Identification and typing of Salmonella for public health surveillance using whole genome sequencing

Philip M Ashton, Satheesh Nair, Tansy Peters, Janet Bale, David G Powell, Anaïs Painset, Rediat Tewolde, Claire Jenkins, Timothy J Dallman, Elizabeth M DePinna, Kathie A Grant

In April 2015, Public Health England implemented whole genome sequencing (WGS) as a routine typing tool for public health surveillance of Salmonella, adopting a multilocus sequence typing (MLST) approach as a replacement for traditional serotyping. The WGS derived sequence type (ST) was compared to the phenotypic serotype for 6887 isolates of S. enterica subspecies I, and of these, 6616 (96%) were concordant. Of the 4% (n=271) of isolates of subspecies I exhibiting a mismatch, 119 were due to a process error in the laboratory, 26 were likely caused by the serotype designation in the MLST database being incorrect and 126 occurred when two different serovars belonged to the same ST. The population structure of S. enterica subspecies II-IV differs markedly from that of subspecies I and, based on current data, defining the serovar from the clonal complex may be less appropriate for the classification of this group. Novel sequence types that were not present in the MLST database were identified in 8.6% of the total number of samples tested (including S. enterica subspecies I-IV and S. bongori) and these 654 isolates belonged to 326 novel STs. For S. enterica subspecies I, WGS MLST derived serotyping is a high throughput, accurate, robust, reliable typing method, well suited to routine public health surveillance. The combined output of ST and serovar supports the maintenance of traditional serovar nomenclature while providing additional insight on the true phylogenetic relationship between isolates.

1 **Identification and typing of *Salmonella* for public health surveillance using**

2 **whole genome sequencing**

3

4

5 Philip M. Ashton, Satheesh Nair, Tansy Peters, Janet Bale, David Powell, Anais Painset, Rediat Tewolde,

6 Claire Jenkins, Timothy J. Dallman*, Elizabeth M. de Pinna, Kathie A. Grant and the *Salmonella* Whole

7 Genome Sequencing Implementation Group

8

9 Gastrointestinal Bacteria Reference Unit, Public Health England, 61 Colindale Ave, London, NW9 5HT

10

11

12 Corresponding author

13 Timothy J. Dallman, Gastrointestinal Bacteria Reference Unit, Public Health England, 61 Colindale Ave,

14 London, NW9 5HT

15 Email: tim.dallman@phe.gov.uk

16

17

18

19

20

21 Running title: WGS derived MLST serotyping for *Salmonella*

22

23 Key words: whole genome sequencing, multilocus sequence typing, *Salmonella*

24

26 **Abstract**

27 In April 2015, Public Health England implemented whole genome sequencing (WGS) as a routine typing

28 tool for public health surveillance of *Salmonella*, adopting a multilocus sequence typing (MLST)

29 approach as a replacement for traditional serotyping.   The WGS derived sequence type (ST) was

30 compared to the phenotypic serotype for 6887 isolates of *S. enterica* subspecies I, and of these, 6616

31 (96%) were concordant. Of the 4% (n=271) of isolates of subspecies I exhibiting a mismatch, 119 were

32 due to a process error in the laboratory, 26 were likely caused by the serotype designation in the MLST

33 database being incorrect and 126 occurred when two different serovars belonged to the same ST. The

34 population structure of *S. enterica* subspecies II-IV differs markedly from that of subspecies I and, based

35 on current data, defining the serovar from the clonal complex may be less appropriate for the

36 classification of this group. Novel sequence types that were not present in the MLST database were

37 identified in 8.6% of the total number of samples tested (including *S. enterica* subspecies I-IV and *S.*

38 *bongori*) and these 654 isolates belonged to 326 novel STs. For *S. enterica* subspecies I, WGS MLST

39 derived serotyping is a high throughput, accurate, robust, reliable typing method, well suited to routine

40 public health surveillance. The combined output of ST and serovar supports the maintenance of

41 traditional serovar nomenclature while providing additional insight on the true phylogenetic relationship

42 between isolates.

44 **Introduction**

45 The *Salmonella*e are major human pathogens and represent a significant global public health issue

46 causing morbidity and mortality resulting in a high social and economic burden worldwide (*Majowicz et*

47 *al., 2010*). The genus consists of 2 species; *Salmonella enterica* and *S. bongori*. There are six subspecies

48 of *S. enterica* differentiated by biochemical variations, namely subspecies *enterica* (I), *salamae* (II),

49 *arizonae* (IIIa), *diarizonae* (IIIb), *houtenae* (IV) and *indica* (VI) (*Threlfall et al. 1999*). Subspecies I, *S.*

50 *enterica* subsp. *enterica* cause 99% of human and animal infections. The two main pathologies

51 associated with *S. enterica* are gastroenteritis and typhoidal disease. The typhoidal *Salmonella*e include

52 *S*. Typhi and *S*. Paratyphi A, B and C. They are host restricted, monophyletic, rarely undergo

53 recombination events and exhibit convergent evolution driven by genome degradation (*Wain et al.*

54 *2015*). The majority of gastroenteritis in the UK is caused by the host generalist serovars, such as *S.*

55 Typhimurium and *S*. Enteritidis, and host adapted serovars that are adapted to a specific animal

56 reservoir but can infect man and include *S*. Dublin, *S*. Gallinarum *S*. Choleraesuis, and *S*. Bovismorbificans

57 (*Langridge et al. 2015*).

58

59 Approximately 8,000 isolates are referred to the *Salmonella* Reference Service (SRS) at Public Health

60 England (PHE) each year from local and regional hospital laboratories. In April 2015, PHE implemented

61 whole genome sequencing (WGS) as the routine typing tool for public health surveillance of *Salmonella*

62 infections. Prior to April 2015, presumptive *Salmonella* isolates referred to SRS were speciated and sub-

63 speciated using PCR *(Hopkins et al. 2009, 2011)* and grouped into serovars as described in the White-

64 Kauffman-Le Minor scheme (*Grimont & Weill 2007, Guibourdenche et al. 2010, Issenhuth-Jeanjean et al.*

65 *2014*). This methodology is based on reactions of rabbit antisera to the lipopolysaccharide (O antigen

66 encoded by *rfb* genes) and flagellar antigens (phases 1 and 2 of H antigen encoded by *fli*C and *flj*B). The

67 scheme utilises this phenotypic variation, expressed as an antigenic formulae, to divide *Salmonella* into

68 more than 2600 serovars. Epidemiological investigations of *Salmonella* infecting humans and animals

69 have relied on serotyping for over 70 years; national and international governmental agencies base

70 guidelines and regulations on the serotyping method and the use of this nomenclature is a globally

71 recognised form of communication (*Swaminathan et al. 2009, EFSA 2010*). Furthermore, serovars have

72 often been shown to correlate with host range and disease sequelae (*Gordon et al. 2011, Wain et al.*

73 *2015, Langridge et al. 2015*).

74

75   There are, however, a number of issues with the serotyping approach; specifically, the expense and

76   expertise required to produce the antisera and , furthermore, serotyping does not reflect the genetic

77   relatedness between serovars, nor does it provide an evolutionary perspective. Alternative molecular

78   serotyping methods have been described previously including Pulsed-field gel electrophoresis,

79   ribotyping, repetitive extragenic palindromic sequence-based PCR (rep-PCR) and combined PCR- and

80   sequencing-based approach that directly targets O- and H-antigen-encoding genes (*Ranieri et al. 2013,*

81   *Shi et al. 2015*).   In 2012, Achtman *et al.* proposed a sequenced based approach, multilocus sequence

82   typing (MLST), based on the sequences of multiple house-keeping genes. Isolates that possess identical

83   alleles for the seven gene fragments analysed are assigned a common sequence type (ST) and related

84   STs from clonal complexes are termed e-Burst Groups (eBGs). They showed that ST and eBGs strongly

85   correlated with serovar and so utilising this approach would facilitate backward compatibility with

86   historical data, minimise disruption for reference laboratory service users and facilitate data exchange

87   with other colleagues in the field.

88

89   Advances in whole genome sequencing (WGS) methodologies have resulted in the ability to perform

90   high throughput sequencing of bacterial genomes at low cost making WGS an economically viable

91   alternative to traditional typing methods for public health surveillance and outbreak detection (*Koser et*

92   *al. 2012*). Whilst WGS provides the opportunity to resolve bacterial strains to the single nucleotide

93   resolution needed for identifying cases linked to a common source of infection (*Dallman et al. 2015*),

94   grouping isolates into higher taxonomical clones (e.g. those defined by serotyping) is an important step.

95   The decision to adopt WGS as a routine typing method at PHE provided the opportunity to review our

96   approach to typing *Salmonella* and to implement the MLST approach in parallel with WGS.

97

98   The aim of this study was to evaluate MLST, as derived from WGS data, as a replacement for

99   conventional serotyping of *Salmonella* for routine public health surveillance and to provide insight into

100   the genetic population structure of all *Salmonella* species in England and Wales during a 12 month

101   period.

