

Elucidation of global and local genomic epidemiology of *Salmonella enterica* serovar Enteritidis through multilevel genome typing

Lijuan Luo¹, Michael Payne¹, Sandeep Kaur¹, Dalong Hu¹, Liam Cheney¹, Sophie Octavia¹, Qinning Wang², Mark M. Tanaka¹, Vitali Sintchenko^{2,3} and Ruiting Lan^{1,*}

¹School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, New South Wales, Australia

²Centre for Infectious Diseases and Microbiology–Public Health, Institute of Clinical Pathology and Medical Research – NSW Health Pathology, Westmead Hospital, New South Wales, Australia

³Marie Bashir Institute for Infectious Diseases and Biosecurity, Sydney Medical School, University of Sydney, New South Wales, Australia

*Corresponding Author

Email: r.lan@unsw.edu.au

Phone: 61-2-9385 2095

Fax: 61-2-9385 1483

1 **Keywords:** *Salmonella enterica* serovar Enteritidis; MGT; genomic epidemiology; population
2 structure; global database; virulence

3

4 **Repositories:** There were no newly sequenced data in this study.

5

6 **Abstract**

7 *Salmonella enterica* serovar Enteritidis is a major cause of foodborne *Salmonella* infections and
8 outbreaks in humans. Effective surveillance and timely outbreak detection are essential for public
9 health control. Multilevel genome typing (MGT) with multiple levels of resolution has been
10 previously demonstrated as a promising tool for this purpose. In this study, we developed MGT with
11 nine levels for *S. Enteritidis* and characterised the genomic epidemiology of *S. Enteritidis* in detail.
12 We examined 26,670 publicly available *S. Enteritidis* genome sequences from isolates spanning 101
13 years from 86 countries to reveal their spatial and temporal distributions. Using the lower resolution
14 MGT levels, globally prevalent and regionally restricted sequence types (STs) were identified; avian
15 associated MGT4-STs were found that were common in human cases in the USA were identified;
16 temporal trends were observed in the UK with MGT5-STs from 2014 to 2018, revealing both long
17 lived endemic STs and the rapid expansion of new STs. Using MGT3 to MGT6, we identified MDR
18 associated STs at various MGT levels, which improves precision of detection and global tracking of
19 MDR clones. We also found that the majority of the global *S. Enteritidis* population fell within two
20 predominant lineages, which had significantly different propensity of causing large scale outbreaks.
21 An online open MGT database has been established for unified international surveillance of *S.*
22 *Enteritidis*. We demonstrated that MGT provides a flexible and high-resolution genome typing tool
23 for *S. Enteritidis* surveillance and outbreak detection.

24

25 **Impact statement**

26 *Salmonella enterica* serovar Enteritidis is a common foodborne pathogen that can cause large
27 outbreaks. Surveillance and high-resolution typing are essential for outbreak prevention and control.
28 Genome sequencing offers unprecedented power for these purposes and a standardised method or
29 platform for the interpretation, comparison and communication of genomic typing data is highly
30 desirable. In this work, we developed a genomic typing scheme called Multilevel Genome Typing
31 (MGT) for *S. Enteritidis*. We analysed 26,670 publicly available genomes of *S. Enteritidis* using

32 MGT. We characterised the geographic and temporal distribution of *S. Enteritidis* MGT types as well
33 as their association with multidrug resistance (MDR) and virulence genes. A publicly available MGT
34 database for *S. Enteritidis* was established, which has the potential facilitate the unified global public
35 health surveillance for this pathogen.

36

37 **Abbreviations**

38 MGT: Multilevel genome typing; ST: Sequence type; CC: Clonal complex; ODC: Outbreak
39 detection cluster; SNP: Single nucleotide polymorphism; MLST: Multi-locus sequence typing;
40 cgMLST: core genome multi-locus sequence typing; HierCC: Hierarchical clustering of cgMLST.
41 AR: Antibiotic resistance; MDR: Multi-drug resistance.

42

43 **Data Summary**

- 44 1. The MGT database for *S. Enteritidis* is available at <https://mgtdb.unsw.edu.au/enteritidis/>.
- 45 2. All accession numbers of the public available genomes were available in the MGT database and
46 Data Set S1, Tab 1. And there were no newly sequenced data in this study.
- 47 3. Supplementary material: Supplementary Fig. S1 to S7, supplementary methods and supporting
48 results about the evaluation of potential repeat sequencing bias.
- 49 4. Data Set S1: Supporting tables of the main results.
- 50 5. Data Set S2. Supporting tables of the repeat sequencing bias evaluation by removing the potential
51 repeat sequencing isolates. Note outbreak isolates may also be removed.

