bioRxiv preprint doi: https://doi.org/10.1101/537654; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 **A real-time PCR for the differentiation of typhoidal and non-typhoidal**
2 ***Salmonella***
3

4 [1]Satheesh Nair*, [1&2]Vineet Patel, [1&2]Tadgh Hickey, [1]Clare Maguire, [1]David R Greig,
5 [1]Winnie Lee, [1]Gauri Godbole, [1]Kathie Grant and [1]Marie Anne Chattaway

6 [1]Gastrointestinal Bacteria Reference Unit, Public Health England, 61 Colindale
7 Avenue, London, NW8 5EQ

8 [2]University of Hertfordshire, School of Life and Medical Sciences, College Lane,
9 Hatfield, AL10 9AB, UK

10

11 *Corresponding Author

12 Satheesh Nair

13 Gastrointestinal Bacteria Reference Unit

14 Public Health England

15 61 Colindale Avenue

16 London, NW8 5EQ

17 satheesh.nair@phe.gov.uk

18

19 Keywords: Real-time PCR, Salmonella, Typhi, Paratyphi, WGS

20 Running title: Real-time PCR for typhoidal *Salmonella* detection

21

bioRxiv preprint doi: https://doi.org/10.1101/537654; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

22 **Abstract**

23 Rapid and accurate differentiation of *Salmonella* spp. causing enteric fever from non-
24 typhoidal *Salmonella* is essential for clinical management of cases, laboratory risk
25 management and implementation of public health measures. Current methods used
26 for confirmation of identification including biochemistry and serotyping as well as
27 whole genome sequencing analyses, takes several days. Here we report the
28 development and evaluation of a real-time PCR assay that can be performed directly
29 on crude DNA extracts from bacterial colonies, for the rapid identification of typhoidal
30 and non-typhoidal *Salmonella*.

31 This novel two-hour assay identifies the genus *Salmonella* by detecting the *ttr* gene,
32 encoding tetrathionate reductase, and defines typhoidal *Salmonella* by the detection
33 of *S.* Typhi and Paratyphi-specific gene combinations. PCR assay performance was
34 determined using 211 clinical cultures of *Salmonella* (114 non-typhoidal and 97
35 Typhoidal strains) and 7 clinical non-*Salmonella* cultures. In addition, the specificity
36 of the assay was evaluated *in silico* using a diverse in-house collection of 1882
37 *Salmonella* whole genome sequences. The real-time PCR results for 218 isolates
38 and the genomic analysis of the 1882 isolates produced 100% sensitivity and 100%
39 specificity (based on a 7 gene profile) for identifying typhoidal *Salmonella* compared
40 to the *Salmonella* whole genome sequening identification methods currently used at
41 Public Health England.

42 This paper describes a robust real-time PCR assay for the rapid, accurate
43 identification of typhoidal and non-typhoidal *Salmonella* which will be invaluable for
44 the urgent screening of isolates from symptomatic individuals, the safe processing of
45 isolates in laboratories and for assisting the management of public health risks.

bioRxiv preprint doi: https://doi.org/10.1101/537654; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

46 **Introduction**

47 *Salmonella* are a diverse genus of gastrointestinal pathogens that cause a wide
48 spectrum of disease from self-limiting gastroenteritis (non-typhoidal salmonellae,
49 NTS) to systemic enteric fever (typhoidal salmonellae - *Salmonella enterica* serovar
50 Typhi, Paratyphi A, B and C). Salmonellosis is global but typhoidal *Salmonella* are
51 found mainly in sub-Saharan Africa and South Asia where enteric fever is endemic
52 (1); although the detailed local surveillance data from endemic regions remains poor
53 (2). A current concern is the increase in bacteraemia (and focal infections)
54 associated with multi-drug resistant NTS infection in sub-Saharan Africa. In high
55 income countries such as the UK, invasive NTS infection is mainly confined to
56 immune-compromised hosts and so the major risks are local outbreaks of NTS
57 through poor food hygiene and typhoidal infections associated with travel to endemic
58 regions.

59 Diagnostic hospital microbiology laboratories make only a presumptive identification
60 of *Salmonella* spp.: they do not usually hold a sufficient range of specific antisera for
61 full identification and rapid identification systems, such as Matrix Assisted Laser
62 Desorption/Ionisation-Time of Flight, Mass Spectroscopy, are unable to fully speciate
63 *Salmonella*. In reference laboratories where definitive microbiological methods for
64 the identification of *Salmonella* by serology and biochemistry (3) do exist, the
65 turnaround times are often lengthy because of weak expression of the somatic (O),
66 flagellar (H) and Vi polysaccharide surface antigens leading to incomplete or
67 incorrect identification of the serovars. Whole Genome Sequencing (WGS) for
68 *Salmonella* (4) has simplified the process for identifying *Salmonella* serovars
69 substantially but still takes days rather than the hours. Currently there are no rapid
70 diagnostic tests for informing clinical and public health management of enteric fever

bioRxiv preprint doi: https://doi.org/10.1101/537654; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

71 or for ensuring *Salmonella* isolates are processed appropriately with respect to
72 laboratory safety.

73 In the UK, salmonellosis is a significant public health problem causing morbidity,
74 financial loss due to sickness and absenteeism until clearance from infection for
75 certain professions. The clinical management of salmonellosis patients depends on
76 diagnosis. Enteric fever is treated with antibiotics but non-typhoidal *Salmonella*
77 gastroenteritis is usually self-limiting. Invasive disease needs to be treated with
78 antibiotics specific to the strain causing infection. In addition, the processing of
79 isolates or specimens in the laboratory from patients with suspected diarrhoeal
80 infection depends on the identification of the causal agent. In the UK microorganisms
81 that pose a risk to human health are classified into one of four hazard groups based
82 on their ability to infect healthy humans. The classification of these organisms allows
83 the risk they pose to laboratory and healthcare workers to be controlled by
84 implementing safety measures proscribed by law. *S.* Typhi, *S*. Paratyphi A, B and C
85 are classified as a Hazard Group 3 (HG3) pathogens requiring processing in a
86 specialised containment level 3 (CL3) laboratory (5). It is clear, therefore, that in
87 order to treat patients effectively and protect healthcare and laboratory staff, the
88 rapid identification of a patient as being infected with a typhoidal salmonella is
89 critical.