102

103   **Methods**

104   **Bacterial strains**

105   All isolates (n=7465) of *Salmonella* from human cases of gastrointestinal disease submitted to SRS from

106   local and regional hospital laboratories in England & Wales between 1st April 2014 and 31st March 2015

107   were sequenced in parallel with phenotypic serotyping (Supplementary Table). Of these, 7338 were

108   identified as subspecies I and included 263 different serovars. The ten most common serovars in this

109   dataset were Enteriditis (2310), Typhimurium (1407), Infantis (184), Typhi (184), Newport (173), Virchow

110   (162), Kentucky (160), Stanley (146), Paratyphi A (135) and Java (99). One hundred and twenty seven

111   isolates were identified as subspecies II-IV (*S. enterica* subspecies *salamae* n=28; *S. enterica* subspecies

112   *arizonae* n=25; *S. enterica* subspecies *diarizonae* n=49; *S. enterica* subspecies *houtenae* n=20) and there

113   was one isolate of *S. bongori*. No isolates belonging to subspecies VI (*S. enterica* subspecies *indica*) were

114   submitted to SRS during the study period.

115

116   **DNA extraction for WGS**

117   DNA extraction of *Salmonella* isolates was carried out using a modified protocol of the Qiasymphony

118   DSP DNA midi kit (Qiagen). In brief, 0.7 ml of overnight *Salmonella* culture in a 96 deep well plate was

119   harvested. Bacterial cells were pre-lysed in 220 µl of ATL buffer (Qiagen) and 20 µl Proteinase K

120   (Qiagen), and incubated shaking for 30 mins at 56°C. Four µl of RNase at 100 mg/ml (Qiagen) was added

121   to the lysed cells and re-incubated for a further 15 mins at 37°C. This step increases the purity of the

122   DNA for further downstream sequencing. Extraction of DNA from the treated cells was performed on

123   the Qiasymphony SP platform (Qiagen) and eluted in 100 µl of water. DNA concentration using the

124   GloMax system (Promega) was determined for the following sequencing steps.

125

126   **DNA sequencing**

127   Extracted DNA was then processed using the NexteraXT sample preparation method and sequenced

128   with a standard 2x101 base protocol on a HiSeq 2500 Instrument in fast mode (Illumina, San Diego).

129

130   **Bioinformatics Service Workflow**

131   FASTQ reads were quality trimmed using Trimomatic (PMID 24695404) with bases removed from the

132   trailing end that fell below a PHRED score of 30. If the read length post trimming was less than 50bp the

133   read and its pair were discarded. A K-mer identification step was used to compare the sequenced reads

134   with 1769 published genomes to identify the bacterial species (and *Salmonella* subspecies) and to detect

135   cultures submitted by the local and regional hospital laboratories that contained more than one

136   bacterial species (mixed cultures). ST assignment was performed using a modified version of SRST

137   (PMID 25422674). Preliminary analysis was undertaken using the MLST database described in Achtman

138   *et al.* (2012).

139

140 For isolates that had novel STs, or a ST but no associated serovar in the Achtman MLST database, the

141 serovar was determined by phenotypic serotyping at PHE. STs and corresponding serovars of isolates

142 serotyped and sequenced during this study were added to a modified version of the Achtman MLST

143 database, held and curated at PHE. These novel STs were assigned a preliminary ST (PST) and an

144 inferred serovar was determined. The PHE MLST database currently holds 7000 strains and 1,200

145 serovars and is up-dated every three months.

146

147 **Results**

148 Achtman *et al.* (2012) described the population structure of *Salmonella enterica* as monophyletic

149 lineages of STs that have evolved from a single founder node and termed these discrete clusters eBGs.

150 The population structure of all the *Salmonella* species submitted to PHE between April 2014 and March

151 2015 is illustrated by the minimum spanning tree in Figure 1.

152

153 *Salmonella subspecies I*

154 The ST and corresponding serovar designation obtained from the MLST database were used to compare

155 the WGS derived ST to the phenotypic serotype for 6887 (94%) of 7338 isolates of subspecies 1, and of

156 these, 6616 (96%) had the same result by both methods (Supplementary Table). It was not possible to

157 compare phenotypic serotyping with MLST-based serotyping for 451 (6%) subspecies I isolates because

158 either the phenotypic serotype could not be determined due to an incomplete antigenic structure (*S.*

159 Unnamed) (n=423) or the serovar could not be determined because the ST did not have a designated

160 serotype in the MLST database (n=70). Forty-two isolates were both *S.* Unnamed and had no MLST

161 designated serotype.

162

163 For the 423 (5.8%) subspecies I isolates reported as *S.* Unnamed, 318 (90%) were designated a serotype

164 from the WGS derived MLST data. The most common serovars typed in this way included *S.*

165 Typhimurium (118), *S.* Virchow (30), S. Stanley (17), *S.* Enteritidis (16), *S.* Infantis (14) and *S.* Thompson

166 (13). Of the 7338 strains tested, 70 (1%) had no serotype designation in the MLST database, of which 28

167 (40%) were serotyped phenotypically (Supplementary Table).

168

169 *Subspecies I novel sequence types*

170   Novel sequence types that were not present in the MLST database were identified in 8.6% (n=654) of

171   the strains (Supplementary Table). These 654 isolates belonged to a total of 326 novel STs, designated

172   PST; the modal number of isolates identified per PST was one (Figure 2a). There was no difference in

173   the distribution of number of isolates per PST depending on whether the PST had a known serovar or

174   belonged to an unnamed or ambiguous serotype. The rate at which PSTs were received throughout the

175   year was plotted and revealed a linear relationship ($R^2 = 0.98$, y = 1.04 * x, where x = number of days

176   since April 1st 2014) (Figure 2b).

177

178   The serovars with the highest number of new PSTs were *S*. Typhimurium (n=9), *S*. Stanley (n=9), *S*.

179   Enteritidis (n=9) and *S*. Newport (n=8), although the majority of these PSTs were single locus variants

180   (SLVs) of established STs, belonging to these serovars (*S*. Typhimurium 8/9, *S*. Stanley 7/9 and *S*.

181   Newport 7/8). There were also serovars for which a large number of PSTs were identified that were not

182   SLVs of established STs (S. Agama 5/5, *S*. Agbeni 5/5, *S*. Saint-Paul 5/5, *S*. Enteritidis 4/9) which may

183   represent new eBGs that share these serotypes.

184

185   *Subspecies I mismatches*

186   Four percent (n= 271) of the isolates tested exhibited a mismatch between the WGS MLST derived

187   serovar and the phenotypic serotyping results. Of the 271 mismatches, 119 were due to a process error

188   in the laboratory either in the phenotypic serotyping or the DNA extraction part of the WGS pipeline.

189   With respect to the phenotypic serotyping, common errors included mislabelling samples and

190   misinterpreting or incorrectly transcribing the antigenic structure, especially when the antigenic

191   structures were similar. For example, *S*. Agona (I 4, 12:f, g, s:-) and *S*. Derby (I 4, 12:f, g:-). DNA

192   extraction errors were associated with mislabelled samples.

193

194   Twenty-six mismatches were potentially caused by the predicted serotype designation in the Achtman

195   MLST database being incorrect which may be attributed to single entries that had been misidentified at

196   the laboratory from which the MLST data was submitted. For example, in the original database ST1499

197   is represented by one entry identified by the submitter as *S*. Litchfield. Subsequently, phenotypic

198   serotyping at PHE identified this ST as *S.* Bovis-morbificans in more than five isolates. ST1499 belongs to

199   eBG34 which comprises two other STs both associated with S. Bovis-morbificans, indicating that the

200   original entry in the MLST database is likely to be incorrect.

201

202   The most common reason for mismatches occurring between the WGS MLST derived serotype and the

203   phenotypic serotype (n=126) occurred when two different serovars belonged to the same eBG and the

204   same ST (see Table 1 and discussed in more detail below).

205

206   *Serovars Enteritidis and Dublin*

207   Of the 2308 isolates of *S.* Enteritidis identified by both phenotypic serotyping and WGS MLST derived

208   serotyping, 2296 belonged to eBG4, including 2200 ST11 and 76 ST183 (Figure 1). There were five

209   additional SLVs of ST11, four of which were novel types. *S.* Gallinarum and *S.* Pulloram can be difficult to

210   distinguish from *S.* Enteritidis (*Thomson et al. 2008*) but neither of these serovars were identified in this

211   study. Serologically, *S.* Dublin ([1],9,12:g,p:-) has a similar antigenic structure to *S.* Enteritidis

212   ([1],9,12:g,m:-), and in Achtman *et al.* (2012), eBG32 (ST74) contained both *S.* Enteritidis and *S.* Dublin.