52 **Introduction**

53 Nontyphoidal *Salmonella* is ranked second in foodborne disease in Europe and North America [1, 2],
54 with *Salmonella enterica* serovar Enteritidis a dominant serovar in many countries [3, 4]. *S.*
55 Enteritidis mainly causes human gastrointestinal infections leading to diarrhoea. However, invasive

56 infections manifesting as sepsis, meningitis and pneumonia, have recently been reported [5, 6]. As
57 many *S. Enteritidis* strains causing invasive infections carry plasmids which confer multidrug
58 resistance (MDR), the spread of these strains internationally is a major threat to global health [7, 8].
59 Additionally, *S. Enteritidis* has caused numerous large-scale national and international outbreaks
60 with complex transmission pathways [3, 9, 10]. While there have been limited studies of the total
61 global epidemiology of *S. Enteritidis* genomic subtypes [5, 11], the geographic distribution, outbreak
62 propensity, MDR and evolutionary characteristics of different lineages of *S. Enteritidis* have not been
63 systematically evaluated using the large numbers of newly available public genome sequences. A
64 rapid, stable and standardized global genomic typing strategy for *S. Enteritidis* is required for the
65 high-resolution and scalable surveillance of outbreaks, tracking of international spread and MDR
66 profiles of *S. Enteritidis*.

67 Sequence types (STs) are based on exact matching of genes between isolates. The well-established
68 seven gene multi-locus sequence typing (MLST) of *Salmonella* is widely used [12], while the vast
69 majority of *S. Enteritidis* are assigned to ST11, as only seven housekeeping genes are compared. STs
70 of higher resolution are required. With this in mind core genome MLST (cgMLST) and whole
71 genome MLST (wgMLST) schemes of *Salmonella* were developed including 3002 and 21,065 loci,
72 respectively [13]. However, STs based on cgMLST and wgMLST are too diverse to offer useful
73 relatedness information at anything but the finest scales. This issue was addressed by single linkage
74 hierarchic typing systems, i.e., Single Nucleotide Polymorphism (SNP) address and Hierarchical
75 Clustering of cgMLST (HierCC) [14-16]. SNP address is based on the single linkage hierarchic
76 clustering method, which allows for different SNP differences 250, 100, 50, 25, 10, 5 and 0 SNPs
77 [15]. These methods had been successfully used in outbreak tracing and epidemiology [14, 17-19].
78 However, one major disadvantage of the pairwise comparison based single linkage clustering method
79 is that the cluster types called may merge when additional data is added and differences in order of
80 addition can result in different clusters [20]. For example, when an isolate meets the SNP/allele

81 difference threshold to two clusters, this isolate would act as a bridge connecting the two clusters and
82 merging them into a single cluster. This scenario will almost certainly occur in a large database with
83 ever expanding numbers of sequences, especially at finer resolutions, and would be an obstacle for
84 long term epidemiological investigation and comparison [20]. We have recently developed a novel
85 core genome based method called multilevel genome typing (MGT), an exact matching method to
86 assign multiple resolutions of relatedness without the need for clustering [20]. MGT is a genome
87 sequence based typing system, including multiple MLST schemes with increasing resolution from
88 the classic seven gene MLST to cgMLST where each scheme is used to independently assign an ST.
89 An MGT scheme and a public database were established for *Salmonella enterica* serovar
90 Typhimurium where seven to 5268 loci were included in MGT1 to 9 [20]. MGT1 corresponds to the
91 classic seven gene MLST scheme of *S. enterica* [12]. For *S. Typhimurium*, the stable STs from MGT
92 levels of appropriate resolution were found to be useful in identifying DT104, the MDR lineage, as
93 well as African invasive lineages [20]. MGT of *S. Typhimurium* was also shown to perform well in
94 outbreak identification and source tracing [20].

95 Here, we describe an MGT scheme and public database for *S. Enteritidis*. By using MGT, we
96 systematically examine the global and local epidemiology of *S. Enteritidis* over time, and evaluate
97 the application of this scheme to the outbreak evaluation and the monitoring of multidrug resistant
98 STs.