90 At present there is no single rapid method to identify all typhoidal (HG3 in the UK)
91 *Salmonella*, even though genomic data on the presence or absence of genes in both
92 typhoidal and non-typhoidal *Salmonella* are in abundance. The *ttr* gene, encoding
93 tetrathionate reductase has been used as a PCR gene target to detect and identify
94 *Salmonella* since it is present in all *Salmonella* spp. (6). However, it is not intended
95 to distinguish typhoid and non-typhoidal subspecies. A few potential candidate

bioRxiv preprint doi: https://doi.org/10.1101/537654; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

96 genes for identifying HG3 *Salmonella* have been described previously. For example,
97 the *tviB* gene, encoding a Vi polysaccharide capsule, which is present in *S.* Typhi
98 and *S.* Paratyphi C, (7) but not in *S.* Paratyphi A or *S.* Paratyphi B, can identify a
99 subset of typhoidal *Salmonella* but doesn't distinguish *S.* Typhi or *S.* Paratyphi C.
100 (8). In order to differentiate *Salmonella* serovars causing enteric fever, additional
101 genes are required. Nga *et al* (2010) proposed using *SPA2308*, encoding a
102 hypothetical protein, for the detection of *S.* Paratyphi A and STY0201 (also known as
103 the *staG* gene), encoding a putative fimbrial protein, for the detection of *S*. Typhi in
104 clinical blood samples via PCR (9). Connor *et al* (2016) suggested that *S.* Paratyphi
105 B (HG3) could be distinguished from *S.* Java (HG2) using two genes encoding Type
106 III Secretion System (TTSS) effector proteins; *sseJ* and *srfJ* (10): with *S.* Paratyphi
107 B possessing only *srfJ* but *S.* Java possessing both *sseJ* and *srfJ*. However, as *sseJ*
108 is also absent in *S.* Typhi and *S.* Paratyphi A, this gene cannot be used to
109 differentiate all HG3 serovars or used alone as an HG2 marker. A potential gene
110 target for *S*. Paratyphi C identification is the *SPC0869* target, a gene encoding a
111 hypothetical protein, shown to be present only in *S*. Paratyphi C (8). The use of this
112 gene requires further assessment to ensure it is a unique target amongst the *S.*
113 Paratyphi C population as only five serovars were investigated in the study by Lui *et*
114 *al*., 2009.

115 The design of a PCR assay to identify *Salmonella* and differentiate HG2 and HG3
116 *Salmonella* requires a multi-targeted approach with defined gene profiles and
117 rigorous validation. The aim of this study was to develop and validate a real-time
118 PCR assay to distinguish HG3 (Typhoidal) and HG2 (Non-typhoidal) *Salmonella* and
119 identify specific serovars of HG3 *Salmonella*.

bioRxiv preprint doi: https://doi.org/10.1101/537654; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

120 ## Methodology

121 ### Bacterial strains

122 A total of 211 *Salmonella enterica subsp* I isolates, received at the Gastrointestinal
123 Bacterial Reference Unit (GBRU), Public Health England (PHE) between 2008 -
124 2017, (Table 1a) were used in this PCR study. Representative HG2 isolates from
125 the two most common serovars, *S.* Enteritidis and *S.* Typhimurium, as well as
126 serovars that can be difficult to distinguish from HG3 isolates by traditional methods,
127 including *S.* Dublin, *S.* Java and *S.* Choleraesuis, were selected (Table 1a). Assay
128 specificity was further investigated by the inclusion of four *Shigella* isolates (*S.*
129 *flexneri, S. sonnei, S. dysenteriae, S. boydii*) and three *E. coli* isolates (containing
130 either *eae* or *stx* genes) as representatives to test the specificity against other
131 *Enterobacteriaceae* that are occasionally misidentified by referring clinical
132 laboratories using automated identification platforms (Supp table 1).

133 ### *Salmonella* whole genome sequence data

134 1882 *Salmonella* whole genome sequences (including the 211 *Salmonella* isolates),
135 representing the diversity of Salmonellae tested by GBRU, were included in an *in*
136 *silico* validation of the specificity of the selected target genes (Figure 1). This dataset
137 included representative sequence types (ST) of the 19,221 strains validated and
138 reported at GBRU between 2016-2017. The strains selected included all sub-species
139 of *Salmonella* and the common (3 or more isolates received between 2016-2017 at
140 PHE) *Salmonella* Serovars, representing in total 477 different sequence types (Table
141 1b, Supp table 1).

bioRxiv preprint doi: https://doi.org/10.1101/537654; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

142 **DNA extraction and real-time PCR assays**

143 DNA from 218 isolates was extracted via a crude extraction method in which a single
144 colony from MacConkey agar [ThermoFisher Scientific, Waltham, USA] was
145 inoculated into 490 µL of sterile distilled water in a screw cap microtube [Eppendorf,
146 Hamburg, Germany] and placed in a boiling water bath for 20 minutes. Primers and
147 probes for *ttr* (detection of all *Salmonella*), *tviB* (detection of *S.* Typhi and *S.*
148 Paratyphi C), *SPA2308* (detection of *S.* Paratyphi A) and *staG* (detection of *S.* Typhi)
149 were based on previous published studies (Table 2). Primers and probes for
150 *SPC0869* (detection of *S.* Paratyphi C) *sseJ* and *srfJ* (detection of *S.* Paratyphi B)
151 were designed using the PrimerQuest Tool V8
152 (https://www.idtdna.com/PrimerQuest/Home/Index) using sequences obtained from
153 the NCBI nucleotide database (https://www.ncbi.nlm.nih.gov/nucleotide/) (Table 2).

