



Swansea University  
Prifysgol Abertawe



## Cronfa - Swansea University Open Access Repository

---

This is an author produced version of a paper published in :  
*Nature Genetics*

Cronfa URL for this paper:  
<http://cronfa.swan.ac.uk/Record/cronfa29259>

---

### **Paper:**

Humphrey, T. Distinct Salmonella Enteritidis lineages associated with enterocolitis in high-income settings and invasive disease in low-income settings. *Nature Genetics*

---

This article is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence. Authors are personally responsible for adhering to publisher restrictions or conditions. When uploading content they are required to comply with their publisher agreement and the SHERPA RoMEO database to judge whether or not it is copyright safe to add this version of the paper to this repository.  
<http://www.swansea.ac.uk/iss/researchsupport/cronfa-support/>

1 Distinct *Salmonella* Enteritidis lineages associated with enterocolitis in high-income  
2 settings and invasive disease in low-income settings

3

4

5 Authors

6 Nicholas A. Feasey <sup>1,2,3,4</sup>, James Hadfield <sup>2</sup>, Karen H. Keddy <sup>5,6</sup>, Timothy J Dallman  
7 <sup>7</sup>, Jan Jacobs <sup>8,9</sup>, Xiangyu Deng <sup>10,11</sup>, Paul Wigley <sup>4</sup>, Lars Barquist <sup>12</sup>, Gemma C.  
8 Langridge <sup>2</sup>, Theresa Feltwell <sup>2</sup>, Simon R. Harris <sup>2</sup>, Alison E. Mather <sup>2\*</sup>, Maria Fookes  
9 <sup>2</sup>, Martin Aslett <sup>2</sup>, Chisomo Msefula <sup>3,13</sup>, Samuel Kariuki <sup>14</sup>, Calman A. Maclennan <sup>2,15</sup>,  
10 Robert S. Onsare <sup>14</sup>, François-Xavier Weill <sup>16</sup>, Simon Le Hello <sup>16</sup>, Anthony M. Smith <sup>5</sup>,  
11 Michael McClelland <sup>17</sup>, Prerak Desai <sup>17</sup>, Christopher M. Parry <sup>1,18</sup>, John Cheesbrough  
12 <sup>20</sup>, Neil French <sup>4</sup>, Josefina Campos <sup>21</sup>, Jose A. Chabalgoity <sup>22</sup>, Laura Betancor <sup>22</sup>, Katie  
13 L. Hopkins <sup>23</sup>, Satheesh Nair <sup>7</sup>, Tom J. Humphrey <sup>24</sup>, Octavie Lunguya <sup>25,26</sup>, Tristan A.  
14 Cogan <sup>27</sup>, Milagritos D. Tapia <sup>28</sup>, Samba O. Sow <sup>29</sup>, Sharon M. Tennant <sup>28</sup>, Kristin  
15 Bornstein <sup>28</sup>, Myron M. Levine <sup>28</sup>, Lizeth Lacharme-Lora <sup>4</sup>, Dean B. Everett <sup>4</sup>, Robert  
16 A. Kingsley <sup>2,30</sup>, Julian Parkhill <sup>2</sup>, Robert S. Heyderman <sup>3,31</sup>, Gordon Dougan <sup>2</sup>, Melita A  
17 Gordon† <sup>3,4</sup> & Nicholas R. Thomson† <sup>2,18</sup>

18

19 Affiliations

- 20 1. Liverpool School of Tropical Medicine, Liverpool, UK
- 21 2. Wellcome Trust Sanger Institute, Cambridge, UK
- 22 3. Malawi Liverpool Wellcome Trust Clinical Research Programme, University of Malawi College of  
23 Medicine, Blantyre, Malawi
- 24 4. Institute of Infection and Global Health, University of Liverpool, Liverpool, UK
- 25 5. Centre for Enteric Diseases, National Institute for Communicable Diseases, Johannesburg, South Africa
- 26 6. Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa
- 27 7. Gastrointestinal Bacteria Reference Unit, Public Health England, Colindale, UK
- 28 8. Institute of Tropical Medicine, Antwerp, Belgium
- 29 9. Department of Microbiology and Immunology, University of Leuven, Belgium
- 30 10. Center for Food Safety, Department of Food Science and Technology, University of Georgia
- 31 11. Centers for Disease Control and Prevention, Atlanta, USA
- 32 12. Institute for Molecular Infection Biology, University of Würzburg, Würzburg, Germany
- 33 13. University of Malawi, The College of Medicine, Blantyre, Malawi
- 34 14. Centre for Microbiology Research, Kenya Medical Research Institute, Nairobi, Kenya

- 35 15. Jenner Institute, Nuffield Department of Medicine, University of Oxford, Oxford, UK  
36 16. Institut Pasteur, Paris, France  
37 17. Department of Microbiology and Molecular Genetics, University of California, Irvine, CA, USA  
38 18. London School of Hygiene & Tropical Medicine, London, UK  
39 19. School of Tropical Medicine and Global Health, Nagasaki University, Japan  
40 20. Department of Epidemiology and Population Health, University of Liverpool, Liverpool, UK  
41 21. Enteropathogen Division, Administración Nacional de Laboratorios e Institutos de Salud (ANLIS)  
42 Carlos G. Malbran Institute, Buenos Aires, Argentina  
43 22. Instituto de Higiene, Facultad de Medicina, Universidad de la Republica, Uruguay  
44 23. Antimicrobial Resistance and Healthcare-Associated Infections Reference Unit, Public Health England,  
45 Colindale, UK  
46 24. Swansea Medical School, Swansea University, Swansea, UK  
47 25. National Institute of Biomedical Research, Kinshasa, the Democratic Republic of the Congo  
48 26. University Hospital of Kinshasa, Kinshasa, the Democratic Republic of the Congo  
49 27. School of Veterinary Sciences, University of Bristol, Bristol, UK  
50 28. Center for Vaccine Development, University of Maryland School of Medicine, Baltimore, Maryland, USA  
51 29. Centre pour le Développement des Vaccins, Bamako, Mali  
52 30. Institute of Food Research, Colney, Norwich, UK  
53 31. Division of Infection and Immunity, University College London, London, UK  
54 \* Current address: Department of Veterinary Medicine, University of Cambridge, Cambridge, UK

55 †These authors contributed equally

56  
57 Corresponding Author  
58 Nicholas Feasey  
59 Liverpool School of Tropical Medicine  
60 5 Pembroke Place  
61 Liverpool  
62 L3 5QA  
63 Tel: 0151 705 3214  
64 E-mail: Nicholas.feasey@lstmed.ac.uk  
65

66 Running Title: Emergence of distinct lineages of *S. Enteritidis*

67 Key words: *Salmonella* Enteritidis, global, multidrug resistant, bacteraemia,  
68 gastroenteritis  
69

70 **Abstract**

71 An epidemiological paradox surrounds *Salmonella enterica* serovar Enteritidis. In  
72 high-income settings, it has been responsible for an epidemic of poultry-associated,  
73 self-limiting enterocolitis, whilst in sub-Saharan Africa it is a major cause of invasive  
74 nontyphoidal *Salmonella* disease, associated with high case-fatality. Whole-genome  
75 sequence analysis of 675 isolates of *S. Enteritidis* from 45 countries reveals the  
76 existence of a global epidemic clade and two novel clades of *S. Enteritidis* that are  
77 each geographically restricted to distinct regions of Africa. The African isolates  
78 display genomic degradation, a novel prophage repertoire and have an expanded,  
79 multidrug resistance plasmid. *S. Enteritidis* is a further example of a *Salmonella*  
80 serotype that displays niche plasticity, with distinct clades that enable it to become a  
81 prominent cause of gastroenteritis in association with the industrial production of  
82 eggs, and of multidrug resistant, bloodstream invasive infection in Africa.

83

## 84 **Introduction**

85

86 *Salmonella enterica* serovar Enteritidis (hereafter referred to as *S. Enteritidis*) has  
87 been a global cause of major epidemics of enterocolitis, which have been strongly  
88 associated with intensive poultry farming and egg production<sup>1</sup>. The serovar is  
89 usually considered to be a generalist in terms of host range and has a low human  
90 invasiveness index, typically causing self-limiting enterocolitis<sup>2</sup>. Following a number  
91 of interventions in the farming industry involving both improved hygiene and  
92 poultry vaccination, epidemic *S. Enteritidis* has been in decline in many countries  
93 including the United Kingdom and USA<sup>3,4</sup>. *S. Enteritidis* has also been used  
94 extensively since the early 1900s as a rodenticide (named the “Danysz virus”),  
95 following development at Institut Pasteur, France. Although by the 1960s,  
96 *Salmonella*-based rodenticides had been banned in the US, Germany and the UK, *S.*  
97 *Enteritidis* is still produced as a rodenticide in Cuba, under the name Biorat®<sup>5</sup>.

98

99 Serovars of *Salmonella* that cause enterocolitis in industrialised settings are strongly  
100 associated with life-threatening invasive nontyphoidal *Salmonella* (iNTS) disease in  
101 sub-Saharan Africa (SSA). *S. Enteritidis* and *Salmonella enterica* serovar  
102 Typhimurium (*S. Typhimurium*) are the two leading causes of iNTS disease in SSA<sup>6</sup>  
103 and both are associated with multidrug resistance (MDR)<sup>7</sup>. The clinical syndrome  
104 iNTS disease is associated with immunosuppression in the human host, particularly  
105 malnutrition, severe malaria and advanced HIV in young children and advanced HIV  
106 in adults<sup>8</sup>. It has been estimated to cause 681,000 deaths per year<sup>9</sup>.