213   However, in this study both isolates belonging to ST74 eBG32 typed as *S.* Enteriditis. Of the 2308

214   isolates, 26 belonged to nine new PSTs. The most common was P3147, a previously undescribed SLV of

215   ST11, comprising 16 cases including 10 known to have travelled to Malaysia or Singapore.

216

217   *Serovar Typhimurium*

218   In this study, eBG1 contained 1392 isolates of *S.* Typhimurium and monophasic *S.* Typhimurium (rough

219   and non-motile variants) (*Hopkins et al. 2012*). The monophasic variants also belong to eBG138

220   (primarily ST 36) and eBG243. In contrast to eBG1 described in Achtman *et al.* (2012), which was

221   represented by a large central ST19 node with at least 27 SLV STs comprising much smaller numbers of

222   strains, eBG1 in the PHE dataset shows a predominance of both ST19 and ST34 and less allelic variation.

223   Only nine SLVs to ST19 were identified including three undesignated STs (Figure 1).

224

225   *Serovars Java/Paratyphi B data*

226   Despite the different disease outcomes associated with *S.* Paratyphi B (most commonly associated with

227   invasive disease and paratyphoid fever) and *S.* Java (most commonly associated with gastroenteritis) it is

228   not possible to differentiate the two serotypes by serotyping alone. *S.* Java and *S.* Paratyphi B are

229   therefore differentiated in the laboratory by their ability to ferment dextrorotatory tartrate (*S.* Java *d*Ta+

230   and *S.* Paratyphi B *d*Ta-) (*Malorny et al. 2003*).

231

232   The 99 isolates identified by both phenotypic serotyping and WGS MLST derived serotyping as *S.* Java,

233   belonged to a diverse range of eBGs, STs and PSTs (Table 2 and Figure 1). Two of these 99 isolates

234   (marked with * in Table 2) belonged to ST86 and the predicted serotype from the MLST database was *S.*

235   Paratyphi B. One of these isolates was from a blood culture (associated with invasive disease) and,

236   therefore, likely to have been misidentified phenotypically. All 12 isolates identified as *S.* Paratyphi B

237   phenotypically, were identified as *S.* Paratyphi B ST86 by WGS MLST.

238

239   *Subspecies II-IV and S. bongori*

240   Isolates from subspecies II, III, IV, VI and *S. bongori* were not well represented in Achtman's MLST

241   database and thus the majority of isolates from these sub-species sequenced in this study did not

242   belong to a previously designated eBG or ST. The population structure of the 127 non-subspecies I

243   isolates differs markedly from that of *Salmonella enterica* (subspecies 1) (Figure 1) and shows some

244   similarity to the population structure of lineage 3 in being a connected network of STs.

245

246   Sixteen of the 28 isolates belonging to subspecies II were previously designated *S.* Unnamed and the 28

247   strains belonged to 20 different STs. There were 25 isolates classed as subspecies IIIa (belonging to 10

248   different STs) and 49 in subspecies IIIb (belonging to 27 different STs). Of the 20 isolates identified as

249   subspecies IV, 10 were designated *S.* Wassenaar (P3029) by phenotypic serotyping and the 20 isolates

250   belonged to five different STs. All isolates of subspecies II-IV and *S. bongori* were correctly speciated

251   using the k-mer ID approach.

252

253   *Population structure*

254   As highlighted by Achtman *et al.* (2012), the majority of isolates in the dataset belong to eBGs that have

255   a one-to-one relationship with a specific serovar including *S.* Typhi, *S.* Paratyphi A and *S.* Heidelberg. In

256   this study, of the serovars comprising more than 25 isolates, there were 17 serovar specific eBGs and 10

257   examples of a single serovar being associated with multiple eBGs (Figure 1). There were at least six

258   examples of more than one serovar belonging to the same eBG but different STs, for example *S.* Hadar

259   (ST33) and *S.* Kottbus (ST582) both belong to eBG22 and *S.* Bredeney (ST306) and *S.* Schwarzengrund

260   (ST96) both belong to eBG33 (Supplementary Table).

261

262   There were seven examples where two serovars belonged to the same eBG and the same ST (Table 1).

263   In all of these examples, the antigenic structures of the two serovars were similar with only one antigen

264   differentiating the two serovars. Further analysis was carried out on two examples to determine

265   whether this difference in antigenic structure represented a true difference in strain relatedness or a

266   random change that is not reflected in phylogeny (for example, the insertion of phage encoded antigen).

267   The analysis showed that the change in antigenic structure in *S.* Richmond (I 6,7:y:1,2) and *S.* Bareilly (I

268   6, 7:y:1,5), both ST 909, and in *S.* Saintpaul (I 4,5,12: e,h: 1,2) and *S.* Haifa (I 4,5,12: z,10: 1,2), both ST49,

269   reflected a true phylogenetic difference (Figures 3a and 3b).

270

271   The same higher strata population structure referred to as lineage 3 for *S. enterica* subspecies I, as

272   described by Achtman *et al.* (2012), was observed in this dataset (Figure 4). Genomes of these

273   *Salmonella*e are in constant flux and homologous recombination among unrelated eBGs is frequent

274   (*Achtman et al. 2012, Didelot et al. 2011*). Serovars in this lineage mainly consists of multiple eBGs and

275   are polyphyletic by nature. Achtman *et al.* (2012) suggested that the population structure of lineage 3

276   does not comprise of independent startbursts, as observed with other serovars of subspecies I, but

277   rather a connected network (Figure 4). The five most common examples of this in the current study,

278   were *S.* Oranienburg, *S.* Montevideo, *S.* Chester, *S.* Poona and *S.* Bredeney (Figure 4 and Supplementary

279   Table). These five serovars are not represented in the top 10 serovars submitted to SRS during this

280   surveillance period.

281

282   *K-mer identification*

283   There were 249 cultures submitted to SRS by the local hospital and regional laboratories for *Salmonella*

284   typing that were a mix of *Salmonella* and non-*Salmonella* species. These were identified by the k-mer

285   identification step and included 138 *Escherichia coli*, 40 *Morganella morganii*, 11 *Citrobacter species* and

286   four *Escherichia albertii*.

287

288   **Discussion**

289   In their seminal 2012 paper Achtman and colleagues argued convincingly for replacing serotyping with a

290   MLST approach based on genetic population groupings for typing *S. enterica* (Achtman *et al. 2012*). The

291   key aspects of this approach that led PHE to adopt this strategy were (i) the robustness of the

292   population structure as defined by the natural eBG clusters (ii) the fact that eBG designation provides an

293   accurate representation of strain relatedness and (iii) that this approach lends itself to automation. At

294   the same time, it was necessary for PHE to maintain serovar nomenclature in order to facilitate data

295   exchange with other colleagues in the field and maintain backward compatibility with historical data. It

296   was suggested that by using the MLST approach to infer serovar, and by reporting both inferred serovar

297   and ST, it would be possible to utilise the advantages of both methods and implement a state-of-the-art

298   typing system while keeping disruption for reference laboratory service users to a minimum.

299

300   The PHE dataset of 6887 subspecies I isolates that were serotyped using both traditional phenotypic

301   methods and a derived serotype based on MLST data extracted from the genome during a 12 month

302   time frame, provided further evidence of the robustness of the ST/eBG approach to typing. The 96%

303   concordance between the two techniques in a reference laboratory setting is evidence of the validity

304   and suitability of this approach. There were 451 isolates that had to be excluded from the comparison

305   because both types of data (phenotypic and genotypic) were not available. Of these, for 94% of the

306   isolates, it was the phenotypic serotype that could not be determined indicating that WGS MLST derived

307   serotyping is more robust.

308

309   The PHE dataset included single serovars associated with multiple eBG, for example *S.* Typhimurium and

310   *S.* Newport (*Sangal et al. 2010, Achtman et al. 2012*) and multiple serovars belonging to the same eBG

311   but with different STs, for example *S.* Java (ST43) and *S.* Paratyphi (ST86) both belong to eBG5 (*Achtman*

312   *et al. 2012*). In both these scenarios, the correct serovar was determined from the MLST WGS data and

313   the combination of serovar and ST/eBG provided insight into the true phylogenetic relationship between

314   isolates. This data clearly supports Achtman and colleagues argument that eBG and ST designation

315   provides a more accurate representation of strain relatedness than the traditional serovar designation.

316   The phenomenon of multiple serovars belonging to the same ST (for example *S.* Richmond/*S.* Bareilly

317   and *S.* Haifa/*S.* Saintpaul) was a rare but important example of serotyping providing a higher level of

318   strain discrimination within a ST. These strains could be differentiated *in silico* using a tool to infer

319   serovar from the genes that determine antigenic struture, such as seqsero (*Zhang et al. 2015*).