154 The real-time PCR reported here was carried out as seven individual monoplex
155 reactions but also worked as duplex and triplex PCR assays with interchangeable
156 primers and probes targets (and probe dyes) depending on the target gene
157 combination required. Mastermix for the monoplex assay consisted of 12.5 µL
158 Takyon Low Rox probe mastermix [Eurogentec, Liège, Belgium], 8 µL Nuclease free
159 water, 0.5 µL each of 20 µM forward and reverse primers, 1 µL of 5 µM probe and
160 2.5 µL DNA to a final reaction volume of 25 µL. A negative control was run with each
161 PCR using 2.5 µL nuclease free water for the template [Severn Biotech,
162 Kidderminster, UK] and the following positive controls were used: NCTC 8385 – *S.*
163 Typhi (*ttr, tviB, staG*) NCTC 11803 - *S.* Paratyphi A (*ttr, SPA2308*), NCTC 8299 - *S.*
164 Paratyphi B (*ttr, srfj*), NCTC 96 – *S.* Paratyphi C (*ttr, sseJ, tviB, SPC0869),* NCTC
165 6676 – *S.* Enteritidis (*ttr, sseJ*) and NCTC 14013 – *S.* Typhimurium (*ttr, sseJ, srfJ*).
166 The PCR was run on the ABI Prism 7500 Real-Time PCR System [Applied

bioRxiv preprint doi: https://doi.org/10.1101/537654; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

167 Biosystems, Foster City, USA]. The conditions for the PCR were an initial activation
168 of 95°C for 3 minutes, followed by 40 cycles of: Denaturation at 95°C for 30 seconds,
169 Annealing at 60°C for 30 seconds, Extension at 72°C for 10 seconds. A positive
170 result was assigned when a Ct value was achieved between 12-30 with a threshold
171 set at 0.03ΔR.

172 Identification of HG3 *Salmonella* and differentiation from HG2 *Salmonella* was based
173 on a profile of seven genes (Table 3). The molecular and/or *in silico* PCR
174 identification was compared with the original identification of the serovar obtained via
175 a combination of WGS identification, phenotype and serology carried out by the
176 *Salmonella* laboratory as described previously (4) (Supp table 1).

177 **PCR assay evaluation**

178 The sensitivity and specificity of the *ttr, sseJ, srfJ, tviB, staG, SPA2308* and
179 *SPC0869* primers and probes (Table 3) used in the real-time PCR assays were
180 calculated according to Martin, 1984 (11).

181 In addition, PCR assay specificity was assessed by *in silico* genomic analysis using
182 a diverse in-house WGS dataset covering the population structure of *Salmonella*
183 (Figure 1). A total of 1882 *Salmonella* sequences (Supp table 1, Figure 1) which
184 includes the 211 Salmonella isolates tested by PCR were screened for the presence
185 of seven target genes (*ttr, sseJ, srfJ, tviB, staG, SPA2308* and *SPC0869)* using a
186 PHE in-house bioinformatics tool called GeneFinder (developed by Doumith M, *et al,*
187 unpublished). This tool takes paired-end Illumina FASTQ reads and aligns them to a
188 reference sequence of the target genes, as a multi-FASTA file, (Accession numbers
189 in table 2) using Bowtie2 v2.1.0 (12) and Samtools v1.0.18 (13) and determines
190 metrics such as coverage, presence of indels (an insertion or deletion), amino acid

bioRxiv preprint doi: https://doi.org/10.1101/537654; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

191 alterations, presence of single nucleotide polymorphisms and overall sequence
192 similarity of the test sequence to the reference gene sequence. Target genes were
193 designated as present when sequences achieved a detection threshold of 80%
194 sequence similarity to the reference gene, apart from *ttr* where the threshold was set
195 at 70% sequence similarity, due to the size and variability of this particular gene. Any
196 discrepant results between GeneFinder and the PCR were investigated further by
197 assembling the sequence data using Spades v3.1.1 to default parameters and
198 examining the variability of primer and probe binding sites.

199 Assay reproducibility was determined by testing 20 of the 211 *Salmonella* isolates in
200 triplicate. Precision was evaluated by the standard deviation of Ct values of n=10
201 replicates of each of the positive controls for each target. Each target was assessed
202 individually and as a multiplex in separate assay runs by different individuals and
203 had the threshold set at 25% of the maximal fluorescence (ΔR) of each respective
204 target.

205 **Results**

206 **Comparison of real-time PCR and current PHE methods for distinguishing HG2**
207 **and HG2 Salmonellae**

208 Of the 211 *Salmonella* isolates subjected to PCR identification, all gave the expected
209 gene profile identification (Table 4, Supp table 1), matching the original identification,
210 except for three *S*. Typhi isolates where the *tviB* gene was not detected. This was
211 confirmed by *in silico* analysis (see below). Previous described 'HG3' gene targets
212 *SPA0869, staG and SPA2308* were found sporadically in 41/114 (35%) of the HG2
213 *Salmonella* tested (two isolates had two HG3 gene targets present) confirming that
214 use of single targets to differentiate HG3 from HG2 *Salmonella* is not appropriate

bioRxiv preprint doi: https://doi.org/10.1101/537654; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

215 (Table 5, Supp table 1). None of the 7 target genes were detected in the four
216 *Shigellae* and three *E. coli* isolates that were tested.