107

108 *Salmonella* is a key example of a bacterial genus in which there is a recognizable  
109 genomic signature that distinguishes between a gastrointestinal and an extra-  
110 intestinal/invasive lifestyle<sup>10</sup>, whereby functions required for escalating growth in  
111 an inflamed gut are lost when the lineage becomes invasive<sup>11</sup>. In order to investigate  
112 whether there were distinct bacterial characteristics explaining the very different  
113 epidemiological and clinical profile of epidemic isolates of serotype *S. Typhimurium*

114 from SSA and industrialised settings, whole-genome sequence (WGS) investigations  
115 of this serovar were previously undertaken. These revealed a novel pathotype of  
116 multilocus sequence type (MLST) ST313 from SSA, which differed from clades that  
117 cause enterocolitis in industrialised settings, by showing patterns of genomic  
118 degradation potentially associated with more invasive disease and differential host  
119 adaptation<sup>12-17</sup>.

120

121 In relation to *S. Enteritidis*, there is a growing body of literature on the evolutionary  
122 history, phylogeny and utility of WGS for surveillance of *S. Enteritidis* outbreaks<sup>18-20</sup>.  
123 The broadest study of the phylogeny to date revealed five major lineages, but  
124 contained only two African isolates<sup>21</sup>. There have also been limited reports of  
125 isolates of *S. Enteritidis* from African patients living in Europe that are MDR and  
126 which display a distinct phage type (PT 42)<sup>22,23</sup>. We therefore hypothesized that  
127 there are distinct lineages of *S. Enteritidis* circulating in both the industrialised and  
128 developing world with different origins, likely distinct routes of spread and that are  
129 associated with different patterns of disease, which will display the distinct genomic  
130 signatures characteristic of differential adaptation. To investigate this we have  
131 collected a highly diverse global collection of *S. Enteritidis* isolates and compared  
132 them using whole-genome sequencing, the highest possible resolution typing  
133 methodology.

134

135

136

137

138

139

140

141

142 **Results**

143

144 **Isolate collection**

145

146 In total, 675 isolates of *S. Enteritidis* isolated between 1948 and 2013 were  
147 sequenced. The collection originated from 45 countries and six continents (Table 1).  
148 496/675 isolates were from Africa, with 131 from the Republic of South Africa  
149 (RSA), a further 353 from the rest of SSA, and 12 from North Africa (Table 1). There  
150 were 343 isolates from normally sterile human sites (invasive), 124 non-invasive  
151 human isolates (predominantly stool samples) and 40 from animal, food or  
152 environmental sources. The full metadata are described in Supplementary Table 1  
153 and have been uploaded to the publically available database Enterobase.

154

155 **Phylogeny**

156

157 675 *S. Enteritidis* genomes and one *Salmonella enterica* serovar Gallinarum were  
158 mapped to the *S. Enteritidis* strain P125109 reference sequence, variable regions  
159 excluded and the remaining sites were screened for single nucleotide  
160 polymorphisms (SNPs). This left an alignment containing a total of 42,373 variable  
161 sites, from which a maximum likelihood (ML)-phylogeny was constructed using *S.*  
162 Gallinarum, which is a closely related serovar, as an out-group (Figure 1). HierBAPS  
163 was run over two rounds, which provided clear distinction between  
164 clades/clusters<sup>24</sup>. The phylogeny of *S. Enteritidis* revealed evidence of three clades  
165 associated with epidemics, one which we have termed the 'global epidemic clade'  
166 and includes the reference PT4 isolate P125109 and two African clades: one  
167 predominantly composed of West African isolates (labeled the 'West African clade')  
168 and a second composed of isolates predominantly originating in Central and Eastern  
169 Africa, called the 'Central/Eastern African clade'). Figure 1 also shows the other  
170 clades and clusters predicted by HierBAPS, the largest of which is a paraphyletic

171 cluster from which the global epidemic clade emerged (Outlier Cluster in Figure 1),  
172 and a further five smaller clades or clusters predicted by HierBAPS.

173

174 The global epidemic clade contains isolates of multiple phage types, including 4 and  
175 1, which have been linked to the global epidemic of poultry associated human  
176 enterocolitis<sup>25</sup>. It comprised 250 isolates from 28 countries, including 43 from  
177 Malawi and 82 from RSA. They were isolated from across a 63-year period (1948-  
178 2013). Antimicrobial susceptibility testing had been performed on 144 isolates and  
179 104 were susceptible to all antimicrobials tested, five were multidrug resistant  
180 (MDR: resistant to 3 or more antimicrobial classes), one was nalidixic acid resistant  
181 and none were extended-spectrum beta-lactamase (ESBL)-producing isolates.  
182 Database comparison of the genomes from this clade revealed that 221 (88%) of  
183 them contained no predicted antimicrobial resistance (AMR) genes apart from the  
184 cryptic resistance gene *aac(6')-Iy*<sup>26</sup>.

185

186 The global epidemic clade has emerged from a diverse cluster previously described  
187 by Zheng<sup>27</sup>, which encompassed 131 isolates (Figure 1: 'Outlier Cluster'). In addition  
188 to being paraphyletic, this group was geographically and temporally diverse, and  
189 predominantly drug susceptible (59/71 isolates). Whilst the majority of the  
190 diversity of phage typed isolates was contained within the global epidemic clade,  
191 this cluster alone contained isolates of phage type 14b, which was recently  
192 associated with a multi-country outbreak of *S. Enteritidis* enterocolitis in Europe  
193 associated with chicken eggs from Germany<sup>28</sup>. There were also 41 isolates from RSA  
194 in this clade, where it has been a common cause of bloodstream infection, and 39  
195 bloodstream isolates from Malawi. Database comparison of the genomes from this  
196 clade revealed that 122 (82%) of these genomes contained no predicted AMR genes  
197 apart from the cryptic resistance gene *aac(6')-Iy*.

198

199 There were two related, but phylogenetically and geographically distinct, epidemic  
200 clades that largely originated from SSA. The Central/Eastern African clade included



201 166 isolates, all but two of which (from RSA) came from this region. Of these,  
202 126/155 (82%) were MDR and 148/153 (97%) displayed phenotypic resistance to  
203 between one and four antimicrobial classes. All of these genomes contained at least  
204 five predicted resistance genes and 128 (77%) contained nine (Table 2 and  
205 Supplementary Table 2). 155/165 (94%) of these isolates were cultured from a  
206 normally sterile compartment of a human (i.e. blood or cerebrospinal fluid) and  
207 were considered to be causing invasive disease (Table 2). The second African  
208 epidemic clade was significantly associated with West Africa with 65/66 isolates  
209 coming from this region and one isolate from USA. This clade was also associated  
210 with drug resistance (62 [94%] resistant to  $\geq 1$  antimicrobial class by phenotype and  
211 genotype) and human invasive disease (61 [92%]). It also included two isolates that  
212 were subtyped as phage type 4.

213

214 The remaining 58 isolates included in this study were extremely diverse,  
215 phylogenetically, temporally and geographically. Only two displayed any phenotypic  
216 AMR, one of which was MDR. Inspection of the genome revealed that five had  
217 predicted AMR genes in addition to *aac(6')-Iy*, four of which were isolated in sub-  
218 Saharan Africa. Twenty were associated with invasive human disease, and six were  
219 recovered from stool. Three isolates were from stocks of rodenticide and these were  
220 phylogenetically remote from both global-epidemic and the two African epidemic  
221 clades.

222

223 To add further context to these findings we screened the entire publically available  
224 Public Health England (PHE) sequenced *Salmonella* routine surveillance collection,  
225 which includes 2,986 *S. Enteritidis* genomes, 265 of which were associated with  
226 travel to Africa (Supplementary Figure 1). Within this huge collection, including 61  
227 (2.0%) bloodstream isolates and 2670 (89.4%) stool isolates, only 6 isolates (4 from  
228 blood culture, 1 from stool) fell within to the West African clade and 1 (from stool)  
229 belonged to the Central/Eastern African clade. Notably, these isolates were all either  
230 associated with travel to Africa and/or taken from patients of African origin.

231 It is apparent from the location of the archetypal reference isolate and archetypal  
232 phage types in the phylogeny (Supplementary Figure 2) that the majority of *S.*  
233 *Enteritidis* studied previously belonged to the global epidemic clade associated with  
234 enterocolitis in industrialised countries. Furthermore, it is also clear that two  
235 additional, previously unrecognized *S. Enteritidis* lineages have emerged, largely  
236 restricted to Africa, that are strongly associated with MDR and invasive disease.

237

238 To understand how recently these Africa-associated lineages emerged we used  
239 Bayesian Evolutionary Analysis by Sampling Trees (BEAST) to reconstruct the  
240 temporal history of the epidemic clades<sup>29</sup>. These data (Supplementary Figure 3)  
241 estimate the most recent common ancestor (MRCA) of the Central/Eastern African  
242 clade dates to 1945 (95% Credible Interval [CrI]: 1924-1951) and for the West  
243 African clade it was 1933 (95% CrI: 1901-1956). We estimate the MRCA of the  
244 global epidemic clade originated around 1918 (95% CrI: 1879-1942 –  
245 Supplementary Figure 4), with a modern expansion occurring in 1976 (95% CrI:  
246 1968-1983), whereas the paraphyletic cluster from which it emerged dates to  
247 approximately 1711 (95% CrI: 1420-1868).