99 **Methods**

100 ***S. Enteritidis* sequences used in this study**

101 Raw reads of 26,670 *S. Enteritidis* whole genome sequences (WGS) downloaded from the European
102 Nucleotide Archive (ENA) passed the quality filtering criteria by Quast [21, 22] (**Data Set S1, Tab**
103 **1**). The epidemiological metadata of those isolates were collated from NCBI and EnteroBase [13].
104 Trimmomatic v0.39 was used with default settings to trim raw reads [23]. The trimmed raw reads

105 were then assembled using either SPAdes v3.13.0 for core genome definition or SKESA v2.3 for
106 MGT typing [24, 25]. Quality assessment for the assemblies were performed using Quast v5.0.2
107 [21]. Thresholds of the assembly quality were in accordance with Robertson *et al* 's criteria [22].
108 Kraken v0.10.5 was used to identify contamination in the assembled genomic sequences or
109 problematic taxonomic identification [26]. SeqSero v1.0 and SISTR v1.0.2 were used to verify
110 serotype assignments [27, 28].

111 **Core gene definition of *S. Enteritidis***

112 To define the core genome of *S. Enteritidis*, a total of 1801 genomes were selected based on the
113 ribosomal sequence type [13] and strain metadata. The trimmed raw reads were then assembled
114 using SPAdes v3.13.0 [24]. The assemblies were annotated with Prokka v1.12[29]. The core genes
115 and core intergenic regions were defined with Roary v3.11.2 and Piggy v3.11.2, respectively (Fig.
116 S1). A total of 113 sRNA regions of SL1344 were searched in the 1054 intergenic regions using
117 BLAST, with identity and coverage threshold of 70% and 60%, respectively [30]. Both SeqSero v1.0
118 and SISTR v1.0.2 were used to verify whether the isolates were *S. Enteritidis* [27, 28]. The detailed
119 methods for isolates selection and core genome identification were described in Supplementary
120 material.

121 **Establishment of MGT scheme for *S. Enteritidis* and MGT allele calling**

122 The initial allele sequence of each locus used for MGT were extracted from the complete genome
123 sequence of *S. Enteritidis* strain P125109 (Genbank accession NC_011294.1). The first eight levels
124 of MGT were identical to the MGT scheme for *S. Typhimurium* [20], in which MGT1 refers to the
125 classical seven gene MLST for the *Salmonella* genus [12]. Loci used for levels MGT2 to MGT8
126 were orthologous to those in the *S. Typhimurium* MGT with the exception of one gene which was
127 removed at MGT8 because of a paralogue in *S. Enteritidis*. MGT9 contains all loci in MGT8 as well
128 as the core genes and core intergenic regions of *S. Enteritidis*.

129 The raw read processing and genome assembly procedures were the same as above, except that
130 SKESA v2.3 was used to assemble the raw reads and only SISTR v1.0.2 was used for the molecular
131 serotype identification [25, 28]. SKESA v2.3 offers higher per base accuracy and speed than
132 SPADES and was therefore used in the allele calling pipeline [25]. For the assemblies that passed
133 quality control by the criteria of Robertson *et al.* [22], a custom script was used to assign alleles, STs
134 and clonal complexes (CCs) to those isolates at levels from MGT2 to MGT9 [20]. Allele, ST, CC
135 and outbreak detection cluster (ODC) assignments were performed as previously described [20].
136 Briefly, if one or more allele difference was observed, a new ST was assigned to the isolate. The
137 same strategy was also used for assigning a new CC, if at least two allele differences were observed
138 comparing to any of STs in existing CCs in the database [20]. ODC is an extension of the CC
139 clustering method, but it was performed only at MGT9 level [20]. For example, any two MGT9 STs
140 with no more than 2, 5, or 10 allele differences were assigned the same ODC2, ODC5 or ODC10
141 type [20].

142 **Geographic distribution and temporal variation analysis**

143 We first determined which MGT level would be the most useful in describing the global
144 epidemiology of *S. Enteritidis* by summarizing STs that contained at least 10 isolates for their
145 distribution across continents. STs where more than 70% of isolates were from a single continent
146 were identified at each level of MGT (Fig. 2a). MGT4 STs matching this criterion had the highest
147 percentage of all the levels at 76.3% of the 26,670 isolates, thus MGT4 was chosen to describe the
148 continental distribution of *S. Enteritidis*. The USA states variation map and the UK monthly variation
149 column chart were produced using Tableau v2019.2 [31].