217 **Whole genome sequencing *in silico* analysis**

218 Of the 1882 Salmonella analysis subjected to *in silico* analysis, identification based
219 on gene profiles (Table 3) matched the original identification but did highlight that
220 individual gene targets could be found sporadically across the *Salmonella*
221 population. *In silico* analysis identified 952/1882 non-typhoidal *Salmonella* isolates
222 that were positive for *ttr* and a combination of other 'HG3' gene targets (Table 5,
223 Supp table 1), designated as profile 5 (Table 4). None of the gene profiles of these
224 isolates matched the designated HG3 profiles (profile 1-4) and thus our interpretation
225 is that *ttr* positive strains with a profile not matching the HG3 profiles should be
226 classified as HG2 *Salmonella* (Table 3, Table 4, Supp table 1).

227 As with the real-time PCR assay, the *Salmonella* processed via *in silico* analysis
228 identified the three SPI-7 negative *S.* Typhi isolates. The real-time PCR and
229 GeneFinder correctly identified the deletion of this gene.

230 In this study 8 of the 1882 sequences were positive by PCR and yet negative for the
231 same gene by GeneFinder. Further *in silico* analysis revealed that the genes
232 concerned had an intact primer and probe binding site, thus confirming the PCR
233 result but variation outside of these regions resulted in average similarity values
234 below the GeneFinder threshold value (Supp table 1).

235 **Reproducibility and precision of PCR assay**

236 Reproducibility was assessed by performing the PCR 3 times on 20 isolates. Results
237 indicated that the PCR was reproducible for differentiating between HG2 and HG3

bioRxiv preprint doi: https://doi.org/10.1101/537654; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

238 salmonellae and for the identification of serovars within HG3 *Salmonella* (Supp table
239 1). The precision analysis demonstrated that five out of seven of the gene targets
240 were considered precise (i.e. standard deviation <0.167). The following results show
241 the gene, average Ct (and standard deviation): *ttr* - 25.12 (0.154), *sseJ* - 23.59
242 (0.127), *SrfJ* - 24.51 (0.179), *tviB* - 25.01 (0.115), *StaG* - 24.97 (0.121), *SPC0869* -
243 25.68 (0.142) and *SPA2308* - 20.59 (0.248). Both *SrfJ* and *SPA2308* have standard
244 deviations above the 0.167 value that is considered precise. The explanation for this
245 is that these two primer/probe sets are more susceptible to variation due to the *SrfJ*
246 reverse primer having no G/C's in the GC clamp therefore increasing the possibility
247 of variable binding to the target gene. The *SPA2308* forward primer has less than
248 40% GC content making it more thermally variable and both reverse primer and
249 probe's do self-anneal and form hairpins. This is the case as the *SPA2308* gene has
250 a very low GC content of 32.25% and as a result will lead to more variable results.
251 Another important note is that this validation process occurred using boiled cells as
252 the DNA extraction method (as this is the intended use for rapidity) and there is
253 always the possibility of slight levels of PCR inhibition, in comparison to using
254 purified DNA, which will also affect the precision results. The lower precision levels
255 did not affect the molecular PCR in practice and was deemed suitable for use.

256 Reproducibility was not affected when targets were tested as a multiplex assay,
257 however the precision assay in the molecular multiplex PCR proved to be better than
258 the molecular monoplex reactions (Supp data 1).

259 **Sensitivity and specificity**

260 Sensitivity and specificity were based on the 7 gene profiles (and not individual gene
261 markers) detected by real-time PCR and GeneFinder (Table 4). It showed 100%

bioRxiv preprint doi: https://doi.org/10.1101/537654; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

262 sensitivity and specificity for the detection of HG3 *Salmonella* as compared to the
263 routine reference identification by WGS and serotyping.

264 **Discussion**

265 This study describes for the first time a robust real-time PCR assay for the specific
266 identification of each of the four typhoidal *Salmonella* serovars: *S.* Typhi and *S.*
267 Paratyphi A, B and C and is 100% reliable (Figure 1, Table 4, Table 5, Supp table 1).
268 This assay was validated as a monoplex PCR providing the flexibility to use
269 individual targets of interest but the assay was also found to work equally well as a
270 multiplex assay (Supp data 1) and is now in use routinely at PHE. The rapid
271 turnaround time of this PCR assay has potential for expediting the management of
272 suspected cases of typhoid fever. With additional optimisation, the application of this
273 assay could be extended to direct testing of clinical specimens (blood and stool) as
274 well as food, water and environmental specimens. This would further increase the
275 value of the assay although such use may risk the possibility of less isolates being
276 referred to reference laboratories for further characterisation leading to loss of typing
277 for surveillance purposes, including antimicrobial resistance monitoring, as well as
278 outbreak detection and investigation. Thus, it is essential that isolates continue to be
279 isolated and referred to reference laboratories.

280 Many assays for identifying typhoidal Salmonellae have been described previously
281 but these are usually single gene methods with much lower specificity and sensitivity
282 or are aimed at just one or two of the typhoidal serovars (9, 14-16). However, these
283 important studies have provided input for the selection of candidate gene targets in
284 designing a gene profile-based PCR assay, the validation of this PCR assay was

bioRxiv preprint doi: https://doi.org/10.1101/537654; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

285 strengthened by the use of WGS sequence data for high-through put testing on a
286 more diverse collection of *Salmonella*.