248

#### 249 **Contribution of the accessory genome**

250

251 Prophages have the potential to carry non-essential "cargo" genes, which suggests  
252 they confer a level of specialization to their host bacterial species, whilst plasmids  
253 may confer a diverse array of virulence factors and AMR<sup>30,31</sup>. Therefore it is critical  
254 to evaluate the accessory genome in parallel with the core. 622 sequenced genomes  
255 were used to determine a pangenome, which yielded a core genome comprising  
256 4,076 predicted genes present in  $\geq 90\%$  isolates, including all 12 recognised  
257 *Salmonella* Pathogenicity Islands as well as all 13 fimbrial operons found in the  
258 P125109 reference<sup>32</sup>. The core gene definition was set to minimize stochastic loss of  
259 genes from the core due to errors in individual assemblies across such a large  
260 dataset. The accessory genome consisted of 14,015 predicted genes. Of the

261 accessory genes, 324 were highly conserved across the global and two African  
262 epidemic clades, as well as the outlier cluster. Almost all were associated with the  
263 acquisition or loss of mobile genetic elements (MGEs) such as prophage or plasmids.  
264 Prophage regions have been shown to be stable in *Salmonella* genomes and are  
265 potential molecular markers, the presence of which has previously been used to  
266 distinguish specific clades<sup>13,33</sup>.

267

268 The lineage-specific whole gene differences of the major clades are summarized in  
269 Figure 2 and plotted against the representatives of the four major clades in  
270 Supplementary Figure 5. The lineage specific sequence regions include 57 predicted  
271 genes found to be unique to the global epidemic clade (Figure 2), all of which were  
272 associated with prophage  $\phi$ SE20, a region shown to be essential for invasion of  
273 chicken ova and mice in one previous study<sup>34</sup>. There were a further 39 genes  
274 conserved in the global epidemic and the paraphyletic outlying cluster, which were  
275 absent from both African clades, 26 of which correspond to region of difference  
276 (ROD) 21<sup>32</sup>. The Central/Eastern Africa clade contained 77 predicted genes that  
277 were absent in the other clades. 33 were associated with the virulence plasmid and  
278 a further 40 chromosomal genes were associated with a novel, Fels-2 like prophage  
279 region ( $\phi$ fels-BT). The West African clade had only 15 distinct predicted genes, 11 of  
280 which were plasmid-associated. The two African clades shared a further 102 genes:  
281 48, including a leucine-rich repeat region, were associated with a novel prophage  
282 region closely related to *Enterobacter* phage P88, 44 were associated with a Gifsy-1  
283 prophage found in *S. Bovismorbificans* and eight were associated with a Gifsy-2  
284 prophage which has degenerated in the reference P125109.

285

286 The *S. Enteritidis* plasmid is the smallest of the generic *Salmonella* virulence  
287 plasmids at 58 kb and is unusual in that it contains an incomplete set of *tra* genes  
288 that are responsible for conjugative gene transfer. The phylogeny of the *S.*  
289 *Enteritidis* virulence plasmid backbone was reconstructed using reads that mapped  
290 to the *S. Enteritidis* reference virulence plasmid, pSENV. 120/675 (18%) genomes

291 lacked pSENV. The virulence plasmid phylogeny is similar to that of the  
292 chromosome, suggesting that they have been stably maintained by each lineage and  
293 diversified alongside them (Supplementary Figure 6).

294

295 The virulence plasmids from the African clades were much larger than those held in  
296 the other clades at ~90 kb. A representative example was extracted from Malawian  
297 isolate D7795, sequenced using long read technology to accurately reconstruct it  
298 (PacBio; see methods) and denoted pSEN-BT (Accession number LN879484). pSEN-  
299 BT is composed of a backbone of pSENV with additional regions that are highly  
300 similar to recently sequenced fragments of an novel *S. Enteritidis* virulence plasmid  
301 (pUO-SeVR) isolated from an African patient presenting with MDR invasive *S.*  
302 *Enteritidis* in Spain<sup>22</sup>. Plasmid pSEN-BT harbours nine AMR genes (full list in  
303 Supplementary Table 2), plus additional genes associated with virulence and a  
304 toxin/antitoxin plasmid addiction system. Of note, plasmids from the West African  
305 isolates carry resistance gene chloramphenicol acetyl transferase A1 (*catA1*),  
306 whereas the Central/Eastern African strains carry *catA2* and tetracycline resistance  
307 gene *tet(A)*. Like pSENV, the African virulence plasmid contained an incomplete set  
308 of *tra* genes and so is not self-transmissible. This was confirmed by conjugation  
309 experiments and is consistent with previous reports<sup>22,23</sup>. These observations  
310 suggest that the evolution of the *S. Enteritidis* plasmid mirrors that of the  
311 chromosome; it is thus not a 'novel' plasmid, but in different SSA locations has  
312 acquired different AMR genes.

313

### 314 **Multiple signatures of differential host adaptation**

315

316 It has been observed in multiple serovars of *Salmonella* including *S. Typhi*, *S.*  
317 *Gallinarum* and *S. Typhimurium* ST313 that the degradation of genes necessary for  
318 the utilization of inflammation-derived nutrients is a marker of that lineage having  
319 moved from an intestinal to a more invasive lifestyle<sup>13,14,32,35</sup>. Accordingly, we have  
320 looked for similar evidence within a representative example of a MDR, invasive,

321 Central/Eastern African clade isolate, D7795, that was isolated from the blood of a  
322 Malawian child in 2000. The draft genome sequence of D7795 closely resembles  
323 that of P125109, however, in addition to the novel prophage repertoire and plasmid  
324 genes described above, it harbours a number of predicted pseudogenes or  
325 hypothetically disrupted genes (HDGs)<sup>11</sup>.

326

327 In total, there were 42 putative HDGs in D7795, many of which are found in genes  
328 involved in gut colonisation and fecal shedding as well as various metabolic  
329 processes such as cobalamine biosynthesis which is a cofactor for anaerobic  
330 catabolism of inflammation-derived nutrients, such as ethanolamine, following  
331 infection<sup>36</sup>. Curation of the SNPs and insertions or deletions (indels) predicted to be  
332 responsible for pseudogenisation across the Central/Eastern African clade and West  
333 African clade revealed 37/42 predicted HDGs were fixed in other representatives of  
334 the Central/East African clade, with 27 of them being present in over 90% of  
335 isolates from that clade. Relatively fewer HDGs in D7795 (19/42) were present in  
336 representatives of the West African clade, although 13 were present in  $\geq 90\%$  of  
337 isolates (Supplementary Table 3).

338

339 In addition to this evidence of reductive evolution in D7795, there were 363 genes  
340 containing non-synonymous (NS)-SNPs, which change the amino acid sequence and  
341 so may have functional consequences<sup>37</sup>. The two African clades were screened for  
342 the presence of these NS-SNPs and 131 were found to be present and completely  
343 conserved across both clades, including NS-SNPs in 43 genes encoding predicted  
344 membrane proteins, 36 metabolic genes and 23 conserved hypothetical genes  
345 (Supplementary Table 4). Furthermore many of these NS-SNPs fall in genes within  
346 the same metabolic pathways as the HDGs (see Supplementary Note for detailed  
347 description). Supplementary Table 5 provides a list of some of the common traits  
348 identified amongst the functions of genes lost independently by D7795, *S. Typhi* and  
349 *S. Gallinarum*. The disproportionate clustering of mutations in membrane structures

350 observed in the African clades is yet another sign of differential host adaptation  
351 analogous to that reported in both *S. Typhi*<sup>35</sup> and *S. Gallinarum*<sup>32</sup>.

352

### 353 **Biolog™ growth substrate platform profiling**

354

355 The Biolog™ platform was utilized to generate a substrate growth utilisation profile  
356 for selected *S. Enteritidis* isolates. Corresponding signal values of replicate pairs of a  
357 Central/Eastern African isolate (D7795) and a global epidemic isolate (A1636) were  
358 compared using principal component analysis and found to be highly consistent. In  
359 total, 80 metabolites showed evidence of differential metabolic activity (Figure 3).

360 Evaluation of data from the Central/Eastern African isolate using Pathway Tools  
361 software revealed that 14/27 (52%) of pathways with evidence of decreased  
362 metabolic activity at 28°C had a corresponding component of genomic degradation.  
363 This was also true for 12/30 (40%) of pathways with evidence of decreased  
364 metabolic activity at 37°C.

365

366 Instances of reduced metabolic activity in a Central/Eastern African strain (D7795)  
367 compared to a global epidemic strain (A1636) included dulcitol and glycolic acid in  
368 the glycerol degradation pathway, propionic acid in the propanediol pathway and  
369 ethylamine and ethanolamine. These are all vitamin B12 (cobalamin) dependent  
370 reactions, for which there was a corresponding signature of genomic degradation.  
371 Also there was reduced activity in response to three forms of butyric acid, alloxan  
372 and allantoic acid metabolism. Allantoin can be found in the serum of birds, but not  
373 humans and is utilised as a carbon source during *S. Enteritidis* infection of  
374 chickens<sup>38</sup>, and HDGs relating to allantoin have been noted in *S. Typhimurium*  
375 ST313<sup>13</sup>. The full list of differences is detailed in Supplementary Table 6 and 7. This  
376 is a further sign of decreased metabolism of the Central/Eastern African isolate in  
377 the anaerobic environment of the gut.

378

### 379 **Chick infection model suggests divide in host range**

380

381 Given the phenotypic differences observed in the genotypically distinct global and  
382 African clades, we hypothesized that these lineages could have differing infection  
383 phenotypes in an *in vivo* challenge model. We compared the infection profile of a  
384 member of the Central/Eastern African clade (D7795) to the reference global  
385 epidemic strain P125109 in an avian host. The chicken group infected with P125109  
386 showed mild hepatosplenomegaly consistent with infection by this *Salmonella*  
387 serovar and cecal colonization (Figure 4A-C). In contrast, the Central/Eastern  
388 African strain displayed significantly reduced invasion at 7 dpi of both liver  
389 ( $p=0.027$ ) and spleen ( $p=0.007$ ), however cecal colonization was not significantly  
390 reduced ( $p=0.160$ ). This is in marked contrast to the behavior of *S. Typhimurium*  
391 ST313, which is more invasive in a chick infection model<sup>12</sup>.