150 **Antibiotic resistance gene and plasmid specific gene identification**

151 Abricate v1.0.1 (<https://github.com/tseemann/abricate>) was used for identification of antibiotic
152 resistance genes with the ResFinder v3.0 database [32], and plasmid specific genes with the

153 PlasmidFinder v2.1 database [33]. The cut-off of the presence of a gene was set as both identity and
154 coverage $\geq 90\%$ [32]. At each MGT level, STs (≥ 10 isolates each) with more than 80% of isolates
155 harbouring antibiotic resistance genes, were collected from MGT2 to MGT9 and were defined as
156 high-antibiotic-resistant STs. The STs at each level of MDR were non-redundant meaning that no
157 isolate was counted more than once.

158 **Phylogenetic analysis**

159 A phylogenetic tree was constructed using Parsnp v1.2, which called core-genome SNPs [34, 35].
160 Potential recombinant SNPs were removed using both Gubbins v2.0.0 and Recdetect v6.0 [36, 37].
161 Beast v1.10.4 was then used to estimate the mutation rate based on mutational SNPs [38]. A total of
162 24 combinations of clock and population size models were evaluated with the MCMC chain of 100
163 million states. Tracer v1.7.1 was used to identify the optimal model and to estimate population
164 expansion over time [39]. The detailed methods for phylogenetic analysis were described in

165 **Supplementary material.**

166 **Virulence genes and *Salmonella* pathogenicity islands (SPIs) distribution in the *S. Enteritidis*** 167 **population**

168 We compared the presence of virulence determinants in the main lineages of *S. Enteritidis*. Virulence
169 genes from the Virulence Factor Database (VFDB) were identified in all the *S. Enteritidis* genomes
170 using Abricate v1.0.1 (<https://github.com/tseemann/abicate>), with identity threshold of 70% and
171 coverage of 50% [40]. Using the same blast threshold, a total of 23 reported SPIs from SPI-1 to SPI-
172 23 were also identified in all the *S. Enteritidis* genomes [41].

173 **Evaluation of the effect of repeat sequencing on the dataset for epidemiological analysis**

174 To evaluate any bias that may be caused by resequencing of the same isolate, we identified all
175 isolates of the same ST based on MGT9 and same metadata based on collection country, collection

176 year and month, and source type. Such isolates were conservatively treated as repeat sequencing of
177 the same isolate and such “duplicates” were removed from the dataset. We re-analysed the reduced
178 dataset and compared against the original dataset by calculating Kendall's tau [42]. The detailed
179 results are shown in Supplementary material and Data Set S2, Tab 1 to Tab 6.

180 **Results**

181 **Establishing the MGT system for *S. Enteritidis***

182 The core genome of *S. Enteritidis* was defined at $\geq 96\%$ identity and presence rate of $\geq 99\%$ of the
183 1801 sampled isolates (Fig. S1). The core genome of *S. Enteritidis* included 3932 genes, with 977 not
184 found in the *Salmonella enterica* core as well as 1054 *S. Enteritidis* core intergenic regions (Fig. 1a)
185 (Data Set S1, Tab 2). We also searched for the presence of small RNA (sRNA) in the intergenic
186 regions with 37 (32.7%, 37/113) sRNAs observed in 36 (3.4%, 36/1054) intergenic regions.

187 The first eight schemes of *S. Enteritidis* MGT used the same 2955 core genes of *Salmonella*
188 *enterica* as the MGT of *S. Typhimurium*, and the rationale of locus selection has been fully described
189 by Payne *et al* [20]. MGT1 corresponded to the classic seven gene MLST of *Salmonella* [12]. A total
190 of 4986 loci were incorporated in the MGT9 scheme, including the 3932 *S. Enteritidis* core genes
191 and 1054 core intergenic regions defined (Fig. 1a).