320

321   Despite the implementation of WGS, a limited phenotypic serotyping facility continues to be maintained

322   at PHE in order to serotype isolates that cannot be matched to a serovar; either because the ST in the

323   MLST database has no serovar designation or the ST is a novel type. Additionally, it ensures that we

324   maintain the ability to perform the standard reference method for serotyping *Salmonella*. The PHE MLST

325   database is regularly up-dated to include STs recently matched to a serotype by linking the ST to PHE

326   phenotypic serotyping data and novel PSTs. This approach was adopted because at the time of analysis,

327   the Achtman MLST database was not accepting submissions generated by WGS. There was no decrease

328   in the rate at which PSTs were observed during the 12 month study period and the majority of PSTs

329   were only sampled once in that time frame. Many PSTs were SLVs of known STs, indicating that we have

330   not yet sampled the full diversity of known eBGs. New PSTs, not part of any previously identified eBG,

331   were also observed and further diversity was found within *S. enterica* subspecies II-IV and the lineage 3

332   population. This suggests that there is a large amount of previously unidentified diversity within the

333   species *Salmonellae* associated with both domestically acquired and travel related gastrointestinal

334   disease in human cases resident in England and Wales.

335

336   Isolates exhibiting monophasic properties that could not be fully serotyped phenotypically because they

337   had an incomplete antigenic structure were matched to a ST derived serotype. The monophasic variants

338   in this study mainly belonged to eBG1, eBG138 and eBG243 and previous studies have also shown that

339   monophasic variants of *S.* Typhimurium have emerged as a result of multiple independent genetic

340   events (*Soyer et al. 2009, Switt et al. 2009, Tennant et al. 2010*). Strains with monophasic properties are

341   reportable to European Centre for Disease Prevention and Control (ECDC) but cannot be determined

342   using the ST approach. Alternative strategies for determining monophasic characteristics by PCR are

343   available (*Prendergast et al. 2013*) and methods for extracting this information from the genome

344   sequencing data have been developed at PHE (Personal communication: Philip Ashton & Anna Lewis,

345   publication in preparation).

346

347   In contrast to *S.* Typhurmurium, where ST could not be used to determine monophasic characteristics, in

348   this study ST was able to differentiate the complex relationship between *S.* Java (Hazard Group (HG) 2

349   organism) and *S.* Paratyphi B (HG3) with the latter belonging to either ST42 or ST86. If this ST

350   designation proves to be robust, MLST will facilitate the diagnosis of invasive disease and life

351   threatening paratyphoid fever.

352

353   The MLST derived serovar correlated well with the traditional serovar designation and demonstrated

354   many advantages over traditional phenotypic serotyping. Monophasic strains with incomplete antigenic

355   structures were accurately assigned to serotypes. Phenotypic serotyping errors, such as misinterpreting

356   or incorrectly transcribing the antigenic structure, were avoided. Novel types were identified, confirmed

357   and given a PST designation. Finally, this approach lends itself to automation and rapid, high-throughput

358   processing.

359

360 Two main issues arose during the evaluation of the MLST approach: (i) a number of STs did not have a

361 serovar designation in the MLST database (including subspecies II to IV) and (ii) the unexpectedly large

362 number of novel STs identified. Traditional phenotypic serotyping was required to type these isolates

363 and the MLST database was modified and up-dated to incorporate the new data. Clearly, as we move

364 forward the PHE MLST database will be constantly evolving and this data will be shared with colleagues

365 in the field via existing MLST databases and their WGS compliant successors e.g. EnteroBase & BIGSdb.

366 While it is difficult to draw conclusions based on our small sample size, MLST may not currently be an

367 appropriate tool for the classification of *Salmonella* sub-species II-IV, due to the lack of a discrete

368 population structure of EBGs. However, non-subspecies I isolates which are mainly adapted to cold

369 blooded animals and/or reptiles contributed to less than 1.7% of the workload during the time frame of

370 the study. Although MLST approach is generally more discriminatory than serovar, it does not always

371 provide the fine resolution required for public health surveillance. Further analysis based on single

372 nucleotide polymorphisms in the core genome compared to a type strain representing the most

373 common eBGs is performed for outbreak detection and investigation (*Ashton et al. 2014*).

374

375 In conclusion, WGS MLST derived serotyping is an accurate, robust, reliable, high throughput typing

376 method that is well suited to routine public health surveillance of *Salmonella*. This approach supports

377 the maintenance of traditional serovar nomenclature and provides further insight on the true

378 evolutionary relationship between isolates, as well as a framework for fine level typing within eGBs for

379 surveillance, outbreak detection and source attribution.

380

381

382 **Acknowledgements**

388

389 **Data Deposition**

390 All data from the *Salmonella* surveillance project are deposited in the BioProject of the SRA

391 PRJNA248792.

References

Achtman M, Wain J, Weill FX, Nair S, Zhou Z, Sangal V, Krauland MG, Hale JL, Harbottle H, Uesbeck A, Dougan G, Harrison LH, Brisse S, S. Enterica MLST Study Group. 2012. Multilocus sequence typing as a replacement for serotyping in *Salmonella* enterica. *PLoS Pathogens* 8(6):e1002776.

Ashton PM, Peters T, Ameh L, McAleer R, Petrie S, Nair S, Muscat I, de Pinna E, Dallman T. 2015. Whole Genome Sequencing for the Retrospective Investigation of an Outbreak of Salmonella Typhimurium DT 8. *PLoS Curr* 10:7. doi: 10.1371/currents.outbreaks.2c05a47d292f376afc5a6fcdd8a7a3b6.

Dallman TJ, Byrne L, Ashton PM, Cowley LA, Perry NT, Adak G, Petrovska L, Ellis RJ, Elson R, Underwood A, Green J, Hanage WP, Jenkins C, Grant K, Wain J. 2015. Whole-genome sequencing for national surveillance of Shiga Toxin-producing *Escherichia coli* O157. *Clinical Infectious Diseases* 61(3):305-12.

Didelot X, Bowden R, Street T, Golubchik T, Spencer C, McVean G, Sangal V, Anjum MF, Achtman M, Falush D, Donnelly P. 2011. Recombination and population structure in *Salmonella enterica*. *PLoS Pathogens* 7:e1002191

European Food Safety Authority. 2010. Scientific opinion on monitoring and assessment of public health risk of "*Salmonella* Typhimurium" strains. *EFSA Journal* 8.

Grimont PAD, Weill FX. 2007. Antigenic Formulae of the *Salmonella* Serovars (ninth ed.)WHO Collaborating Center for Reference and Research on Salmonella, Institut Pasteur, Paris (2007) http://www.pasteur.fr/sante/clre/cadrecnr/salmoms/WKLM_En.pdf

Gordon MA. 2011. Invasive nontyphoidal *Salmonella* disease: epidemiology, pathogenesis and diagnosis. *Current Opinion Infectious Diseases* 24:484-489.

Guibourdenche M, Roggentin P, Mikoleit M, Fields PI, Bockemuhl J, Grimont PA, Weill FX. 2010. Supplement 2003-2007 (No. 47) to the White-Kauffmann-Le Minor scheme. *Research in Microbiology* 161:26-29.

Hopkins KL, Lawson AJ, Connell S, Peters TM, de Pinna E. 2011. A novel real-time polymerase chain reaction for identification of *Salmonella enterica* subspecies enterica. *Diagnostic Microbiology & Infectious Disease* 70(2):278-80.

Hopkins KL, Peters TM, Lawson AJ, Owen RJ. 2009. Rapid identification of *Salmonella enterica s*ubsp. arizonae and S. enterica subsp. diarizonae by real-time polymerase chain reaction. *Diagnostic Microbiology & Infectious Disease* 64(4):452-4.

Hopkins KL, de Pinna E, Wain J. 2012. Prevalence of *Salmonella enterica* serovar 4,[5],12:i:- in England and Wales, 2010. *Euro Surveillance* 17(37). pii: 20275

Issenhuth-Jeanjean S, Roggentin P, Mikoleit M, Guibourdenche M, de Pinna E, Nair S, Fields PI, Weill FX. 2014. Supplement 2008-2010 (no. 48) to the White-Kauffmann-Le Minor scheme. *Research in Microbiology* 165(7):526-30.

440

441 **Koser CU, Ellington MJ, Cartwright EJ, Gillespie SH, Brown NM, Farrington M, Holden MT, Dougan G,**
442 **Bentley SD, Parkhill J, Peacock SJ. 2012.** Routine use of microbial whole genome sequencing in
443 diagnostic and public health microbiology. *PLoS Pathogens* 8:e1002824.

444

445 **Langridge GC, Fookes M, Connor TR, Feltwell T, Feasey N, Parsons BN, Seth-Smith HM, Barquist L,**
446 **Stedman A, Humphrey T, Wigley P, Peters SE, Maskell DJ, Corander J, Chabalgoity JA, Barrow P,**
447 **Parkhill J, Dougan G, Thomson NR. 2015.** Patterns of genome evolution that have accompanied host
448 adaptation in *Salmonella*. *Proceedings of the National Academy of Sciences USA* 112(3):863-8.

449

450 **Majowicz SE, Musto J, Scallan E, Angulo FJ, Kirk M, O'Brien SJ, Jones TF, Fazil A, Hoekstra RM;**
451 **International Collaboration on Enteric Disease 'Burden of Illness' Studies. 2010.** The global burden of
452 nontyphoidal *Salmonella* gastroenteritis. *Clinical Infectious Diseases* (6):882-9.