392

393

## 394 **Discussion**

395

396 *S. Enteritidis* is an example of a successful *Salmonella* lineage with the apparent  
397 ability to adapt to different hosts and transmission niches as and when  
398 opportunities for specialization have presented. Langridge *et al* recently evaluated  
399 the Enteritidis/Gallinarum/Dublin lineage of *Salmonella*, revealing components of  
400 the nature and order of events associated with host-range and restriction<sup>39</sup>. In the  
401 present study, we have highlighted the plasticity of *S. Enteritidis*, providing evidence  
402 of three distinct epidemics of human disease. In addition we show multiple  
403 additional clades and clusters that demonstrate the huge reservoir of diversity  
404 amongst *S. Enteritidis* from which future epidemics might emerge.

405

406 An important question posed by this study is why have distinct clades of *Salmonella*  
407 emerged to become prominent causes of iNTS disease in Africa, from a serotype  
408 normally considered to be weakly invasive? The presence of a highly  
409 immunosuppressed population due to the HIV pandemic is clearly a key host factor

410 that facilitates the clinical syndrome iNTS disease<sup>40,41</sup>. In addition to human host  
411 factors, there are two distinct African epidemic lineages that have emerged in the  
412 last 90 years. Both lineages are significantly associated with a novel prophage  
413 repertoire, an expanded, MDR-augmented virulence plasmid, and patterns of  
414 genomic degradation with similarity to other host-restricted invasive *Salmonella*  
415 serotypes including *S. Typhi* and *S. Gallinarum* and to clades of *S. Typhimurium*  
416 associated with invasive disease in Africa<sup>13,32,35</sup>. This pattern of genomic  
417 degradation is concentrated in pathways specifically associated with an enteric  
418 lifestyle, however it is noteworthy that in the chick infection model, the African *S.*  
419 *Enteritidis* invaded the chick liver and spleen less well than the global pandemic  
420 clade. This raises the possibility that the two clades occupy different ecological  
421 niches outside the human host or that they behave differently within the human  
422 host and screening of the huge *S. Enteritidis* collection from routine *Salmonella*  
423 surveillance by PHE supports the assertion that these lineages are geographically  
424 restricted to Africa. This study therefore indicates a need to understand what these  
425 ecological niches might be, and then to define the transmission pathways of African  
426 clades of *S. Enteritidis*, in order to facilitate public health interventions to prevent  
427 iNTS disease.

428

429 The evolution of the *S. Enteritidis* virulence plasmid is intriguing; pSENV is the  
430 smallest of the known *Salmonella* virulence-associated plasmids, but in SSA, the  
431 plasmid has nearly doubled in size partly through the acquisition of AMR genes. The  
432 absence of *tra* genes necessary for conjugal transfer either indicates that MDR status  
433 has evolved through acquisition of MGEs multiple times or through clonal expansion  
434 and vertical transmission of the plasmid to progeny. The available data suggest that  
435 the former scenario has happened twice, once in West Africa, and once in  
436 Central/Eastern Africa.

437

438 Despite *S. Enteritidis* being reported as a common cause of bloodstream infection  
439 (BSI) in Africa<sup>6,7</sup> the Global Enteric Multicenter Study (GEMS) found that *Salmonella*



440 serotypes were an uncommon cause of moderate to severe diarrhoea in African  
441 children less than 5-years of age<sup>42</sup>. Our data associating the African lineages with  
442 invasive disease is also consistent with data presented in a recent Kenyan study  
443 comparing a limited number and diversity of *S. Enteritidis* isolates from blood and  
444 stool. Applying the lineages defined in this study to the genome data reported from  
445 Kenya showed that 20.4% of isolates from that study belonging to the global clade  
446 were associated with invasive disease, whereas 63.2% of the isolates in that study  
447 belonging to our Central/Eastern African clade were associated with invasive  
448 disease<sup>43</sup>. The remaining isolates were associated with cases of enterocolitis or  
449 asymptomatic carriage, confirming that the Central/Eastern African clade can also  
450 cause enterocolitis. The association of *S. Enteritidis* clades circulating in sub-  
451 Saharan Africa with iNTS disease may reflect the fact that their geographical  
452 distribution permits them to behave as opportunistic invasive pathogens in a setting  
453 where advanced immunosuppressive disease is highly prevalent in human  
454 populations.

455

456 In summary, two clades of *S. Enteritidis* have emerged in Africa, which have  
457 different phenotypes and genotypes to the strains of *S. Enteritidis* circulating in the  
458 industrial world. These strains display evidence of changing host adaptation,  
459 different virulence determinants and multi-drug resistance, a parallel situation to  
460 the evolutionary history of *S. Typhimurium* ST313. They may have different  
461 ecologies and/or host ranges to global strains and have caused epidemics of BSI in  
462 at least three countries in SSA, yet are rarely responsible for disease in South Africa.  
463 An investigation into the environmental reservoirs and transmission of these  
464 pathogens is warranted and urgently required.

465

466

467

468 **URLs:**

469

470 Enterobase: <https://enterobase.warwick.ac.uk/>

471

472 **Data access:** Accession numbers for sequencing data including both raw sequencing  
473 reads and assembled sequences are available in Supplementary Table 1.

474

#### 475 **ACKNOWLEDGEMENTS**

476 This work was supported by The Wellcome Trust. We would like to thank the  
477 members of the Pathogen Informatics Team and the core sequencing teams at the  
478 Wellcome Trust Sanger Institute (Cambridge, UK). We are grateful to David Harris  
479 for work in managing the sequence data.

#### 480 **FUNDING**

481 This work was supported by a number of organizations. The Wellcome Trust Sanger  
482 Institute authors were funded by Wellcome Trust Award 098051; NF was supported  
483 by the Wellcome Trust Research Fellowship WT092152MA. NF, RSH and this work  
484 were supported by a strategic award from the Wellcome Trust for the MLW Clinical  
485 Research Programme (101113/Z/13/Z). The Institut Pasteur (IP) authors were  
486 funded by the IP, the Institut de Veille Sanitaire, and by the French Government  
487 “Investissement d'Avenir” program (Integrative Biology of Emerging Infectious  
488 Diseases” Laboratory of Excellence, grant no. ANR-10-LABX-62-IBEID). JJ was  
489 supported by the antibiotic resistance surveillance project in DR Congo, funded by  
490 Project 2.01 of the Third Framework Agreement between the Belgian Directorate  
491 General of Development Cooperation and the Institute of Tropical Medicine,  
492 Antwerp, Belgium. SK was supported by the NIH Grant Number R01 AI099525-02.  
493 AEM was supported by Wellcome Trust grant 098051 whilst at the Wellcome Trust  
494 Sanger Institute, and Biotechnology and Biological Sciences Research Council grant  
495 BB/M014088/1 at the University of Cambridge. The funders had no role in study  
496 design, data collection and analysis, decision to publish, or preparation of the  
497 manuscript.

#### 498 **AUTHOR CONTRIBUTIONS**

499 Study design: NAF, NRT, MAG, GD, RAK, JP. Data analysis: NAF, NRT, JH, TF, LB, PQ  
500 LLL, GL, SRH, AEM, MF, MA. Isolate acquisition and processing and clinical data  
501 collection: NAF, KHK, JJ, XD, CMe, SK, CMI, RSO, FXW, SLH AMS, MM, PD, CMP, JC, NF,  
502 JC, JAC, LBe, KLH, TJH, OL, TAC, M T, SS, SMT, KB, MML, DBE, RSH. Manuscript  
503 writing: NAF, JH, NRT, MAG. All authors contributed to manuscript editing.

504

## 505 References

- 506 1. Ward, L.R., Threlfall, J., Smith, H.R. & O'Brien, S.J. Salmonella enteritidis  
507 epidemic. *Science* **287**, 1753-4; author reply 1755-6 (2000).
- 508 2. Jones, T.F. *et al.* Salmonellosis outcomes differ substantially by serotype. *J*  
509 *Infect Dis* **198**, 109-14 (2008).
- 510 3. O'Brien, S.J. The "decline and fall" of nontyphoidal salmonella in the United  
511 kingdom. *Clin Infect Dis* **56**, 705-10 (2013).
- 512 4. Braden, C.R. Salmonella enterica serotype Enteritidis and eggs: a national  
513 epidemic in the United States. *Clin Infect Dis* **43**, 512-7 (2006).
- 514 5. Friedman, C.R., Malcolm, G., Rigau-Perez, J.G., Arambulo, P., 3rd & Tauxe,  
515 R.V. Public health risk from Salmonella-based rodenticides. *Lancet* **347**, 1705-6  
516 (1996).
- 517 6. Reddy, E.A., Shaw, A.V. & Crump, J.A. Community-acquired bloodstream  
518 infections in Africa: a systematic review and meta-analysis. *Lancet Infect Dis* **10**,  
519 417-32 (2010).
- 520 7. Feasey, N.A. *et al.* Three Epidemics of Invasive Multidrug-Resistant Salmonella  
521 Bloodstream Infection in Blantyre, Malawi, 1998-2014. *Clin Infect Dis* **61 Suppl 4**,  
522 S363-71 (2015).
- 523 8. Feasey, N.A., Dougan, G., Kingsley, R.A., Heyderman, R.S. & Gordon, M.A.  
524 Invasive non-typhoidal salmonella disease: an emerging and neglected tropical  
525 disease in Africa. *Lancet* **379**, 2489-99 (2012).
- 526 9. Ao, T.T. *et al.* Global burden of invasive nontyphoidal salmonella disease,  
527 2010(1). *Emerg Infect Dis* **21**(2015).
- 528 10. Nuccio, S.P. & Baumber, A.J. Reconstructing pathogen evolution from the ruins.  
529 *Proc Natl Acad Sci U S A* **112**, 647-8 (2015).