192 A total of 26,670 *S. Enteritidis* genomes with publicly available raw reads were analysed using the
193 MGT scheme. These publicly available genomes were collected from 26 source types with 49.9% of
194 isolates collected from humans and 8.9% from avian sources; they were collected from 86 countries
195 with the majority of the isolates from the United States (47.3%) and United Kingdom (35.9%); and
196 they were collected between the years 1917 and 2018 with 2014 (11.5%), 2015 (14.6%), 2016
197 (14.0%), 2017 (12.8%) and 2018 (3.7%) having more than 500 isolates each (Fig. S2a). At each
198 MGT level, each isolate was assigned an ST and a clonal complex (CC). A CC in this study was

199 defined as a group of STs with one allele difference [12, 43]. The number of STs and CCs at each
200 MGT level is shown in Fig. 1b. As the resolution of typing increased from MGT1 to MGT9, an
201 increasing number of STs and CCs were assigned. At MGT2, the 26,670 isolates were subtyped into
202 252 STs and 23 CCs. By contrast, MGT9 divided the isolates into 20,153 different STs and 14,441
203 CCs. The MGT scheme of *S. Enteritidis* is available through a public online database
204 (<https://mgtdb.unsw.edu.au/enteritidis/>).

205 **The international or global epidemiology of *S. Enteritidis* by MGT**

206 Of the 26,670 isolates typed by MGT, 25,207 have country metadata. By the 7-gene MLST (or
207 MGT1) scheme, ST11 was the dominant type representing 94.8% of the isolates, followed by ST183
208 representing 1.6%. ST183 is an endemic ST prevalent in Europe and was divided into two main
209 phage types, PT11 and PT66 [44]. Using MGT, ST183 (or MGT1-ST183) can be divided into seven
210 STs at MGT2 level. Interestingly, 97% (137/141) of the PT11 isolates belonged to MGT2-ST3, and
211 100% (24/24) of the PT66 belonged to MGT2-ST82, highlighting the potential for backward
212 compatibility of MGT with traditional typing data.

213 For the predominant ST11 isolates (or MGT1-ST11), we found that the optimal MGT level to
214 describe their global epidemiology was MGT4, based on the distribution of each ST in different
215 continents (Fig. 2a). At the MGT4 level, the 26,670 isolates were subtyped into 2,236 STs and 423
216 CCs. Among the 2,236 MGT4-STs, 163 STs were predominantly found in only one continent
217 (defined as $\geq 70\%$ from one continent). These 163 STs contained 20,341 of the 26,670 genomes
218 (76.3%).

219 The distribution of STs varied between continents. The following MGT4-CC1 STs, ST171, ST163,
220 ST370, ST1009, and MGT4-CC13 STs, ST99, ST136, ST135, ST396, ST198, ST160, ST416 were
221 the most prevalent STs in North America (Fig. 2b). While in Europe, the following MGT4-CC1 STs,
222 ST15, ST208, ST357, ST237 and MGT4-CC13 STs, ST25, ST13, ST100, ST31, ST29 were

223 common. However, the North America and Europe *S. Enteritidis* sequences analysed were mostly
224 obtained from the USA and UK, which represented 94.6% and 92.5% of isolates from these two
225 continents (Fig. S2a). In Africa, MGT4-ST11 (also described by MGT3-ST10) was more prevalent
226 in West Africa and MGT4-ST16 (also described by MGT3-ST15) in Central/Eastern Africa (Fig.
227 S3). Finally, MGT4-ST15 was a global ST which was observed in all continents (Fig. 2b).

228 These dominant STs can be further grouped into CCs which offered a more inclusive picture.
229 Among the 423 MGT4-CCs at MGT4, 10 represented 94.1% of all the isolates. The top two CCs,
230 MGT4-CC1 and MGT4-CC13 accounted for 88.0% of the isolates (Fig. 2b). MGT4-CC1 was
231 prevalent in all six continents, while MGT4-CC13 was more common in North America and Europe
232 (Fig. 2b, Fig. S4).

233 **The national or local epidemiology of *S. Enteritidis* by MGT**

234 As the majority of the *S. Enteritidis* sequences analysed in this study were from the USA and UK, we
235 compared the distribution of the STs between these two countries. We used MGT4 and MGT5 levels
236 to describe the data. A total of 39 MGT4 STs with more than 50 isolates represented 75.1% of the
237 USA and UK isolates and each country had its own specific types (Fig. 3a). MGT4-ST99, ST136,
238 ST135, ST171, ST163, ST370, ST396 and ST150 were the main STs in the USA, whereas MGT4-
239 ST15, ST25, ST13, ST100, ST31, ST326, ST24, ST208, ST29 were the main STs in the UK. MGT4-
240 ST15 and ST25 were further subtyped into 34 MGT5-STs (each with more than 20 isolates), of
241 which the majority were mainly observed in the UK, except for MGT5-ST412 and ST387 in the
242 USA (Fig. 3b).

