<td>10008/09</td>
<td>I 9,46:y:1,2</td>
<td>1767</td>
<td>Cork</td>
<td>95, 110, 18, 534, 136, 9, 3</td>
<td>SS I</td>
<td><em>S. enterica</em> (gp1) subsp. <em>enterica</em></td>
</tr>
<tr>
<td>S 181143</td>
<td>10010/09</td>
<td>I 39:x:1.5</td>
<td>1769</td>
<td>Firminley</td>
<td>152, 424, 108, 535, 5, 74, 462</td>
<td>SS I</td>
<td><em>S. enterica</em> (gp2) subsp. <em>salamae</em></td>
</tr>
<tr>
<td>S 181968</td>
<td>10011/09</td>
<td>I 4,12z:1,7</td>
<td>1811</td>
<td>Harrow</td>
<td>127, 133, 25, 33, 8, 149, 468</td>
<td>SS I</td>
<td><em>S. enterica</em> (gp1) subsp. <em>enterica</em></td>
</tr>
<tr>
<td>S 183780</td>
<td>10012/09</td>
<td>I 44:te,n:z:5</td>
<td>1770</td>
<td>Goodmayes</td>
<td>2, 157, 44, 536, 163, 9, 463</td>
<td>SS I</td>
<td><em>S. enterica</em> (gp2) subsp. <em>salamae</em></td>
</tr>
<tr>
<td>S 186150</td>
<td>10013/09</td>
<td>I 6,7:z:2,g:z-</td>
<td>1771</td>
<td>Unnamed</td>
<td>127, 90, 329, 532, 7, 22, 464</td>
<td>SS I</td>
<td><em>S. enterica</em> (gp1) subsp. <em>enterica</em></td>
</tr>
<tr>
<td>S 186923</td>
<td>10014/09</td>
<td>II 6,7:x:1,5</td>
<td>1772</td>
<td>II 6,7:x:1,5</td>
<td>25, 77, 27, 22, 18, 81, 91</td>
<td>SS II or IV</td>
<td><em>S. enterica</em> (gp2) subsp. <em>salamae</em></td>
</tr>
<tr>
<td>S 193366</td>
<td>10016/09</td>
<td>II 13,22:gm,mt:z:40</td>
<td>1813</td>
<td>II 13,22:gm,mt:z:40</td>
<td>288, 77, 27, 167, 18, 415, 236</td>
<td>SS II or IV</td>
<td><em>S. enterica</em> (gp2) subsp. <em>salamae</em></td>
</tr>
<tr>
<td>S 199445</td>
<td>10017/09</td>
<td>I 9,46:x:1,6</td>
<td>1773</td>
<td>Poole</td>
<td>37, 20, 330, 62, 115, 221, 22</td>
<td>SS I</td>
<td><em>S. enterica</em> (gp2) subsp. <em>salamae</em></td>
</tr>
<tr>
<td>H04050251†</td>
<td>10018/09</td>
<td>I 3,10:z:en:z:13</td>
<td>877</td>
<td>Kingslynn</td>
<td>13, 208, 49, 141, 41, 19, 270</td>
<td>SS I</td>
<td><em>S. enterica</em> (gp1) subsp. <em>enterica</em></td>
</tr>
<tr>
<td>H041200307‡</td>
<td>10019/09</td>
<td>I 11:z:en:z:15</td>
<td>878</td>
<td>Maybush</td>
<td>127, 273, 40, 72, 304, 282, 196</td>
<td>SS I</td>
<td><em>S. enterica</em> (gp1) subsp. <em>enterica</em></td>
</tr>
<tr>
<td>H043000456</td>
<td>10020/09</td>
<td>I 47:gm:z:1</td>
<td>1774</td>
<td>Carshalton</td>
<td>173, 413, 8, 538, 457, 412, 465</td>
<td>SS I</td>
<td><em>S. enterica</em> (gp1) subsp. <em>enterica</em></td>
</tr>
<tr>
<td>H043700283</td>
<td>10024/09</td>
<td>I 30:z:1,2z:39</td>
<td>1775</td>
<td>Coppetwoodwood</td>
<td>61, 46, 25, 183, 9, 45, 368</td>
<td>SS I</td>
<td><em>S. enterica</em> (gp1) subsp. <em>enterica</em></td>
</tr>
<tr>
<td>H045000667§</td>
<td>10026/09</td>
<td>I 1:1:z:1,2z:36,39:z:15</td>
<td>887</td>
<td>Welwyn</td>
<td>249, 90, 40, 33, 309, 284, 276</td>
<td>SS I</td>
<td><em>S. enterica</em> (gp1) subsp. <em>enterica</em></td>
</tr>
</tbody>
</table>

Existing, *Salmonella* serovars.

Existing, *Salmonella* isolates already present in the WKL scheme. ST data were used to update MLST database.