- 530 11. Nuccio, S.P. & Baumber, A.J. Comparative analysis of Salmonella genomes  
531 identifies a metabolic network for escalating growth in the inflamed gut. *MBio* **5**,  
532 e00929-14 (2014).
- 533 12. Parsons, B.N. *et al.* Invasive non-typhoidal Salmonella typhimurium ST313 are  
534 not host-restricted and have an invasive phenotype in experimentally infected  
535 chickens. *PLoS Negl Trop Dis* **7**, e2487 (2013).
- 536 13. Kingsley, R.A. *et al.* Epidemic multiple drug resistant Salmonella Typhimurium  
537 causing invasive disease in sub-Saharan Africa have a distinct genotype.  
538 *Genome Res* **19**, 2279-87 (2009).
- 539 14. Okoro, C.K. *et al.* Signatures of adaptation in human invasive Salmonella  
540 Typhimurium ST313 populations from sub-Saharan Africa. *PLoS Negl Trop Dis*  
541 **9**, e0003611 (2015).
- 542 15. Herrero-Fresno, A. *et al.* The role of the st313-td gene in virulence of Salmonella  
543 Typhimurium ST313. *PLoS One* **9**, e84566 (2014).
- 544 16. Ramachandran, G., Perkins, D.J., Schmidlein, P.J., Tulapurkar, M.E. & Tennant,  
545 S.M. Invasive Salmonella Typhimurium ST313 with naturally attenuated flagellin  
546 elicits reduced inflammation and replicates within macrophages. *PLoS Negl Trop*  
547 *Dis* **9**, e3394 (2015).
- 548 17. Okoro, C.K. *et al.* Intracontinental spread of human invasive Salmonella  
549 Typhimurium pathovariants in sub-Saharan Africa. *Nat Genet* **44**, 1215-21  
550 (2012).
- 551 18. Allard, M.W. *et al.* On the evolutionary history, population genetics and diversity  
552 among isolates of Salmonella Enteritidis PFGE pattern JEGX01.0004. *PLoS One*  
553 **8**, e55254 (2013).
- 554 19. den Bakker, H.C. *et al.* Rapid whole-genome sequencing for surveillance of  
555 Salmonella enterica serovar enteritidis. *Emerg Infect Dis* **20**, 1306-14 (2014).
- 556 20. Taylor, A.J. *et al.* Characterization of Foodborne Outbreaks of Salmonella  
557 enterica Serovar Enteritidis with Whole-Genome Sequencing Single Nucleotide  
558 Polymorphism-Based Analysis for Surveillance and Outbreak Detection. *J Clin*  
559 *Microbiol* **53**, 3334-40 (2015).
- 560 21. Deng, X. *et al.* Genomic epidemiology of Salmonella enterica serotype Enteritidis  
561 based on population structure of prevalent lineages. *Emerg Infect Dis* **20**, 1481-9  
562 (2014).
- 563 22. Rodriguez, I., Guerra, B., Mendoza, M.C. & Rodicio, M.R. pUO-SeVR1 is an  
564 emergent virulence-resistance complex plasmid of Salmonella enterica serovar  
565 Enteritidis. *J Antimicrob Chemother* **66**, 218-20 (2011).

- 566 23. Rodriguez, I., Rodicio, M.R., Guerra, B. & Hopkins, K.L. Potential international  
567 spread of multidrug-resistant invasive *Salmonella enterica* serovar enteritidis.  
568 *Emerg Infect Dis* **18**, 1173-6 (2012).
- 569 24. Cheng, L., Connor, T.R., Siren, J., Aanensen, D.M. & Corander, J. Hierarchical  
570 and spatially explicit clustering of DNA sequences with BAPS software. *Mol Biol*  
571 *Evol* **30**, 1224-8 (2013).
- 572 25. Rodrigue, D.C., Tauxe, R.V. & Rowe, B. International increase in *Salmonella*  
573 enteritidis: a new pandemic? *Epidemiol Infect* **105**, 21-7 (1990).
- 574 26. Magnet, S., Courvalin, P. & Lambert, T. Activation of the cryptic *aac(6')*-ly  
575 aminoglycoside resistance gene of *Salmonella* by a chromosomal deletion  
576 generating a transcriptional fusion. *J Bacteriol* **181**, 6650-5 (1999).
- 577 27. Zheng, J. *et al.* Genetic diversity and evolution of *Salmonella enterica* serovar  
578 Enteritidis strains with different phage types. *J Clin Microbiol* **52**, 1490-500  
579 (2014).
- 580 28. Inns, T. *et al.* A multi-country *Salmonella* Enteritidis phage type 14b outbreak  
581 associated with eggs from a German producer: 'near real-time' application of  
582 whole genome sequencing and food chain investigations, United Kingdom, May  
583 to September 2014. *Euro Surveill* **20**(2015).
- 584 29. Drummond, A.J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by  
585 sampling trees. *BMC Evol Biol* **7**, 214 (2007).
- 586 30. Thomson, N. *et al.* The role of prophage-like elements in the diversity of  
587 *Salmonella enterica* serovars. *J Mol Biol* **339**, 279-300 (2004).
- 588 31. Rotger, R. & Casadesus, J. The virulence plasmids of *Salmonella*. *Int Microbiol*  
589 **2**, 177-84 (1999).
- 590 32. Thomson, N.R. *et al.* Comparative genome analysis of *Salmonella* Enteritidis  
591 PT4 and *Salmonella Gallinarum* 287/91 provides insights into evolutionary and  
592 host adaptation pathways. *Genome Res* **18**, 1624-37 (2008).
- 593 33. Cooke, F.J. *et al.* Prophage sequences defining hot spots of genome variation in  
594 *Salmonella enterica* serovar Typhimurium can be used to discriminate between  
595 field isolates. *J Clin Microbiol* **45**, 2590-8 (2007).
- 596 34. Shah, D.H., Zhou, X., Kim, H.Y., Call, D.R. & Guard, J. Transposon mutagenesis  
597 of *Salmonella enterica* serovar Enteritidis identifies genes that contribute to  
598 invasiveness in human and chicken cells and survival in egg albumen. *Infect*  
599 *Immun* **80**, 4203-15 (2012).
- 600 35. Parkhill, J. *et al.* Complete genome sequence of a multiple drug resistant  
601 *Salmonella enterica* serovar Typhi CT18. *Nature* **413**, 848-52 (2001).

- 602 36. Thiennimitr, P. *et al.* Intestinal inflammation allows Salmonella to use  
603 ethanolamine to compete with the microbiota. *Proc Natl Acad Sci U S A* **108**,  
604 17480-5 (2011).
- 605 37. Kingsley, R.A. *et al.* Genome and transcriptome adaptation accompanying  
606 emergence of the definitive type 2 host-restricted Salmonella enterica serovar  
607 Typhimurium pathovar. *MBio* **4**, e00565-13 (2013).
- 608 38. Dhawi, A.A. *et al.* Adaptation to the chicken intestine in Salmonella Enteritidis  
609 PT4 studied by transcriptional analysis. *Vet Microbiol* **153**, 198-204 (2011).
- 610 39. Langridge, G.C. *et al.* Patterns of genome evolution that have accompanied host  
611 adaptation in Salmonella. *Proc Natl Acad Sci U S A* **112**, 863-8 (2015).
- 612 40. Feasey, N.A. *et al.* Modelling the Contributions of Malaria, HIV, Malnutrition and  
613 Rainfall to the Decline in Paediatric Invasive Non-typhoidal Salmonella Disease  
614 in Malawi. *PLoS Negl Trop Dis* **9**, e0003979 (2015).
- 615 41. Gordon, M.A. *et al.* Non-typhoidal salmonella bacteraemia among HIV-infected  
616 Malawian adults: high mortality and frequent recrudescence. *AIDS* **16**, 1633-41  
617 (2002).
- 618 42. Kotloff, K.L. *et al.* Burden and aetiology of diarrhoeal disease in infants and  
619 young children in developing countries (the Global Enteric Multicenter Study,  
620 GEMS): a prospective, case-control study. *Lancet* **382**, 209-22 (2013).
- 621 43. Kariuki, S. & Onsare, R.S. Epidemiology and Genomics of Invasive Nontyphoidal  
622 Salmonella Infections in Kenya. *Clin Infect Dis* **61 Suppl 4**, S317-24 (2015).

623

624

625

626

627

628

629

630

631

632

633 **Figures**

634

635 Figure 1: Maximum likelihood phylogeny of *S. Enteritidis* based on 675 isolates  
636 rooted to *S. Gallinarum*. There are 3 epidemic clades; 2 African epidemic clades and  
637 a global epidemic clade. Scale bar indicates nucleotide substitutions per site.