287 *In silico* analysis has its limitations if relying on this approach as a sole method.
288 Although PHE utilise a multilocus sequence type (MLST) based approach with
289 genomic data for *Salmonella* identification (4), other organisations may use a gene
290 based approach for *Salmonella* identification, the use of set thresholds with *in silico*
291 testing in the current study on target genes (i.e. at what threshold is the test positive)
292 may need to be flexible depending on the gene. Unlike detection via PCR, the entire
293 target gene is evaluated using *in silico* analysis and therefore we can draw
294 conclusions on the presence/absence of the target gene. However, selecting a
295 threshold value (and therefore a percentage identify of a match) to which a gene is
296 considered present or absent can be difficult. Discrepancies between, real-time PCR
297 and genomic detection of target genes occur when a gene has less than the set
298 threshold of sequence similarity. There were initially eight negative gene results
299 using GeneFinder that were positive by PCR. These were due to a lower percentage
300 of gene similarity and below the 80% set threshold (Supp table 1) and were positive
301 for the presence of the gene (matching the PCR result). When mismatches between
302 PCR and *in silico* methods occur, explicit consideration is required to ascertain if the
303 PCR primer/probe binding region is intact and how much of the gene is present.
304 Specifically, in our targets, *ttr* showed a large range of variability amongst isolates in
305 terms of sequence similarity to the reference gene with five of eight of these samples
306 having *ttr* <80% sequence similarity. After assessing the primer/probe binding sites
307 of the genes, there were no discrepancies between GeneFinder and the PCR assay.

308 This current study showed that 17 of the 952 NTS isolates were only positive for a
309 single gene target (*ttr* gene) (Supp table 1) and belonged to *Salmonella* subspp. III,

bioRxiv preprint doi: https://doi.org/10.1101/537654; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

310 IV and V. Therefore, most NTS *Salmonella* contain one or more of the other genes
311 markers normally associated with typhoidal *Salmonella* (Table 5). This highlights that
312 a single gene target method is not appropriate for distinguishing between typhoidal
313 and non-typhoidal *Salmonella,* with a gene profile-based method being more
314 accurate for identification and differentiation of typhoidal *Salmonella*. The reassuring
315 finding, however, is that not one of the 935 NTS *Salmonella* had the same gene
316 profile as the typhoidal (HG3) *Salmonella* profiles (Supp table 1, Table 3, Table 4).

317 Another notable observation is that three *S.* Typhi isolates from Pakistan lacked the
318 134kb SPI-7 pathogenicity island harbouring the *ViaB* operon (*tvi* genes – associated
319 with the production of the Vi capsule). Although rare, absence of SPI-7 pathogenicity
320 island including the *tvi* region in *S*. Typhi has previously been described (7). This is
321 potentially an important public health finding as the current typhoid Vi polysaccharide
322 vaccine stimulates immunity against the Vi capsule. It is known that SPI-7 negative
323 (Vi-negative) *S.* Typhi can cause typhoid fever (17) and so there is a need to monitor
324 the loss of the SPI-7 island in endemic regions where *S*. Typhi vaccination programs
325 are being conducted (17). The assay described here could be used to monitor the
326 emergence of Vi-negative *S.* Typhi through the emergence of *ttr* and *staG* positive
327 *tviB* negative strains.

328 Conclusion

329 In conclusion, this is the first real-time PCR assay that can rapidly distinguish
330 between typhoidal ie *S*. Typhi, Paratyphi A, Paratyphi B and Paratyphi C (HG3) and
331 non-typhoidal (HG2) *Salmonella* serovars. The assay has the ability to be
332 implemented in diagnostic and reference laboratories globally as a safe and cost-
333 effective way of differentiating *Salmonella*.

bioRxiv preprint doi: https://doi.org/10.1101/537654; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

334 **Funding**

335 This study was funded by PHE

336 **Acknowledgements**

337 Thank you to Sarah Alexandra and Julie Russell from the National Collection of Type
338 Cultures for providing positive control strains.

339 Thank you to Lailanie Aqunino from GBRU for support in undertaking PCR.

340 **Conflict of interest**

341 Nil

342 **References**

343 1. Crump JA, Sjolund-Karlsson M, Gordon MA, Parry CM. 2015.
344 Epidemiology, Clinical Presentation, Laboratory Diagnosis,
345 Antimicrobial Resistance, and Antimicrobial Management of Invasive
346 Salmonella Infections. Clin Microbiol Rev 28:901-37.
347 2. Wain J, Hendriksen RS, Mikoleit ML, Keddy KH, Ochiai RL. 2015.
348 Typhoid fever. The Lancet 385:1136-1145.
349 3. Bale J, Meunier D, Weill FX, dePinna E, Peters T, Nair S. 2016.
350 Characterization of new Salmonella serovars by whole-genome
351 sequencing and traditional typing techniques. J Med Microbiol
352 65:1074-1078.
353 4. Ashton PM, Nair S, Peters TM, Bale JA, Powell DG, Painset A, Tewolde
354 R, Schaefer U, Jenkins C, Dallman TJ, de Pinna EM, Grant KA,
355 Salmonella Whole Genome Sequencing Implementation G. 2016.
356 Identification of Salmonella for public health surveillance using
357 whole genome sequencing. PeerJ 4:e1752.
358 5. Executive HaS. 2013. The Approved List of biological agents.
359 6. Hopkins KL, Peters TM, Lawson AJ, Owen RJ. 2009. Rapid
360 identification of Salmonella enterica subsp. arizonae and S. enterica
361 subsp. diarizonae by real-time polymerase chain reaction. Diagn
362 Microbiol Infect Dis 64:452-4.
363 7. Nair S, Alokam S, Kothapalli S, Porwollik S, Proctor E, Choy C,
364 McClelland M, Liu SL, Sanderson KE. 2004. Salmonella enterica

bioRxiv preprint doi: https://doi.org/10.1101/537654; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