453

454 **Malorny B, Bunge C, Helmuth R. 2003.** Discrimination of d-tartrate-fermenting and -nonfermenting
455 *Salmonella enterica* subsp. enterica isolates by genotypic and phenotypic methods. *Journal of Clinical*
456 *Microbiology* 41(9):4292-7.

457

458 **Prager R, Rabsch W, Streckel W, Voigt W, Tietze E, Tschäpe H. 2003.** Molecular properties of
459 *Salmonella enterica* serotype paratyphi B distinguish between its systemic and its enteric pathovars.
460 *Journal of Clinical Microbiology* 41(9):4270-8.

461

462 **Prendergast DM, Hand D, Ní Ghallchóir E, McCabe E, Fanning S, Griffin M, Egan J, Gutierrez M. 2013.** A
463 multiplex real-time PCR assay for the identification and differentiation of *Salmonella enterica* serovar
464 Typhimurium and monophasic serovar 4,[5],12:i:-. *International Journal Food Microbiology* 166(1):48-53.
465 doi: 10.1016/j.ijfoodmicro.2013.05.031.

466

467 **Ranieri ML, Shi C, Moreno Switt AI, den Bakker HC, Wiedmann M. 2013.** Comparison of typing methods
468 with a new procedure based on sequence characterization for *Salmonella* serovar prediction. *Journal of*
469 *Clinical Microbiology* 51(6):1786-97. doi: 10.1128/JCM.03201-12.

470

471 **Sangal V, Harbottle H, Mazzoni CJ, Helmuth R, Guerra B, Didelot X, Paglietti B, Rabsch W, Brisse S,**
472 **Weill FX, Roumagnac P, Achtman M. 2010.** Evolution and population structure of *Salmonella enterica*
473 serovar Newport. *Journal of Bacteriology* 192(24):6465-76.

474

475 **Shi C, Singh P, Ranieri ML, Wiedmann M, Moreno Switt AI. 2015.** Molecular methods for serovar
476 determination of *Salmonella*. *Critical Reviews Microbiology* 41(3):309-25. doi:
477 10.3109/1040841X.2013.837862.

478

479 **Soyer Y, Moreno SA, Davis MA, Maurer J, McDonough PL, Schoonmaker-Bopp DJ, Dumas NB, Root T,**
480 **Warnick LD, Grohn YT, Wiedmann M. 2009.** *Salmonella* enterica serotype 4,5,12:i:-, an emerging
481 *Salmonella* serotype that represents multiple distinct clones. *Journal of Clinical Microbiology* 47: 3546–
482 3556.

483

484 **Swaminathan B, Gerner-Smidt P, Ng LK, Lukinmaa S, Kam KM, Rolando S, Gutierrez EP, Binsztein N.**
485 **2006.** Building PulseNet International: an interconnected system of laboratory networks to facilitate
486 timely public health recognition and response to foodborne disease outbreaks and emerging foodborne
487 diseases. *Foodborne Pathogens and Disease* 3:36-50

488
489 **Switt AI, Soyer Y, Warnick LD, Wiedmann M. 2009.** Emergence, distribution, and molecular and
490 phenotypic characteristics of ***Salmonella* enterica** serotype 4,5,12:i:-. *Foodborne Pathogens and Disease*
491 6(4):407-15. doi: 10.1089/fpd.2008.0213.
492
493 **Tennant SM, Diallo S, Levy H, Livio S, Sow SO, Tapia M, Fields PI, Mikoleit M, Tamboura B, Kotloff KL,**
494 **Nataro JP, Galen JE, Levine MM. 2010.** Identification by PCR of non-typhoidal *Salmonella enterica*
495 serovars associated with invasive infections among febrile patients in Mali. *PLOS Neglected Tropical*
496 *Diseases* 4: e621.
497
498 **Thomson NR, Clayton DJ, Windhorst D, Vernikos G, Davidson S, *et al.* 2008.** Comparative genome
499 analysis of *Salmonella* Enteritidis PT4 and *Salmonella* Gallinarum 287/91 provides insights into
500 evolutionary and host adaptation pathways. *Genome Research* 18(10):1624-37.
501
502 **Threlfall J, Ward L, Old D. 1999.** Changing the nomenclature of *Salmonella*. *Communicable Diseases*
503 *Public Health* 2(3):156-7.
504
505 **Wain J, Hendriksen RS, Mikoleit ML, Keddy KH, Ochiai RL. 2015.** Typhoid fever. *Lancet* 385(9973):1136-
506 45.
507
508 **Zhang S, Yin Y, Jones MB, Zhang Z, Deatherage Kaiser BL, Dinsmore BA, Fitzgerald C, Fields PI, Deng X.**
509 **2015.** Salmonella serotype determination utilizing high-throughput genome sequencing data. *J Clin*
510 *Microbiol* 53(5):1685-92. doi: 10.1128/JCM.00323-15.

512 **Tables**
513 Table 1. Examples where two serovars belonged to the same eBG and the same ST
514

Serotype	Antigenic structure	ST
Bareilly Richmond	I 6,7: y: 1,2 I 6,7: y: 1,5	909
Saintpaul Haifa	I 4,5,12: e,h: 1,2 I 4,5,12: z,10: 1,2	49
Sandiego Brandenburg	I 4,12: l,v: e,n,z15 I 4,12: e,h: e,n,z15	20
Uganda Sinstorf	I 3, 10: l,z13: 1,5 I 3, 10: l,v: 1,5	684
Agona Essen	I 4,12: f,g,s:- I 4,12: f,g,m:-	13
Napoli Zaiman	I 1,9,12: l,z13: enx I 1,9,12: l,v: enx	P3141

515
516

517

518 Table 2. *S.* Java isolates in this study belonged to a diverse range of eBGs and STs associated with *S.* Java

519 whereas *S.* Paratyphi B belonged to ST86 only

	eBG5					eBG 9	eBG59	eBG32		eBG95
Phenotypic serovar	ST43	ST149	ST307	ST1577	ST86	ST88/127	ST28	ST423	ST682/1588	1583
Java	45	7	4	3	2*	18	6	5	2	1
Paratyphi B	0	0	0	0	12	0	0	0	0	0

521 **Figures**

522 Figure 1. Population structure of all *Salmonella enterica* isolates submitted to PHE from local and

523 regional hospital laboratories in England and Wales between April 2014 and March 2015 (see

524 Supplementary Table for details)

525

526 Figure 2a. Novel, preliminary STs (PST) and the modal number of isolates identified per PST

527

528 Figure 2b. The rate at which PSTs were identified throughout the time frame of the study

529

530 Figure 3a. Phylogenetic relationship of *S.* Richmond and *S.* Bareilly (ST909) (Figure 3a) and *S.* Saintpaul

531 and *S.* Haifa (ST49) (Figure 3b)

532

533 Figure 4. Serovars in lineage 3 mainly consist of multiple eBGs and are polyphyletic by nature

534

Figure 1(on next page)