638

639 Figure 2: Differences in accessory genomes of 4 major clades. Approximate position  
640 of prophages in chromosome is depicted, although prophages are not drawn to scale

641

642 Figure 3: Heat map revealing changes in metabolic activity of Central/Eastern  
643 African clade isolate D7795 when compared to global epidemic isolate A1636 at 28  
644 and 37°C. The figure also displays whether there are corresponding mutations in  
645 genes related to the affected metabolic pathway. (NSSNP=Non-synonymous single  
646 nucleotide polymorphism, HDG = Hypothetically disrupted gene)

647

648 Figure 4: *Salmonella* isolation from a chick infection model demonstrates failure of  
649 Central/Eastern African clade isolate to invade chicken spleen (4A) and liver (4B) or  
650 to colonize chicken caeca (4C) at 7 days post infection (dpi) (n=24 at this time point)  
651 compared to the global epidemic clade. Numbers are expressed as colony forming  
652 units (CFU) per gram of tissue

653

654

655

656

657

658

659

660 **Tables**

661 Table 1: Summary of metadata (n) by region in numbers

662

Region	Total	Site of isolation			Antimicrobial resistance phenotype				
		Human Invasive	Human non-invasive	Food/Animal/Environment	Drug susceptible	Resistant to 1-2 1st line	MDR*	Fluoroquinolone	ESBL†
<b>Asia</b>	11	5	5	1	0	0	0	0	0
<b>Europe</b>	61	0	16	24	2	0	0	0	0
<b>South America</b>	27‡	3	6	7	8	0	0	0	0
<b>North Africa</b>	12	9	1	1	9	0	0	2	0
<b>Sub-Saharan Africa</b>	353	269	22	7	99	64	14	0	3
<b>Republic of South Africa</b>	131	57	74	0	83	44	4	0	0

663

664 \*Multidrug resistant: resistant to ≥3 antimicrobials

665 †Extended spectrum beta lactamase producing

666 ‡Uruguay strains previously characterised by Betancor <sup>44</sup>

667



668 Table 2: Metadata summarised by clade

Major Clade/cluster	Site of isolation					Number (%) of antimicrobial resistance genes*		
	N (%)					1-3	4-6	7-9
	Human Invasive	Human non-invasive	Food/Animal/Environment	Unknown				
<b>West African</b>	61 (92)	1 (2)	0 (0)	4 (6)	22 (33)	9 (14)	35 (66)	
<b>Central/Eastern African</b>	155 (93)	7 (4)	0 (0)	5 (3)	0 (0)	11 (7)	156 (93)	
<b>Global epidemic</b>	94 (38)	95 (38)	31 (12)	30 (12)	243 (97)	7 (3)	0 (0)	
<b>Outlier cluster</b>	51 (38)	36 (27)	27 (20)	20 (15)	128 (96)	3 (2)	3 (2)	

669

670 \*All isolates contained cryptic aminoglycoside acetyltransferase gene *aac(6')-ly<sup>26</sup>*

671

672

673

674

675

676

677 **Online Methods**

678

679 Bacterial Isolates

680

681 *S. Enteritidis* isolates were selected on the basis of six factors; date of original  
682 isolation, antimicrobial susceptibility pattern, geographic site of original isolation,  
683 source (human [invasive vs stool], animal or environmental), phage type (where  
684 available), and multilocus variable number tandem repeat (MLVA) type (where  
685 available). *S. Enteritidis* P125109 (EMBL accession no. [AM933172](#)) isolated from a  
686 poultry farm from the UK was used as a reference<sup>32</sup>. The full metadata are in  
687 Supplementary Table 1. Isolates have been attributed to region according to United  
688 Nations statistical divisions.

689

690 Sequencing, SNP-calling, construction of phylogeny and comparative genomics

691

692 PCR libraries were prepared from 500 ng of DNA as previously described<sup>45</sup>. Isolates  
693 were sequenced using Illumina GA II, HiSeq 2000 and MiSeq machines (Illumina,  
694 San Diego, CA, USA) and 150 bp paired-end reads were generated. The strains were  
695 aligned to *Salmonella* Enteritidis reference genome P125109 using a pipeline  
696 developed in-house at the Wellcome Trust Sanger Institute (WTSI). For each isolate  
697 sequenced, the raw sequence read pairs were split to reduce the overall memory  
698 usage and allow reads to be aligned using more than one CPU. The reads were then  
699 aligned using SMALT, a hashing based sequence aligner. The aligned and unmapped  
700 reads were combined into a single BAM file. Picard was used to identify and flag  
701 optical duplicates generated during the making of a standard Illumina library, which  
702 reduces possible effects of PCR bias. All of the alignments were created in a  
703 standardized manner, with the commands and parameters stored in the header of  
704 each BAM file, allowing for the results to be easily reproduced.

705

706 The combined BAM file for each isolate was used as input data in the SAMtools

707 mpileup program to call SNPs and small indels, producing a BCF file describing all of  
708 the variant base positions<sup>46</sup>. A pseudo-genome was constructed by substituting the  
709 base call at each variant or non-variant site, defined in the BCF file, in the reference  
710 genome. Only base calls with a depth of coverage >4 or quality >50 were considered  
711 in this analysis. Base calls in the BCF file failing this quality control filter were  
712 replaced with the "N" character in the pseudo-genome sequence.

713

714 All of the software developed is freely available for download from GitHub under an  
715 open source license, GNU GPL 3.

716

717 Phylogenetic modelling was based on the assumption of a single common ancestor,  
718 therefore variable regions where horizontal genetic transfer occurs were  
719 excluded<sup>47,48</sup>. A maximum likelihood (ML) phylogenetic tree was then built from the  
720 alignments of the isolates using RAxML (version 7.0.4) using a GTR+I+G model<sup>49</sup>.

721 The maximum-likelihood phylogeny was supported by 100 bootstrap pseudo-  
722 replicate analyses of the alignment data. Clades were predicted using Hierarchical  
723 Bayesian Analysis of Population Structure (HierBAPS)<sup>24</sup>. This process was repeated  
724 to construct the plasmid phylogeny, using reads that aligned to pSENV.

725 To ascertain the presence of the clusters defined by HierBAPS in the Public Health  
726 England (PHE) routine *Salmonella* surveillance collection, seventeen isolates  
727 representing the diversity of the collection were compared against 2986 *S.*

728 Enteritidis PHE genomes. Single linkage SNP clustering was performed as  
729 previously described<sup>50</sup>. A maximum-likelihood phylogeny showing the integration  
730 of the seventeen isolates with 50-SNP cluster representatives of the PHE *S.*

731 Enteritidis collection was constructed as above. FASTQ reads from all PHE  
732 sequences in this study can be found at the PHE Pathogens BioProject at the  
733 National Center for Biotechnology Information (Accession PRJNA248792).

734

735 Temporal reconstruction was performed using Bayesian Evolutionary Analysis  
736 Sampling Trees (BEAST version 1.8.2)<sup>51</sup>. A relaxed lognormal clock model was

737 initially employed. The results of this model indicated that a constant clock model  
738 was not appropriate, as the posterior of the standard deviation of the clock rate did  
739 not include zero. A range of biologically plausible population models (constant,  
740 exponential and skyline) was investigated. Skyline models can be biased by non-  
741 uniform sampling and we observed a strong similarity between reconstructed  
742 skyline population and the histogram of sampling dates and so this model was  
743 excluded. The exponential models consistently failed to converge and were  
744 excluded. Thus, for all datasets, lognormal clock and constant population size  
745 models were used. The computational expense required for this analysis precluded  
746 running estimators for model selection. However, we note that Deng et al used the  
747 same models in their analysis of 125 *S. Enteritidis* isolates. Default priors were used  
748 except for `uclid.mean`, `Gamma(0.001,1000)`, `initial: 0.0001`; `exponential.popSize`,  
749 `LogNormal(10,1.5)`, `initial: 121`.

750

751 Three chains of 100 million states were run in parallel for each clade of the four  
752 major HierBAPS clades, as well as a fourth chain without genomic data to examine  
753 the influence of the prior, which in all cases was uninformative. The final results, as  
754 used here, all had effective sample sizes (ESS) of over 200 and had convergence  
755 between all three runs. For the Global and Global Outlier lineages, the datasets were  
756 not computationally feasible to analyse. We thus created 3 further random subsets  
757 of the data by drawing *n* isolates from each sampled year where *n* was sampled  
758 from a Poisson distribution where  $\lambda=2$ . The posteriors of all subsets were extremely  
759 similar and runs were combined to produce the final most recent common ancestor  
760 (MRCA) estimates.

761

762 In order to gain a detailed insight into genomic differences, a single high quality  
763 sequence from Malawian *S. Enteritidis* isolate D7795 was aligned against the  
764 P125109 using ABACAS and annotated<sup>52</sup>. Differences were manually curated against  
765 the reference using the Artemis Comparison Tool (ACT)<sup>53</sup>. Sections of contigs which  
766 were incorporated into the alignment, but which did not align with P125109 were

767 manually inspected and compared to the public databases using BLASTn. When  
768 these regions appeared to be novel prophages, they were annotated using the phage  
769 search tool PHAST and manually curated<sup>54</sup>. In order to investigate whether the SNPs  
770 and/or indels that were predicted to be responsible for pseudogene formation in  
771 D7795 were distinct to that isolate or conserved across both African epidemic  
772 clades, all isolates were aligned to P12509 and the relevant SNPs/indels  
773 investigated using *in-silico* PCR of the aligned sequences. Manual curation was  
774 performed to confirm the nature of all pseudogene-associated SNPs/indels. NS-SNPs  
775 identified in D7795 were sorted throughout the African clades by extracting and  
776 aligning the appropriate gene sequences from P125109 and D7795. The coordinates  
777 of the NS-SNPs were then used to identify the relevant sequence and determine the  
778 nature of the base.

779

#### 780 Accessory genome

781 The pangenome for the dataset was predicted using ROARY<sup>55</sup>. Genes were  
782 considered to be core to *S. Enteritidis* if present in  $\geq 90\%$  of isolates. A relaxed  
783 definition of core genome was used as assemblies were used to generate it and the  
784 more assemblies one uses, the more likely it is that a core gene will be missed in one  
785 sample due to an assembly error. The remaining genes were considered to be core  
786 to the clades/clusters predicted by HierBAPS if present in  $\geq 75\%$  of isolates from  
787 within each clade/cluster. These genes were then curated manually using ACT to  
788 search for their presence and position in P125109 or the improved draft assembly  
789 of representative isolates of each of the other clades if not present in P125109. Any  
790 large accessory regions identified were blasted against the assembled genomes of  
791 the entire collection to confirm they were grossly intact.