365 serovar Typhi strains from which SPI7, a 134-kilobase island with
366 genes for Vi exopolysaccharide and other functions, has been deleted.
367 J Bacteriol 186:3214-23.
368 8. Liu WQ, Feng Y, Wang Y, Zou QH, Chen F, Guo JT, Peng YH, Jin Y, Li YG,
369 Hu SN, Johnston RN, Liu GR, Liu SL. 2009. Salmonella paratyphi C:
370 genetic divergence from Salmonella choleraesuis and pathogenic
371 convergence with Salmonella typhi. PLoS One 4:e4510.
372 9. Nga TV, Karkey A, Dongol S, Thuy HN, Dunstan S, Holt K, Tu le TP,
373 Campbell JI, Chau TT, Chau NV, Arjyal A, Koirala S, Basnyat B, Dolecek
374 C, Farrar J, Baker S. 2010. The sensitivity of real-time PCR
375 amplification targeting invasive Salmonella serovars in biological
376 specimens. BMC Infect Dis 10:125.
377 10. Connor TR, Owen SV, Langridge G, Connell S, Nair S, Reuter S,
378 Dallman TJ, Corander J, Tabing KC, Le Hello S, Fookes M, Doublet B,
379 Zhou Z, Feltwell T, Ellington MJ, Herrera S, Gilmour M, Cloeckaert A,
380 Achtman M, Parkhill J, Wain J, De Pinna E, Weill FX, Peters T,
381 Thomson N. 2016. What's in a Name? Species-Wide Whole-Genome
382 Sequencing Resolves Invasive and Noninvasive Lineages of
383 Salmonella enterica Serotype Paratyphi B. MBio 7.
384 11. Martin SW. 1984. Estimating disease prevalence and the
385 interpretation of screening. Preventive Veterinary Medicine 2:463-
386 472.
387 12. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with
388 Bowtie 2. Nat Methods 9:357-9.
389 13. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G,
390 Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The
391 Sequence Alignment/Map format and SAMtools. Bioinformatics
392 25:2078-9.
393 14. Hirose K, Itoh K, Nakajima H, Kurazono T, Yamaguchi M, Moriya K,
394 Ezaki T, Kawamura Y, Tamura K, Watanabe H. 2002. Selective
395 amplification of tyv (rfbE), prt (rfbS), viaB, and fliC genes by
396 multiplex PCR for identification of Salmonella enterica serovars
397 Typhi and Paratyphi A. J Clin Microbiol 40:633-6.
398 15. Levy H, Diallo S, Tennant SM, Livio S, Sow SO, Tapia M, Fields PI,
399 Mikoleit M, Tamboura B, Kotloff KL, Lagos R, Nataro JP, Galen JE,
400 Levine MM. 2008. PCR method to identify Salmonella enterica
401 serovars Typhi, Paratyphi A, and Paratyphi B among Salmonella
402 Isolates from the blood of patients with clinical enteric fever. J Clin
403 Microbiol 46:1861-6.
404 16. Tennant SM, Toema D, Qamar F, Iqbal N, Boyd MA, Marshall JM,
405 Blackwelder WC, Wu Y, Quadri F, Khan A, Aziz F, Ahmad K, Kalam A,
406 Asif E, Qureshi S, Khan E, Zaidi AK, Levine MM. 2015. Detection of

bioRxiv preprint doi: https://doi.org/10.1101/537654; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

407 Typhoidal and Paratyphoidal Salmonella in Blood by Real-time
408 Polymerase Chain Reaction. Clin Infect Dis 61 Suppl 4:S241-50.
409 17. Baker S, Sarwar Y, Aziz H, Haque A, Ali A, Dougan G, Wain J, Haque A.
410 2005. Detection of Vi-negative Salmonella enterica serovar typhi in
411 the peripheral blood of patients with typhoid fever in the Faisalabad
412 region of Pakistan. J Clin Microbiol 43:4418-25.

413

bioRxiv preprint doi: https://doi.org/10.1101/537654; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

414 **Tables and Figures**

415 **Table 1**: Number and type of *Salmonella* serovars tested via molecular PCR and by
416 GeneFinder.
417 (a) Via PCR

418 Footnote: *A random selection of HG2 ST containing sporadic gene targets were
419 chosen. EPEC: Enteropaothogenic *E. coli.* STEC: Shiga toxin-producing *E. coli.*

420 (b) Via GeneFinder

421

422 **Table 2:** Primer and probe sequences used for each gene target with the fluorescent
423 dye coloured (Colour of reporter related to spectrum of detection) and quenchers in
424 bold (BHQ- black hole quencher).
425 Footnote: Note – F – forward primer, R – reverse primer, P- probe.

426

427 **Table 3:** Gene profiles for the identification of *S*. Typhi and *S*. Paratyphi from other
428 Serovars**.**

429 Footnotes

430 #*tviB* + means the strain is genotypically Vi positive.

431 *Footnote: A proportion of HG2 Serovars will be positive for the *ttr* gene and a
432 combination of targets that do not match any of the HG3 profiles (Supp table 1).

433

434 **Table 4 –** Summary of gene profile results.

435

436 **Table 5 –** Summary of individual gene target results.

437 Footnote

438 *The combination of genes present were heterogeneous, please see Supplementary
439 Table 1 for details.

bioRxiv preprint doi: https://doi.org/10.1101/537654; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

440 **Supplementary table 1** - Comparison analysis of reference methods versus
441 *Salmonella* HG3 PCR and GeneFinder.

442

443 **Supplementary data 1 –** Supplementary data on how this PCR was multiplexed into
444 two triplexes and one monoplex. This data includes further precision data on the
445 multiplex version of this PCR, the recipe used to make the mastermixes and
446 associated tables and figures.