Population Structure of *Salmonella enterica* submitted to PHE

Population structure of all *Salmonella enterica* isolates submitted to PHE from local and regional hospital laboratories in England and Wales between April 2014 and March 2015 (see Supplementary Table for details)
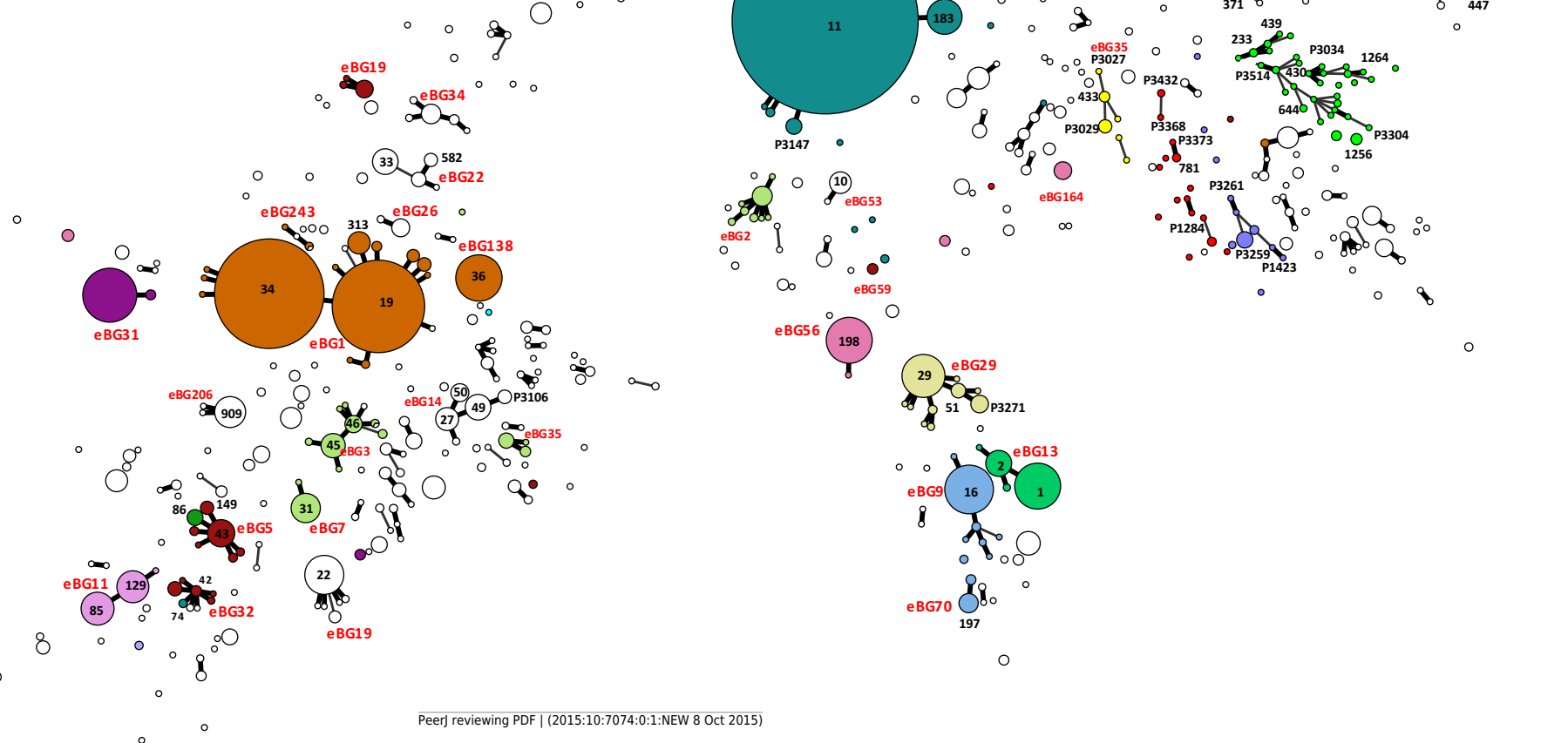
2

Trends in Preliminary Sequence Types

(A) Novel, preliminary STs (PST) and the modal number of isolates identified per PST (B) The rate at which PSTs were identified throughout the time frame of the study.

Figure 3(on next page)

Phylogenetic relationship of S. Richmond and S. Bareilly (ST909) (Figure 3a) and S. Saintpaul and S. Haifa (ST49) (Figure 3b)

H141920572 ['Salmonella Bareilly', '3.4.11.11.11.11.11']
H142600186 ['Salmonella Bareilly', '3.4.11.11.11.11.11']
94226_H15108042301-2 ['Salmonella Bareilly', '3.4.11.11.11.11.11']
90523_H15086058601-1 ['Salmonella Bareilly', '3.4.11.11.11.11.11']
H142600187 ['Salmonella Bareilly', '3.4.11.11.11.11.11']
H142740642 ['Salmonella Bareilly', '3.4.11.25.25.25.26']
H145180448 ['Salmonella Bareilly', '3.4.11.44.45.45.47']
H144340688 ['Salmonella Bareilly', '3.4.30.36.36.36.38']
H143120483 ['Salmonella Bareilly', '3.4.5.27.27.27.28']
H143140494 ['Salmonella Bareilly', '3.4.5.27.27.27.28']
H142920560 ['Salmonella Bareilly', '3.4.5.27.27.27.28']
5829_H14160010201-1 ['Salmonella Bareilly', '3.4.5.5.5.5.5']
5904_H14160010201-2 ['Salmonella Bareilly', '3.4.5.5.5.5.5']
H142860391 ['Salmonella Bareilly', '3.4.21.26.26.26.27']
H143540848 ['Salmonella Bareilly', '3.4.27.33.33.33.34']
H142540393 ['Salmonella Bareilly', '3.4.19.24.24.24.25']
H145060725 ['Salmonella Bareilly', '3.4.34.40.41.41.43']
H141980484 ['Salmonella Bareilly', '3.4.19.22.22.22.23']
H144020700 ['Salmonella Bareilly', '3.4.28.34.34.34.35']
H143340491 ['Salmonella Bareilly', '3.14.24.30.30.30.31']
H150380359 ['Salmonella Bareilly', '3.4.36.45.47.47.49']
94092_H15106049601-1 ['Salmonella Bareilly', '3.7.12.12.12.12.12']
H143340442 ['Salmonella Bareilly', '3.13.23.29.29.29.30']
H143400480 ['Salmonella Bareilly', '3.15.25.31.31.31.32']
H144420895 ['Salmonella Bareilly', '4.6.31.37.37.37.39']
H150600546 ['Salmonella Bareilly', '4.6.31.37.37.37.52']
81995_H15042076301-2 ['Salmonella Bareilly', '4.6.7.7.7.7.7']
99768_H15142058201-2 ['Salmonella Bareilly', '4.6.7.7.7.7.7']
95064_H15118035901-2 ['Salmonella Bareilly', '4.6.7.14.14.14.14']
95067_H15118036201-2 ['Salmonella Bareilly', '2.3.4.15.15.15.15']
5827_H14160009801-1 ['Salmonella Bareilly', '2.3.4.4.4.4.4']
99076_H15138047701-1 ['Salmonella Bareilly', '2.3.14.16.16.16.16']
84462_H15066055705-2 ['Salmonella Bareilly', '2.3.8.8.8.8.8']
H141600103 ['Salmonella Bareilly', '2.3.18.20.20.20.20']
H145120622 ['Salmonella Bareilly', '2.3.35.42.43.43.45']
1857_H14144056501-1 ['Salmonella Bareilly', '2.3.3.3.3.3.3']
86863_H15072060205-1 ['Salmonella Bareilly', '2.3.9.9.9.9.9']
86876_H15072060605-1 ['Salmonella Bareilly', '2.3.10.10.10.10.10']
H141741084 ['Salmonella Bareilly', '2.5.6.21.21.21.21']
H141860448 ['Salmonella Bareilly', '2.5.6.21.21.21.22']
6491_H14180049501-1 ['Salmonella Bareilly', '2.5.6.6.6.6.6']
1840_H14142042201-2 ['Salmonella Bareilly', '2.2.2.2.2.2.2']
H144880872 ['Salmonella Bareilly', '11.19.33.39.40.40.42']
H143320331 ['Salmonella Bareilly', '8.12.22.28.28.28.29']
H133100410 ['Salmonella Richmond', '6.9.16.18.18.18.18']
H122680540 ['Salmonella Richmond', '6.9.15.17.17.17.17']
H141540515 ['Salmonella Richmond', '6.10.17.19.19.19.19']
H144100501 ['Salmonella Richmond', '10.17.29.35.35.35.37']
H144100500 ['Salmonella Richmond', '10.17.29.35.35.35.36']
H145100685 ['Salmonella Richmond', '1.1.1.41.42.42.44']
ref
H150140561 ['Salmonella Richmond', '1.1.1.41.46.46.48']
117225_H15056065001-RPT-1 ['Salmonella Richmond', '1.1.1.1.1.1.1']
H142180431 ['Salmonella Richmond', '7.11.20.23.23.23.24']
95015_H15114064901-2 ['Salmonella Richmond', '5.8.13.13.13.13.13']
H144740697 ['Salmonella Richmond', '5.8.13.13.39.39.41']
H145140598 ['Salmonella Richmond', '5.8.13.43.44.44.46']
H150520680 ['Salmonella Richmond', '5.8.13.47.49.49.51']
H144620692 ['Salmonella Richmond', '5.18.32.38.38.38.40']
H143420593 ['Salmonella Richmond', '9.16.26.32.32.32.33']
H150520679 ['Salmonella Richmond', '12.20.37.46.48.48.50']

0.1

Figure 4(on next page)

Phylogenetic relationship of S. Richmond and S. Bareilly (ST909) (Figure 3a) and S. Saintpaul and S. Haifa (ST49) (Figure 3b)