243 To examine the relationship between MGT-STs, isolation source and location, we examined
244 MGT4-STs of 4383 genome sequences from the USA which contained source and state metadata. Of
245 the 4383 genomes, 46.7% were from avian source while 43.1% were from humans. Six MGT4 STs
246 (with more than 50 isolates each) were isolated from both human and avian sources including

247 MGT4-ST99, ST135, ST136, ST160, ST198 and ST25, while three STs including MGT4-ST15,
248 ST163 and ST171 were mainly from human sources (Fig. 4). All of the human-only STs belonged to
249 MGT4-CC1 whereas the mixed source STs were mostly of MGT4-CC13 origin. STs belonging to
250 MGT4-CC13 were significantly more prevalent in avian sources than STs of MGT4-CC1 ($P < 0.001$,
251 OR = 45.9). For MGT4-CC13, *S. Enteritidis* isolate metadata from 48 states of the USA were
252 available and ST frequencies were similar across the country (Fig. 4). In almost all states, MGT4-
253 ST99 (within MGT4-CC13) was the dominant type, followed by MGT4-ST135 and MGT4-ST136.
254 While for MGT4-CC1 STs, ST15, ST163 and ST171 are predominant in only two states.

255 Most UK isolates contained collection year and month metadata. There were 8,818 human *S.*
256 *Enteritidis* isolates collected from March 2014 to July 2018 in the UK. We chose MGT5-STs to
257 describe the monthly variation of *S. Enteritidis* in the UK. By MGT5, variation in the prevalence of
258 MGT5-STs in different years and months was observed (Fig. 5). There were 13 MGT5-STs with
259 more than 100 isolates, representing 46.3% of the 8,818 isolates. The top five STs over the entire
260 period were MGT5-ST1, ST29, ST79, ST15 and MGT5-ST33, representing 30% of the total isolates.
261 These STs showed differing temporal patterns. MGT5-ST1, ST29, ST79 and ST15 were consistently
262 observed in each month across these four years while MGT5-ST15 included isolates previously
263 reported as part of an outbreak [45]. MGT5-ST156 first appeared in April of 2012 in the UK,
264 increased in frequency in June and July of 2014, and became rare after 2014. MGT5-ST423 was a
265 dominant type from March to September of 2015, then became rare and was dominant again from
266 the September of 2016 to January of 2017.

267 **Detection of potential large outbreak clusters of *S. Enteritidis* using MGT**

268 MGT9 offered highest resolution for tracking strain spread and outbreak detection. By MGT9, there
269 were 124 MGT9-STs including isolates from two or more countries each, indicating international
270 spread (Data Set S1, Tab 3). To facilitate outbreak detection, we identified single linkage clusters of

271 isolates using MGT9 allele differences with a range of cut-offs (0, 1, 2, 5 and 10 allelic differences)
272 that were named as outbreak detection clusters (ODC0, 1, 2, 5 and 10). These clusters can be used to
273 analyse frequencies of closely related isolates at different cut-off levels for population studies and
274 may also be used as dynamic thresholds to detect potential outbreaks [46].

275 In this study, we used ODC2 clusters to detect potential outbreak clusters and to assess whether
276 some STs were more likely to cause large outbreak clusters, based on the number of ODC2 clusters
277 and the total number of isolates in these clusters in different STs. Since the global data may be biased
278 towards outbreak isolates that were preferentially sequenced, we used UK data from 2014 to 2018
279 which included all human isolates referred to public health authorities [47]. There were 17 ODC2
280 clusters of more than 50 isolates representing 1855 isolates in total. The majority of these ODC2
281 clusters belonged to 12 different MGT4-STs and therefore we used MGT4 level to perform
282 comparison (Data Set S1, Tab 4). MGT4-ST15 (1804 isolates) was the dominant ST in UK and
283 contained only one ODC2 cluster of 62 isolates while MGT4-ST25 (1532 isolates), which was the
284 second dominant ST in UK, contained five large ODC clusters containing 58 to 287 isolates (602 in
285 total). MGT4 ST15 is significantly less likely to contain large ODC2 clusters than the other STs (OR
286 = 0.1, P value < 0.001), while MGT4-ST25 is significantly more likely to contain larger ODC2
287 clusters than other STs (OR = 3.1, P value < 0.001). It is noteworthy that by clonal complexes, all of
288 the six top STs belonged to MGT4-CC13 and were positively associated with large scale outbreaks,
289 including three previously reported large scale outbreaks in Europe [19, 45, 48, 49]. Indeed, MGT4-
290 CC13 was significantly more likely to cause larger scale outbreaks than MGT4-CC1 (OR = 6.2, P <
291 0.001). These associations of STs and CCs with large outbreak clusters were also significant when
292 ODC5 clusters was used. Notably the cluster sizes were larger but the trend remains the same (Data
293 Set S1, Tab 4 and Tab 5).