*ST and MLST profile (concordant by both WGS MLST and traditional MLST).
†H04050251 is recorded as isolate S1 in the Warwick MLST Database.
‡H041200307 is recorded as isolate S2 in the Warwick MLST Database.
§H045000667 is recorded as isolate S11 in the Warwick MLST Database.
updated WKL scheme and the current MLST database held at Warwick University, UK. The MLST profiles produced by traditional Sanger sequencing were in concordance with \textit{in silico} MLST profiles obtained from WGS data.

Using a combination of all techniques described in this study, we were able to produce a unique profile for each \textit{Salmonella} isolate (Tables 1 and S1, available in the online Supplementary Material).

**DISCUSSION**

Classification and identification of \textit{Salmonella} are complex and are mainly based on serological detection of antigens plus an array of biochemical tests. This makes the whole identification process laborious and tedious especially when strains do not fully express their antigens or have unique biochemical markers. Many methods have been developed and assessed for the subtyping of \textit{Salmonella}. Each has a degree of advantage or drawback in terms of cost, speed, robustness and sensitivity (Wattiau \textit{et al}., 2011). As a reference laboratory, we have the potential to identify \textit{Salmonella} with new and often unique antigenic structures and are required to be absolutely certain of their determination. This work requires considerable expertise and technical skill to ensure validation of each new isolate. To meet laboratory accreditation criteria, reporting of results to our customers needs to be timely and accurate. Traditional typing of new types often takes $>3$ weeks for confirmation. A further delay may occur when these isolates are sent for independent verification to three further reference laboratories chosen according to the criteria set by the Institut Pasteur, WHO Collaborating Centre for Reference and Research on \textit{Salmonella}, Paris, France. Only once all laboratories are in agreement is the original sending laboratory invited to provide a name for those serovars belonging to \textit{S. enterica} subsp. \textit{enterica}.

Until recently, the cost of WGS and the associated bioinformatics processing have been expensive and resource intensive. With the advent of improved WGS capacity and an efficient bioinformatics pipeline, the processing of sequence data to obtain final allele types can be done in a shorter amount of time with limited user input and an advantageous cost–benefit. It is also no longer necessary to sequence individual alleles as was the case historically with MLST. The combination of low-cost and high-speed WGS opens the opportunity for it to become a very useful and increasingly embraced technique (Luveshi \textit{et al}., 2015).

Indeed, since April 2015, routine WGS of all referred \textit{Salmonella} isolates has largely replaced and enhanced traditional typing methods at the PHE. The Sanger sequencing described in this paper has now been superseded by an automated DNA extraction method in our laboratory and WGS by the specialized PHE Genome Sequencing Unit (Ashton \textit{et al}., 2016). Obtaining ST is relatively straightforward. An algorithmic pipeline has been developed that enables us to retrieve MLST data for serovar identification. It is possible to assign provisional STs to new \textit{Salmonella} and add the relevant data to our PHE database before submission to the MLST database. When required, it is easy to share these data with our colleagues for confirmation. Our

![Fig. 1. Proposed workflow for new serovar identification.](http://jmm.microbiologyresearch.org)
structured and automated approach to identification, where we have adopted a unique workflow combining traditional and molecular methods, has enabled us to obtain accurate, real-time results with minimal technical input (Fig. 1). We have now adopted this protocol for confirming new type status for all in-house and WHO collaborating confirmation work. Our data are published in a format that allows a worldwide understanding and interpretation especially for those laboratories in the *Salmonella* domain. Our approach can be adopted for the identification of new variants of other micro-organisms. WGS has the power to rapidly detect new *Salmonella* serotypes and also allows us access to a host of additional genetic data including fliC, fljB and rfb genes from which we can confer antigenic status (Zhang et al., 2015). Sequence data can also be used to consult online platforms such as SeqSero (Zhang et al., 2015) by which we can confirm phenotype. Single nucleotide polymorphism analysis for high-resolution typing can be acquired for evolutionary and phylogenetic studies as well (Ashton et al., 2015). These methods enable us to define evolutionarily and epidemiologically meaningful groups of *Salmonella* which will provide important tools for public health and food hygiene purposes.

**ACKNOWLEDGEMENTS**

The work presented in this paper was based on an initial concept by John Wain, University of East Anglia. This project could not have been completed without the help of the staff both past and present at the *Salmonella* Reference Service and Sequencing Service, PHE. In particular, we would like to thank Cath Arnold, Steve Connell, Pushpa Gorasia, Martin Day, Andrew Levy, Clare Maguire, Rediat Kidgell, C., Reichard, U., Wain, J., Linz, B., Torpda, M., Dougan, G. & Achtman, M. (2002). *Salmonella* typhi, the causative agent of typhoid fever, is approximately 50,000 years old infection. Infect Genet Evol 2, 39–45.