792

#### 793 Plasmid identification

794 Plasmid DNA was extracted from isolate D7795 using the Kado & Liu method and  
795 separated by gel-electrophoresis alongside plasmids of known size, to estimate the  
796 number and size of plasmids present<sup>56</sup>. Plasmid conjugation was attempted by

797 mixing 100  $\mu$ L of overnight culture of donor and recipient strains (rifampicin  
798 resistant *Escherichia coli* C600) on Luria-Bertani agar plates and incubating  
799 overnight at 26°C and 37°C. The plasmid was sequenced using the PacBio platform  
800 to gain long reads and a single improved draft assembly, which was aligned against  
801 P125109 plasmid pSENV (Accession Number HG970000). For novel regions of the  
802 plasmid from isolate D7795, genes were predicted using GLIMMER and manual  
803 annotations applied based on homology searches against the public databases, using  
804 both BLASTn and FASTA. The plasmid phylogeny was reconstructed using the same  
805 methodology as the chromosome; a maximum likelihood (ML) phylogenetic tree  
806 was built from the alignments of the isolates using RAxML (version 7.0.4) using a  
807 GTR+I+G model

808

809 Identification of AMR genes

810

811 A manually curated version of the Resfinder database was used to investigate the  
812 isolates for the presence of AMR genes<sup>57</sup>. To reduce redundancy, the database was  
813 clustered using CD-HIT-EST<sup>58</sup>, with the alignment length of the shorter sequence  
814 required to be 90% the length of the longer sequence. All other options were left as  
815 the defaults. The representative gene of each cluster was then mapped with SMALT  
816 to the assemblies of each isolate to identify and matches with an identity of 90% or  
817 greater were considered significant, in line with the default clustering parameters of  
818 CD-HIT-EST. Where partial matches were identified at the ends of contigs, having an  
819 identity of 90% or greater to the matched region of the gene, potential AMR gene  
820 presence was recorded. To confirm presence of these partial matches, raw  
821 sequencing reads of the pertinent isolates were mapped using SMALT to these genes  
822 to check for 90% identity across the entire gene.

823

824 Biolog™ growth substrate platform profiling

825

826 The Biolog™ platform enables the simultaneous quantitative measurement of a  
827 number of cellular phenotypes, and therefore the creation of a phenotypic profile of  
828 a variety of assay conditions<sup>59</sup>. Incubation and recording of phenotypic data were  
829 performed using an OmniLog® plate reader.

830

831 In these experiments, two replicates of D7795 were compared to two replicates of a  
832 PT4 like strain at 28°C and 37°C to represent environmental and human  
833 temperatures. Biolog™ plates PM1-4 and 9 (Carbon source [PM1,PM2], nitrogen  
834 source [PM3] and phosphor and sulphur source [PM4] metabolism and osmotic  
835 pressure [PM9]) were used. Each well was inoculated as described below, thereby  
836 testing 475 conditions at once (each plate has one negative control well).

837

838 The isolates were cultured overnight on LB-agar at 37°C in air to exclude  
839 contamination. Colonies were scraped off plates and dispensed into IF-0a solution  
840 (Biolog) to a cell density corresponding to 81% transmittance. For each plate used,  
841 880 µL of this cell suspension was added to 10 mL IF-10b GP/GP solution (Biolog)  
842 and 120 µL dye mix G (Biolog). This was then supplemented with a 1 mL solution of  
843 7.5 mM D-ribose (Sigma), 2 mM magnesium chloride, 1 mM calcium chloride, 2 mM  
844 sodium pyrophosphate (Sigma), 25 µM L-arginine (Sigma), 25 µM L-methionine  
845 (Sigma), 25 µM hypoxanthine (Sigma), 10 µM lipoamide (Sigma), 5 µM nicotine  
846 adenine dinucleotide (Sigma), 0.25 µM riboflavin (Sigma), 0.005% by mass yeast  
847 extract (Fluka) and 0.005% by mass Tween 80 (Sigma). 100µl of this mixture was  
848 dispensed into each well on the assay plate. Plates were then allowed to equilibrate  
849 in air for 5 min prior to being sealed in airtight bags and loaded into the Omnilog  
850 machine (Biolog). Plates were scanned every 15 min for 48 hours while incubated at  
851 28°C and 37°C in air. Culture under anaerobic conditions was unavailable. Two  
852 paired replicates were performed for each of the two isolates.

853 After completion of the run, the signal data were compiled and analysed using the  
854 limma package in 'R' described previously<sup>60</sup>. A log-fold change of 0.5 controlling for

855 a 5% false discovery rate was used as a cut-off for investigating a specific metabolite  
856 further using Pathway Tools<sup>61</sup> and whether the metabolic change was related to  
857 pseudogenes and non synonymous(NS)-SNPs in genes in the respective genomes.

858

#### 859 *In vivo* Infection Model

860

861 Two isolates were used in the animal models: *S. Enteritidis* P125109 and D7795.  
862 Unvaccinated commercial female egg-layer Lohmann Brown chicks (Domestic Fowl  
863 [*Gallus gallus*]) were obtained from a commercial hatchery and housed in secure  
864 floor pens at a temperature of 25°C. Eight chicks per strain per time point were  
865 inoculated by gavage at 10 days (d) of age and received a dose of  $\sim 10^8$  *Salmonella*  
866 colony forming units (CFU) in a volume of 0.2 mL. Subsequently, four to five birds  
867 from each group were humanely killed at 3, 7 or 21 d post-infection (p.i.). At post  
868 mortem, the liver, spleen, and caecal contents were removed aseptically,  
869 homogenised, serially diluted and dispensed onto Brilliant Green agar (Oxoid) to  
870 quantify colony forming units (CFU) as described previously<sup>62</sup>. Statistical analysis  
871 was performed using SPSS, version 20 (IBM). Kruskal-Wallis was used to compare  
872 bacterial loads between infected groups.

873

874 All work was conducted in accordance with the UK legislation governing  
875 experimental animals, Animals (Scientific Procedures) Act 1986, under project  
876 licence 40/3652 and was approved by the University of Liverpool ethical review  
877 process prior to the award of the project license. The licensing procedure requires  
878 power calculations to determine minimal group sizes for each procedure to ensure  
879 results are significant. For these experiments a group size of 8 birds per time point  
880 was chosen, based on a variation in  $1.0 \log_{10}$  in bacterial count between groups as  
881 being significant along with prior experience of *Salmonella* infection studies. Groups  
882 were randomly selected on receipt from the hatchery and investigators conducting  
883 animal experiments were not blinded, as the current UK code of practice requires all  
884 cages or pens to be fully labeled with experimental details. No animals were



885 excluded from the analysis. All animals were checked a minimum of twice daily to  
886 ensure their health and welfare.

887

888 Code availability & URLs

889

890 Software is referenced and URLs are provided below. All software is open source.

891

892 BEAST: <http://beast.bio.ed.ac.uk/>

893 Biolog™: <http://www.biolog.com>

894 BLASTn: <http://blast.ncbi.nlm.nih.gov>

895 limma package: [www.bioconductor.org](http://www.bioconductor.org)

896 PacBio platform: <http://www.pacificbiosciences.com/>

897 Picard: <https://broadinstitute.github.io/picard>

898 'R': [www.R-project.org](http://www.R-project.org)

899 SMALT: [www.sanger.ac.uk/science/tools/smalt-0](http://www.sanger.ac.uk/science/tools/smalt-0)

900 United Nations statistical divisions:

901 [www.unstats.un.org/unsd/methods/m49/m49regin.htm](http://www.unstats.un.org/unsd/methods/m49/m49regin.htm)