294 **Antibiotic resistance (AR) gene profiling of *S. Enteritidis***

295 As nearly all of the isolates (99.98%, 26,665/26,670) harboured the *aac(6')-Iaa_1* gene for
296 aminoglycoside resistance, this gene was excluded from the antibiotic resistance analysis. A total of
297 2505 isolates (9.4% of the total 26,670 isolates) were found to harbor antibiotic resistant genes
298 excluding *aac(6')-Iaa_1*. The most frequent predicted class was tetracyclines (5.1%, 1350/26,670),
299 followed by beta-lactams (4.5%, 1213/26,670) and aminoglycosides (2.8%, 741/2270). Among the
300 2505 isolates with AR genes, 40.4% (1011/2505) were predicted to be MDR, harbouring genes
301 conferring resistance to three or more different antibiotic classes. And 59.6% (1494/2505) of the
302 isolates harboured genes conferring resistance to one to two different antibiotic classes. Although the
303 selection of isolates for genome sequencing in some continents may be biased, African isolates were
304 found to have the highest proportion of AR/MDR isolates, followed by Asia and Oceania (Fig. S5a).

305 STs containing more than 10 isolates at different MGT levels were screened to identify MDR or
306 AR associated STs (defined as $\geq 80\%$ of the isolates are predicted to be MDR or AR). Eleven STs
307 from varied levels were identified as MDR STs, representing 49% (659/1011) of the MDR isolates.
308 These STs were mutually exclusive at different MGT levels. The top two STs, MGT3-ST15 and
309 ST10, were the two invasive types that were prevalent in Africa. For MGT3-ST15, 99.2% of the
310 isolates harboured resistance genes corresponding to as many as six different drug classes (Table 1).
311 For MGT3-ST10, 86.5% of the isolates were MDR, and the antibiotic resistant patterns were similar
312 to those of MGT3-ST15. MGT3-ST10 and ST15 represented 96.3% and 90.2% of the two Africa
313 endemic lineages (MGT3-CC10 and CC15, respectively) as mentioned below. Among the other nine
314 MDR STs, eight belonged to MGT4-CC1, and 93.5% of the isolates (243/260) harboured genes
315 conferring resistance to as many as eight classes of antibiotic drugs. Only one MDR ST (MGT6-
316 ST2698) belonged to MGT4-CC13 with 89.7% (35/39) of the isolates harboured genes conferring
317 resistance to four drug classes. Based on available population sampling, MGT4-CC1 had
318 significantly more isolates with MDR genes (5.6%, 474/8539) than MGT4-CC13 (0.7%, 107/14,972)
319 (Fisher exact test, P value < 0.001 , OR = 12.4).

320 A total of 707 isolates belonging to 47 different STs from MGT3 to MGT7, harboured genes
321 conferring resistance to one or two different antibiotic classes including aminoglycosides, beta-
322 lactams, tetracyclines and quinolones (Data Set S1, Tab 6). Among the 47 STs, 34 (72%, 34/47)
323 belonged to MGT4-CC1, representing 80% (564/707) of isolates, and 12 (26%, 12/47) STs belonged
324 to MGT4-CC13 representing 18% (125/707) of the isolates. Again MGT4-CC1 had significantly
325 more isolates with AR genes than MGT4-CC13 (Fisher exact test, P value < 0.001, OR = 8.4).

326 Plasmid specific genes were identified from the PlasmidFinder database [33] and IncQ1, IncN,
327 IncI1, IncX1 plasmid types were common in *S. Enteritidis* isolates with AR genes (Fig. S5b). In
328 particular, plasmid type IncQ1 was present in MGT3-ST15, ST10 and ST30, which harboured MDR
329 genes up to six drug classes (Table 1). Plasmid type IncI1 was common in MGT3-ST10 and ST161
330 and plasmid type IncX1 was more common in the STs of MGT4-CC1.