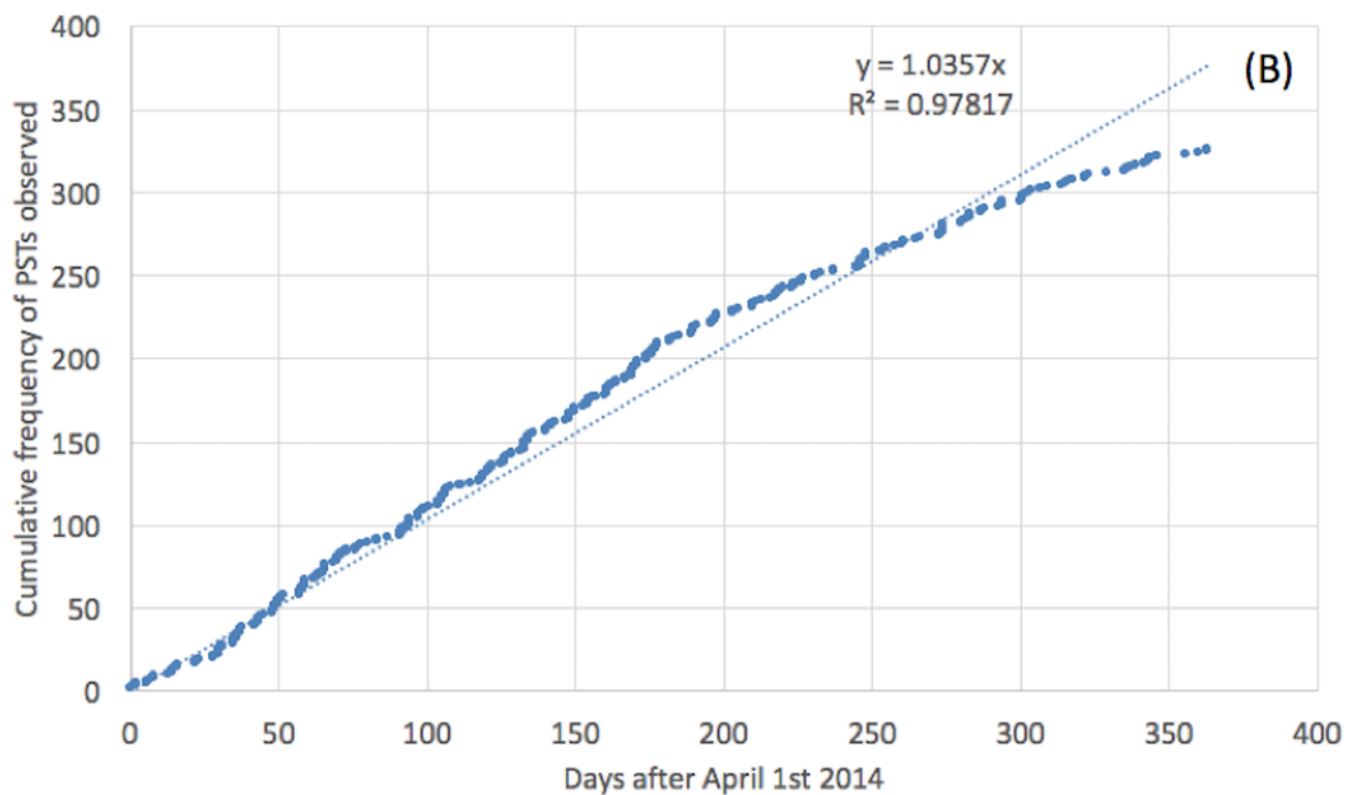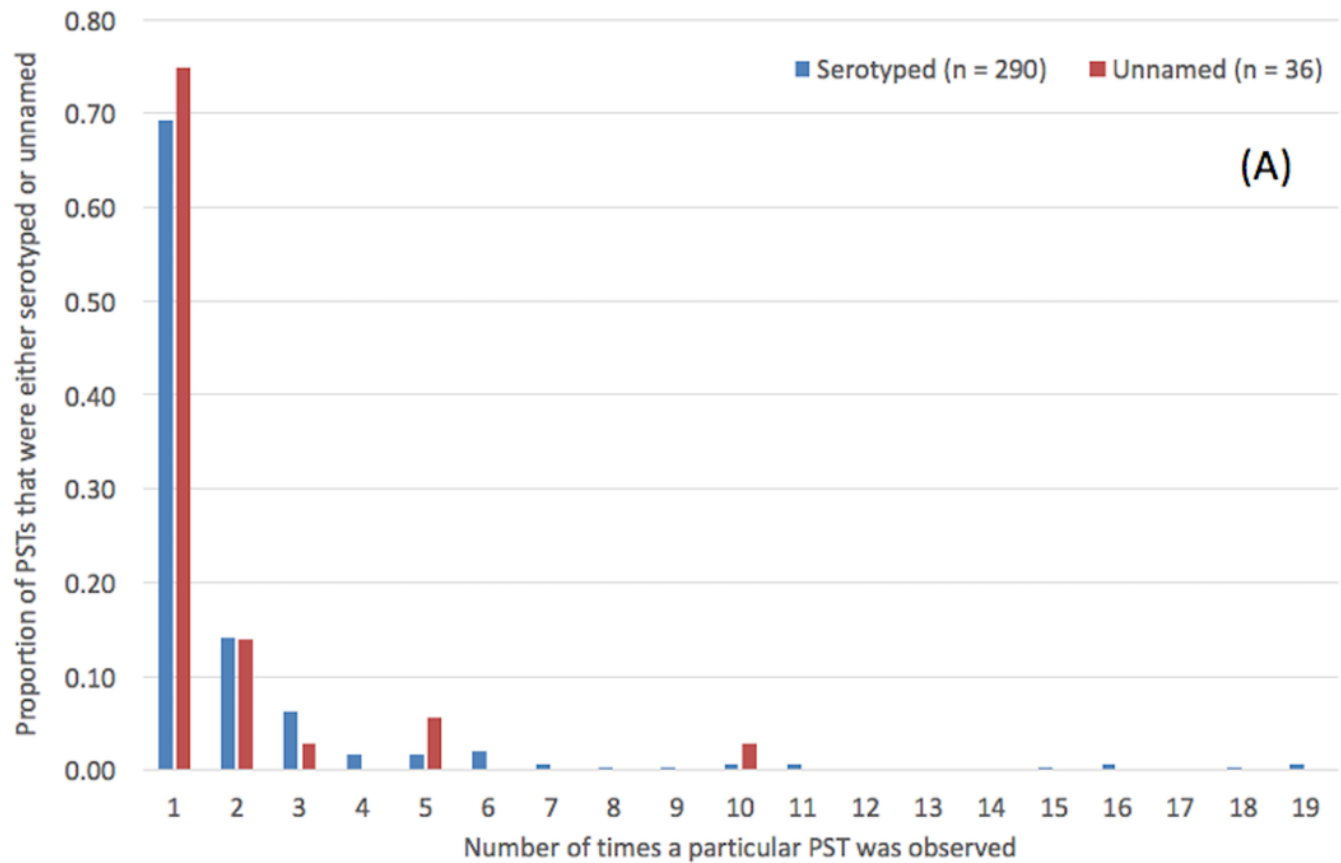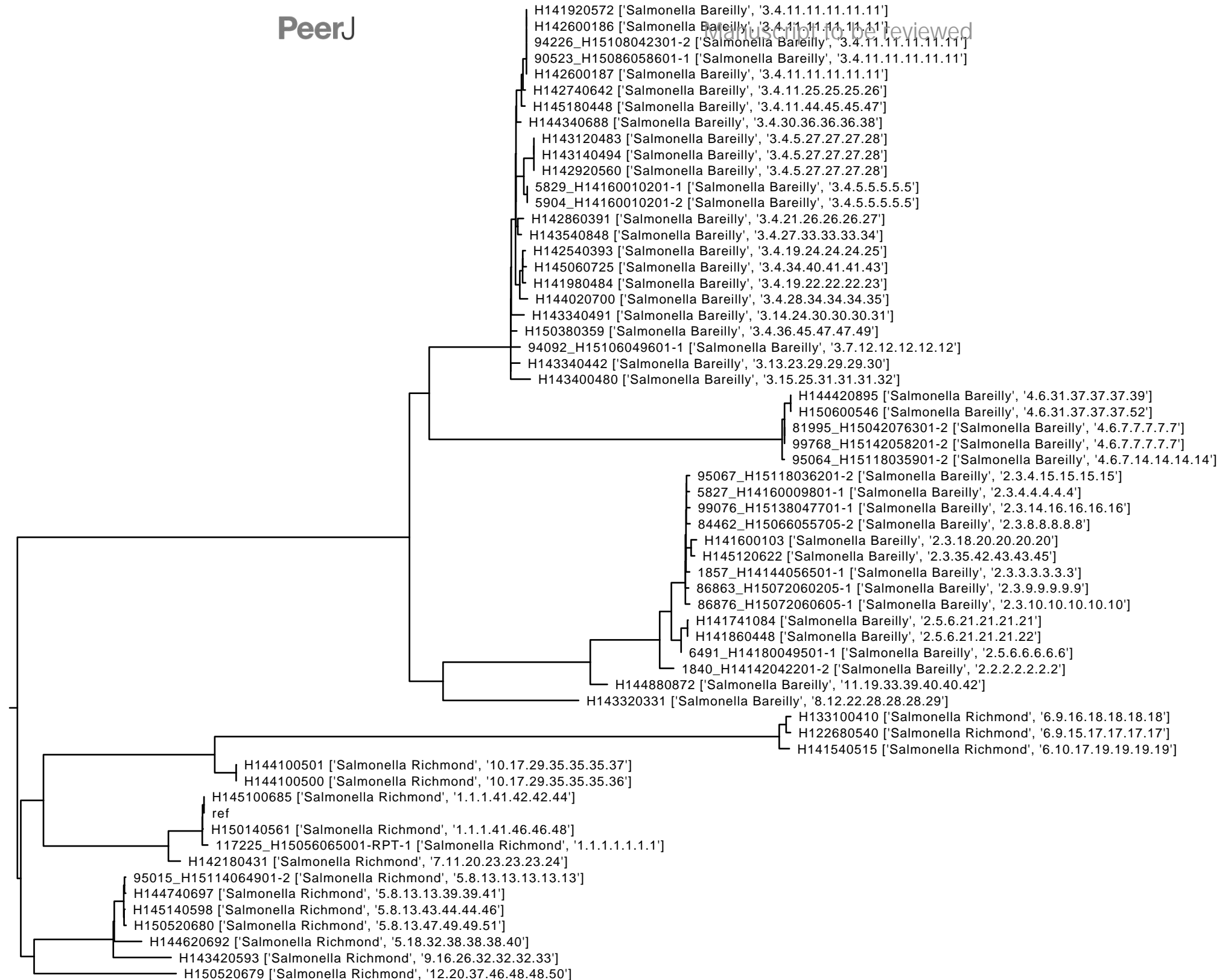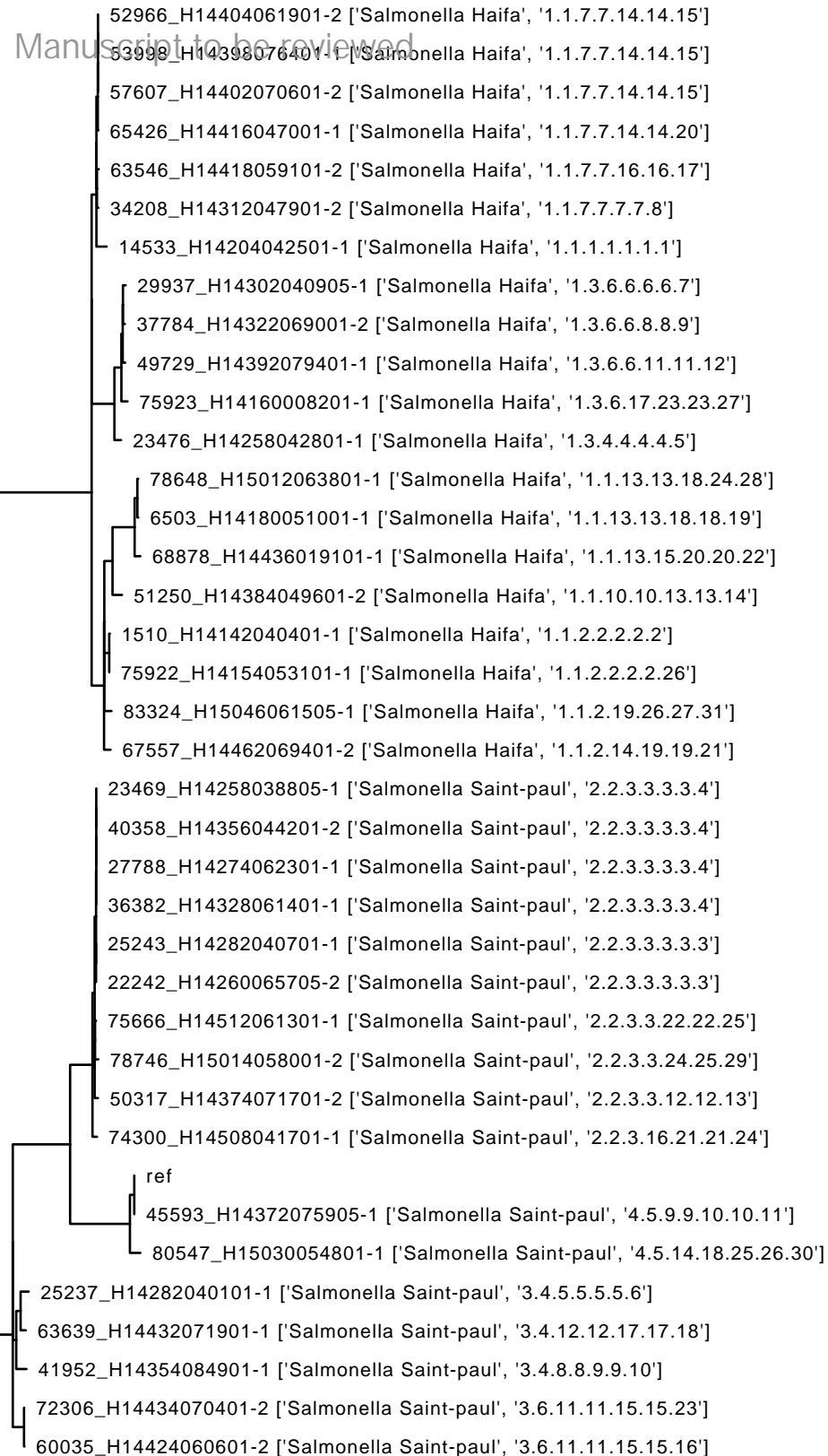52966_H14404061901-2 ['Salmonella Haifa', '1.1.7.7.14.14.15']
53998_H14398076401-1 ['Salmonella Haifa', '1.1.7.7.14.14.15']
57607_H14402070601-2 ['Salmonella Haifa', '1.1.7.7.14.14.15']
65426_H14416047001-1 ['Salmonella Haifa', '1.1.7.7.14.14.20']
63546_H14418059101-2 ['Salmonella Haifa', '1.1.7.7.16.16.17']
34208_H14312047901-2 ['Salmonella Haifa', '1.1.7.7.7.7.8']
14533_H14204042501-1 ['Salmonella Haifa', '1.1.1.1.1.1.1']
29937_H14302040905-1 ['Salmonella Haifa', '1.3.6.6.6.6.7']
37784_H14322069001-2 ['Salmonella Haifa', '1.3.6.6.8.8.9']
49729_H14392079401-1 ['Salmonella Haifa', '1.3.6.6.11.11.12']
75923_H14160008201-1 ['Salmonella Haifa', '1.3.6.17.23.23.27']
23476_H14258042801-1 ['Salmonella Haifa', '1.3.4.4.4.4.5']
78648_H15012063801-1 ['Salmonella Haifa', '1.1.13.13.18.24.28']
6503_H14180051001-1 ['Salmonella Haifa', '1.1.13.13.18.18.19']
68878_H14436019101-1 ['Salmonella Haifa', '1.1.13.15.20.20.22']
51250_H14384049601-2 ['Salmonella Haifa', '1.1.10.10.13.13.14']
1510_H14142040401-1 ['Salmonella Haifa', '1.1.2.2.2.2.2']
75922_H14154053101-1 ['Salmonella Haifa', '1.1.2.2.2.2.26']
83324_H15046061505-1 ['Salmonella Haifa', '1.1.2.19.26.27.31']