447

448 **Figure 1: Selection of representative strains to test *in silico***

449 Footnote: Population structure of *Salmonella* received at PHE between 2016 – 2017
450 and strains tested for PCR in this study totalling to 19,221 strains. Colour coded by
451 main eBURST groups (eBG), a representative strain (highlighted in orange) from
452 each sequence types within an eBG containing 3 or more isolates were selected for
453 *in silco* gene detection of the seven genes (*ttr, sseJ, srfJ, tviB, SPC0869, SPA2308*
454 *& staG*).

bioRxiv preprint doi: https://doi.org/10.1101/537654; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

455 **Table 1:**

456 (a) Via PCR

Sequence Type (ST)	eBURST Group (EBG)	Serovar	Serotype	Hazard Group (HG)	No.
1,2, 2173	13	*Salmonella* Typhi	9,12[Vi]:d: –	HG3	61
85, 129	11	*Salmonella* Paratyphi A	1,2,12:a:[1,5]	HG3	15
86	5	*Salmonella* Paratyphi B	1,4,[5],12:b:1,2	HG3	15
146	20	*Salmonella* Paratyphi C	6,7,[Vi]:c:1,5	HG3	6
Total HG3 *Salmonella*					**97**
11, 183	11, 183	*Salmonella* Enteritidis	1,9,12:g,m:–	HG2	14
19, 34, 36	19, 34, 36	*Salmonella* Typhimurium	1,4,[5],12:i:1,2	HG2	14
10	10	*Salmonella* Dublin	1,9,12[Vi]:g,p:–	HG2	14
43, 88, 2545	43, 88, 0	*Salmonella* Java	1,4,[5],12:b:1,2	HG2	13
2902, 3226, 139, 145	0, 0, 6,6	*Salmonella* Choleraesuis	6,7,:c:1,5	HG2	6
Variable*	Variable*	Selection of *Salmonella* ssp. from GeneFinder analysis*	Variable – see supplementary Table 1	HG2	53
Total HG2 *Salmonella*					**114**
245, 152, 252, 7375	CC245, 152, 145, 0	*Shigella flexneri, S. sonnei, S. boydii, S. dysenteriae*	3a, N/A, O6, O1	HG2	4
11,29, 40	CC11, 21, 40	*E. coli* EPEC - eae +ve, STEC - stx1a, STEC - eae, stx2a	O55:H12, O77:H1, O157:H7	HG2	3
Total of Non-*Salmonella*					**7**
Total of isolates tested					**218**

457 *A random selection of HG2 ST containing sporadic gene targets were chosen. EPEC:
458 Enteropaothogenic *E. coli.* STEC: Shiga toxin-producing *E. coli.*

459

460

bioRxiv preprint doi: https://doi.org/10.1101/537654; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

461 (b) via *in silico* (GeneFinder) analysis

Strains	**No.**
Salmonella Typhi	556
Salmonella Paratyphi A	315
Salmonella Paratyphi B	53
Salmonella Paratyphi C	6
HG2 Serovars	952
Non-Salmonellae	7
Total	1889
Strains	**No.**
No. of different Sequence Types	480
Subspecies I	1821/1889
Subspecies II	14/1889
Subspecies IIIa	14/1889
Subspecies IIIb	29/1889
Subspecies IV	3/1889
Subspecies V	1/1889

462

463

464

465

466

467

468

469

470

471

472

473

bioRxiv preprint doi: https://doi.org/10.1101/537654; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

474 **Table 2:**

Gene	Name	Sequence 5'-3'	NCBI Accession Number	Reference
ttr	*ttr*_F *ttr*_R *ttr*_P	CTCACCAGGAGATTACAACATGG AGCTCAGACCAAAAGTGACCATC **FAM-**CACCGACGGCGAGACCGACTTT-**BHQ1**	AF282268	(6)
sseJ	*sseJ*_F *sseJ*_R *sseJ*_P	CGAGACTGCCGATGCATTTA GTACATAGCCGTGGTGAGTATAAG **CY3-**TGGAGGCGGCCAGTAATATTGGTT-**BHQ1**	AF294582	This study
srfJ	*srfJ*_F *srfJ*_R *srfJ*_P	CTGTCTGTATAGCGTGGAAGAG GTCCACCAGGCCATCTTTAT **JOE**-CGGCAGGGTATGGATGAGATGGAG-**BHQ1**	AF231759	This study
tviB	*tviB*_F *tviB*_R *tviB*_P	TGTGGTAAAGGAACTCGGTAAA GACTTCCGATACCGGGATAATG **JOE-**TGGATGCCGAAGAGGTAAGACGAGA-**BHQ1**	NC_003198	(7) Modified
SPC0869	*SPC0869*_F *SPC0869*_R *SPC0869*_P	CTGGCTGACACATGAACAAATC CCTGAGAACGAGTCAGGTTTAC **CY5**-TGTACGACTGCAAACGCCAAAGTC-**BHQ2**	NC_012125	This study
SPA2308	*SPA2308_F* *SPA2308_R* *SPA2308_P*	ACGATGATGACTGATTTATCGAAC TGAAAAGATATCTCTCAGAGCTGG **CY5-**CCCATACAATTTCATTCTTATTGAGAATGCGC-**BHQ2**	FM200053	(9).
staG	*staG_F* *staG_R* *staG_P*	CGCGAAGTCAGAGTCGACATAG AAGACCTCAACGCCGATCAC **FAM-**CATTTGTTCTGGAGCAGGCTGACGG-**BHQ1**	AL513382	(9).

475 Note – F – forward primer, R – reverse primer, P- probe

bioRxiv preprint doi: https://doi.org/10.1101/537654; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

476 **Table 3:**

477

Profile	***Salmonella* identification**	***ttr***	***sseJ***	***tviB#***	***srfJ***	***SPC0869***	***SPA2308***	***staG***
1	HG3 *Salmonella* Typhi	**+**	**-**	**+/-**	**-**	**-**	**-**	**+**
2	HG3 *Salmonella* Paratyphi A	**+**	**-**	**-**	**-**	**-**	**+**	**-**
3	HG3 *Salmonella* Paratyphi B	**+**	**-**	**-**	**+**	**-**	**-**	**-**
4	HG3 *Salmonella* Paratyphi C	**+**	**+**	**+**	**-**	**+**	**-**	**-**
5	HG2 Serovar*	**+**	**+**	**-**	**+/-**	**-**	**-**	**-**
6	Non-*Salmonella* spp.	-	-	-	-	-	-	-

478 #*tviB* + means the strain is genotypically Vi positive

479 *Footnote: A proportion of HG2 Serovars will be positive for the *ttr* gene and a combination
480 of targets that do not match any of the HG3 profiles (Supp table 1).