902 Supplementary References

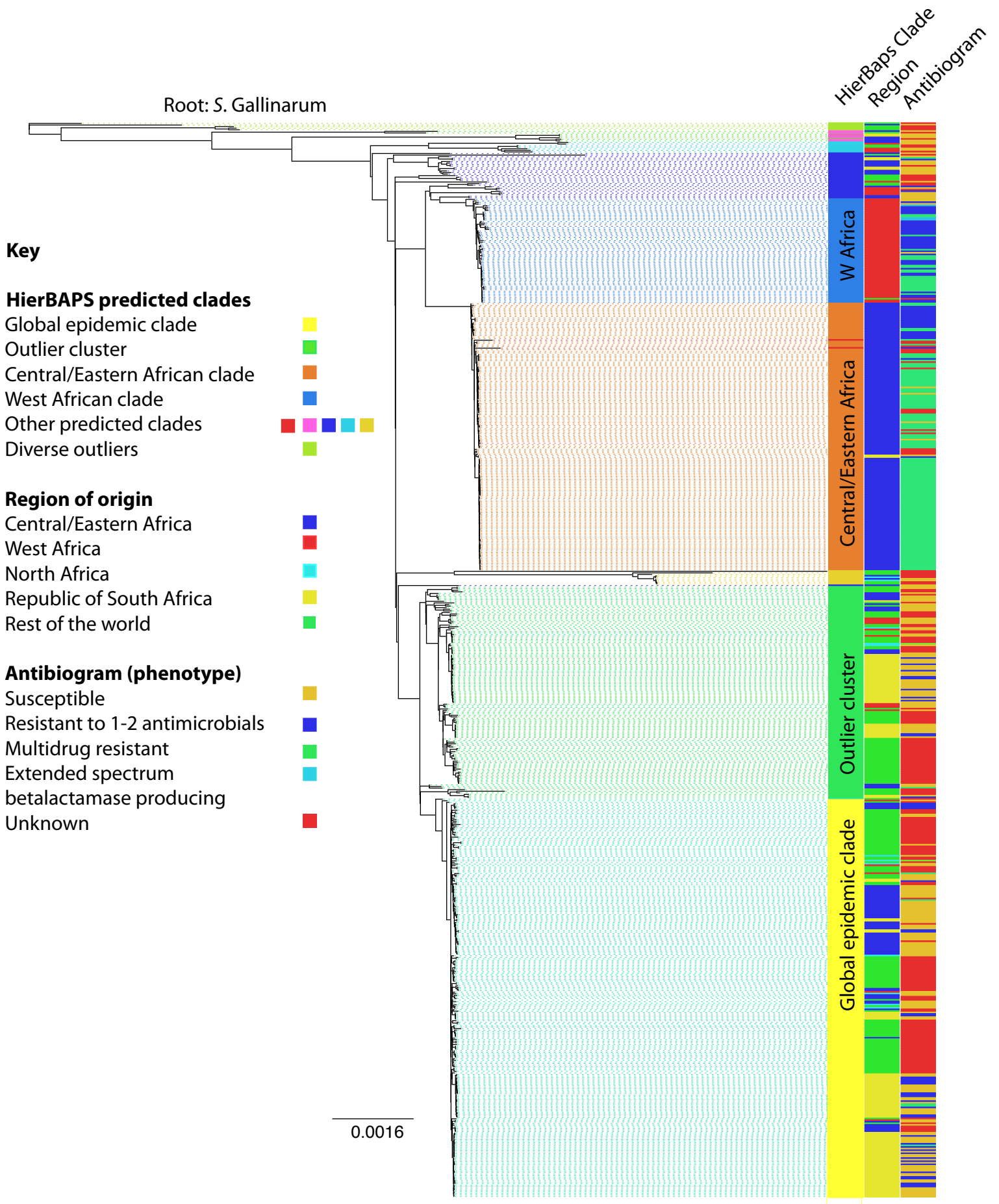
- 903 44. Betancor, L. *et al.* Genomic and phenotypic variation in epidemic-spanning  
904 *Salmonella enterica* serovar Enteritidis isolates. *BMC Microbiol* **9**, 237 (2009).
- 905 45. Harris, S.R. *et al.* Whole-genome sequencing for analysis of an outbreak of  
906 meticillin-resistant *Staphylococcus aureus*: a descriptive study. *Lancet Infect Dis*  
907 **13**, 130-6 (2013).
- 908 46. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics*  
909 **25**, 2078-9 (2009).
- 910 47. Holt, K.E. *et al.* High-throughput sequencing provides insights into genome  
911 variation and evolution in *Salmonella* Typhi. *Nat Genet* **40**, 987-93 (2008).
- 912 48. Croucher, N.J. *et al.* Rapid pneumococcal evolution in response to clinical  
913 interventions. *Science* **331**, 430-4 (2011).
- 914 49. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic  
915 analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-90  
916 (2006).
- 917 50. Philip Ashton, S.N., Tansy Peters, Rediat Tewolde, Martin Day, Michel  
918 Doumith, Jonathan Green, Claire Jenkins, Anthony Underwood, Catherine  
919 Arnold, Elizabeth de Pinna, Tim Dallman, Kathie Grant. Revolutionising  
920 Public Health Reference Microbiology using Whole Genome Sequencing:  
921 *Salmonella* as an exemplar. *bioRxiv* (2015).
- 922 51. Drummond, A.J., Suchard, M.A., Xie, D. & Rambaut, A. Bayesian phylogenetics  
923 with BEAUti and the BEAST 1.7. *Mol Biol Evol* **29**, 1969-73 (2012).
- 924 52. Assefa, S., Keane, T.M., Otto, T.D., Newbold, C. & Berriman, M. ABACAS:  
925 algorithm-based automatic contiguation of assembled sequences. *Bioinformatics*  
926 **25**, 1968-9 (2009).
- 927 53. Carver, T.J. *et al.* ACT: the Artemis Comparison Tool. *Bioinformatics* **21**, 3422-3  
928 (2005).
- 929 54. Zhou, Y., Liang, Y., Lynch, K.H., Dennis, J.J. & Wishart, D.S. PHAST: a fast  
930 phage search tool. *Nucleic Acids Res* **39**, W347-52 (2011).
- 931 55. Page, A.J. *et al.* Roary: Rapid large-scale prokaryote pan genome analysis.  
932 *Bioinformatics* (2015).
- 933 56. Kado, C.I. & Liu, S.T. Rapid procedure for detection and isolation of large and  
934 small plasmids. *J Bacteriol* **145**, 1365-73 (1981).

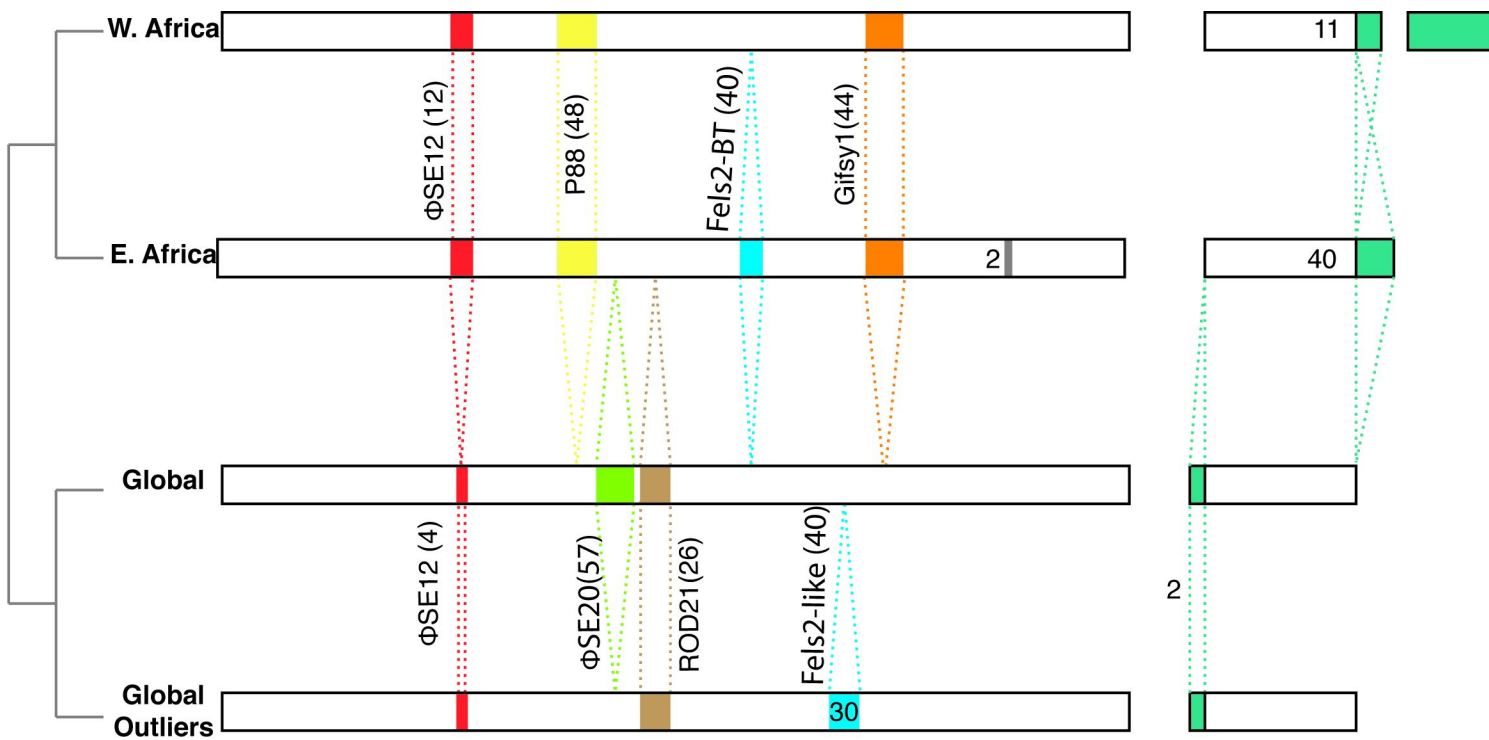
- 935 57. Zankari, E. *et al.* Identification of acquired antimicrobial resistance genes. *J*  
936 *Antimicrob Chemother* **67**, 2640-4 (2012).
- 937 58. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets  
938 of protein or nucleotide sequences. *Bioinformatics* **22**, 1658-9 (2006).
- 939 59. Tracy, B.S., Edwards, K.K. & Eisenstark, A. Carbon and nitrogen substrate  
940 utilization by archival *Salmonella typhimurium* LT2 cells. *BMC Evol Biol* **2**, 14  
941 (2002).
- 942 60. Croucher, N.J. *et al.* Dominant role of nucleotide substitution in the diversification  
943 of serotype 3 pneumococci over decades and during a single infection. *PLoS*  
944 *Genet* **9**, e1003868 (2013).
- 945 61. Karp, P.D. *et al.* Pathway Tools version 13.0: integrated software for  
946 pathway/genome informatics and systems biology. *Brief Bioinform* **11**, 40-79  
947 (2010).
- 948 62. Salisbury, A.M., Bronowski, C. & Wigley, P. *Salmonella* Virchow isolates from  
949 human and avian origins in England--molecular characterization and infection of  
950 epithelial cells and poultry. *J Appl Microbiol* **111**, 1505-14 (2011).

951

952 Competing financial interests:

953 The authors declare no competing financial interests.





- Gifsy-1
- $\Phi$ SE12
- ROD-21
- plasmid
- Enterobacter p88
- $\Phi$ SE20
- Fels-2 like prophage regions
- pseudogenes