67557_H14462069401-2 ['Salmonella Haifa', '1.1.2.14.19.19.21']
23469_H14258038805-1 ['Salmonella Saint-paul', '2.2.3.3.3.3.4']
40358_H14356044201-2 ['Salmonella Saint-paul', '2.2.3.3.3.3.4']
27788_H14274062301-1 ['Salmonella Saint-paul', '2.2.3.3.3.3.4']
36382_H14328061401-1 ['Salmonella Saint-paul', '2.2.3.3.3.3.4']
25243_H14282040701-1 ['Salmonella Saint-paul', '2.2.3.3.3.3.3']
22242_H14260065705-2 ['Salmonella Saint-paul', '2.2.3.3.3.3.3']
75666_H14512061301-1 ['Salmonella Saint-paul', '2.2.3.3.22.22.25']
78746_H15014058001-2 ['Salmonella Saint-paul', '2.2.3.3.24.25.29']
50317_H14374071701-2 ['Salmonella Saint-paul', '2.2.3.3.12.12.13']
74300_H14508041701-1 ['Salmonella Saint-paul', '2.2.3.16.21.21.24']
ref
45593_H14372075905-1 ['Salmonella Saint-paul', '4.5.9.9.10.10.11']
80547_H15030054801-1 ['Salmonella Saint-paul', '4.5.14.18.25.26.30']
25237_H14282040101-1 ['Salmonella Saint-paul', '3.4.5.5.5.5.6']
63639_H144432071901-1 ['Salmonella Saint-paul', '3.4.12.12.17.17.18']
41952_H14354084901-1 ['Salmonella Saint-paul', '3.4.8.8.9.9.10']
72306_H14434070401-2 ['Salmonella Saint-paul', '3.6.11.11.15.15.23']
60035_H14424060601-2 ['Salmonella Saint-paul', '3.6.11.11.15.15.16']

0.2

5

Salmonella lineage 3 population structure

Serovars in lineage 3 mainly consist of multiple eBGs and are polyphyletic by nature