481

482

483

484 **Table 4**

Profile	*Salmonella* serovar (No. tested)	Expected Genes present	Result	Matches
1	HG3 *Salmonella* Typhi (556)	*ttr*, (*tviB* +/-), *staG*	556	100%
2	HG3 *Salmonella* Paratyphi A (315)	*ttr, SPA2308*	315	100%
3	HG3 *Salmonella* Paratyphi B (53)	*ttr, srfJ*	53	100%
4	HG3 *Salmonella* Paratyphi C (6)	*ttr, sseJ, tviB, SPC0869*	6	100%
5	HG2 Serovar (952)	*ttr*, (+ combination of any of the following not fitting the above profiles: *sseJ, srfJ, SPC0869, SPA2308, staG*),	952	100%
6	Non-*Salmonella* spp. (7)	negative for all genes	7	100%

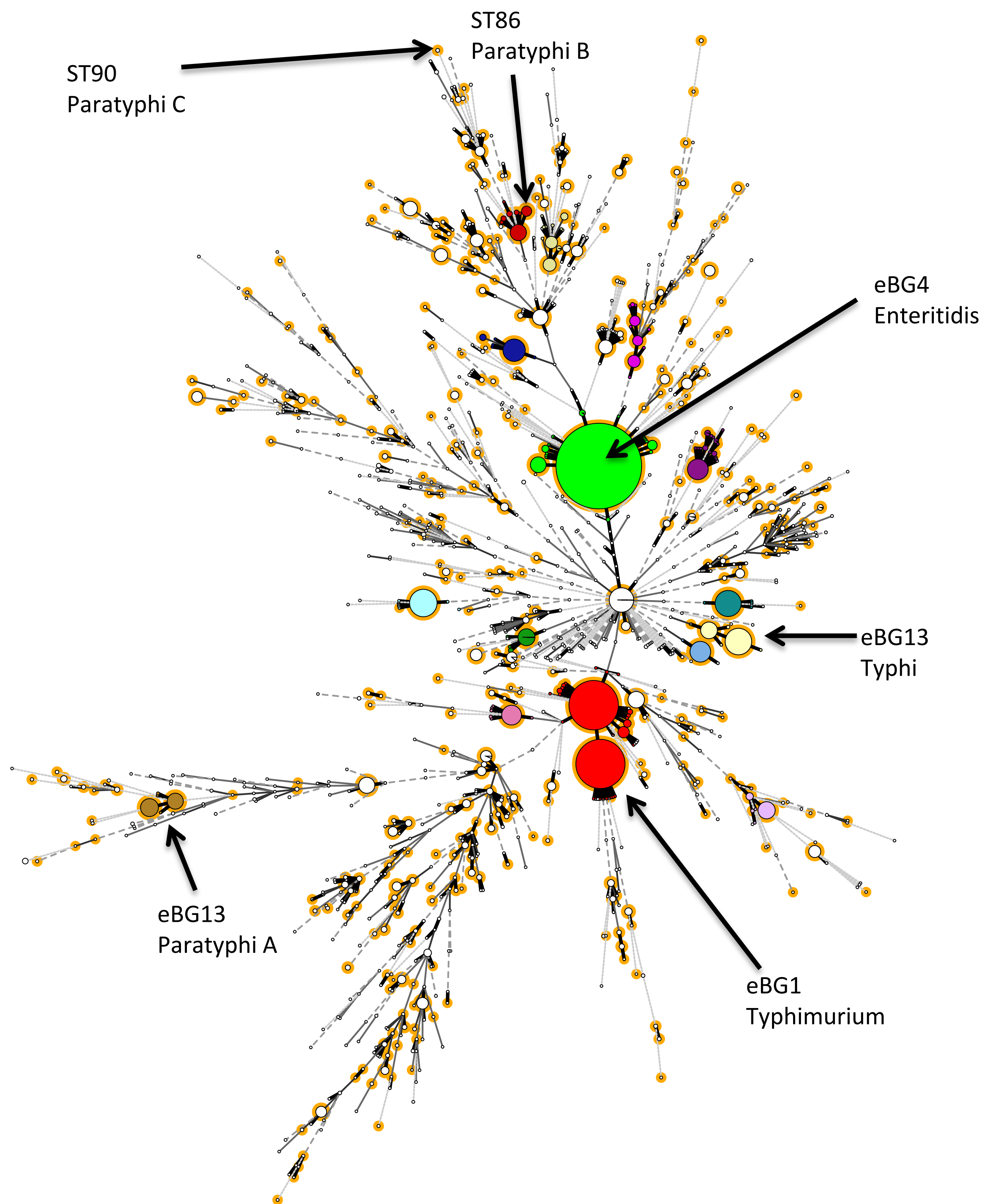
485

486 **Table 5**

Salmonella strain (No. tested)	*ttr*	*sseJ*	*tviB*	*srfJ*	*SPC0869*	*SPA2308*	*staG*
HG3 *Salmonella* Typhi (556)	556	0	553	0	0	0	556
HG3 Atypical *Salmonella* Typhi (3)	3	0	0	0	0	0	3
HG3 *Salmonella* Paratyphi A (315)	315	0	0	0	0	315	0
HG3 *Salmonella* Paratyphi B (53)	53	0	0	53	0	0	0
HG3 *Salmonella* Paratyphi C (6)	6	6	6	0	6	0	0
*HG2 Serovar (952)	952	938	0	380	19	50	41
Non-*Salmonella* spp. (7)	0	0	0	0	0	0	0

487 *The combination of genes present were heterogeneous, please see Supplementary Table 1 for details

488


ST90
Paratyphi C
ST86
Paratyphi B
eBG4
Enteritidis
eBG13
Typhi
eBG13
Paratyphi A
eBG1
Typhimurium