331 **The population structure and evolution of major *S. Enteritidis* STs/CCs**

332 The majority of the STs from different MGT levels analysed belonged to the two MGT4 CCs,
333 MGT4-CC1 and CC13. To describe the global phylogenetic structure of *S. Enteritidis* and explore
334 the phylogenetic relationship of the two lineages, 1506 representative isolates were selected using
335 representatives of MGT6-STs to encompass the diversity of the serovar. A previous study had
336 suggested that *S. Enteritidis* has three clades, A, B and C. Clade A and C appeared to diverge earlier
337 and were phylogenetically more distant to the global clade B than *Salmonella enterica* serovar
338 Gallinarum [11] (Fig. 6a). *S. Enteritidis* clade B and *S. Gallinarum* were sister clades. The vast
339 majority of the genomes analysed belonged to clade B.

340 Within clade B, we identified 10 main lineages, which were concordant with MGT-CCs from
341 MGT1 to MGT4 (Fig. 6b). These lineages can be described at different MGT levels as shown in Fig.
342 6b. The lineages were consistent with the progressive division from lower to higher MGT levels
343 grouped by CCs. MGT1-ST11 can be represented by MGT2-CC1 and MGT1-ST183 by MGT2-CC3.

344 At MGT3, four lineages named MGT3-CC10, CC15, CC18 and CC107 defined separate lineages,
345 while MGT3-CC1 included several lineages represented by MGT4-CCs. MGT3-CC10 and CC15
346 represented 100% of the two previously reported Africa lineages associated with invasive infection
347 [5]. MGT3-CC1 included five main MGT4-CCs, with two, MGT4-CC1 and MGT4-CC13,
348 representing 88.0% of all 26,670 *S. Enteritidis* isolates (Fig. 6b). MGT4-CC30 and CC129, both of
349 which were phylogenetically closer to MGT3-CC13, were two endemic lineages in Europe and
350 North America, respectively (Fig. 6b). The population structure of *S. Enteritidis* defined by MGT,
351 were generally in accordance with cgMLST HierCC HC100 (Fig. S6). The association between
352 MGT-STs/CCs with previously reported SNP analysis-based nomenclature (lineages/clades) and
353 phage types were summarized (Dataset 1, Tab 7) [5, 19, 44, 45, 49].

354 We performed BEAST analysis using the *S. Enteritidis* core genome to determine the evolutionary
355 rate of the clade B *S. Enteritidis* which was estimated to be 1.9×10^{-7} substitution/site/year (95% CI of
356 1.6×10^{-7} to 2.3×10^{-7}), corresponding to 0.8 SNPs per genome per year for the core genome (95% CI
357 of 0.6 to 0.9 SNPs). The mutation rate of MGT4-CC13 lineage core genome was estimated to be
358 2.5×10^{-7} substitution/site/year (95% CI of 2.3×10^{-7} to 2.7×10^{-7}) or 1.0 SNP per genome per year
359 (95% CI of 0.9 to 1.1 SNPs). The mutation rate of MGT4-CC1 lineage core genome was 1.7×10^{-7}
360 substitution/site/year (95% CI of 1.6×10^{-7} to 2.0×10^{-7}), or 0.7 SNP per genome per year (95% CI of
361 0.6 to 0.8 SNPs). The mutation rate of MGT4-CC13 was significantly faster (1.5 times) than that of
362 MGT4-CC1.

363 The most recent common ancestor (MRCA) of the nine lineages belonging to MGT2-CC1, was
364 estimated to have existed in the 1460s (95% CI 1323 to 1580) (Fig. S7). The two global epidemic
365 lineages MGT4-CC1 and CC13 are estimated to have diverged at around 1687 (95% CI 1608 to
366 1760). In 1869 (95% CI=1829-1900), MGT4-CC13 diverged into two sub-lineages of various
367 MGT4-STs, with one more prevalent in North America than in other continents (labelled with red
368 arrow) and the other sub-lineage more prevalent in Europe (labelled with blue arrow). We further

369 estimated the population expansion of the two global epidemic lineages MGT4-CC1 and CC13 (Fig.
370 7a, b). For MGT4-CC1, there were two large expansions around 1950 and 1970. For MGT4-CC13,
371 the population gradually increased until a more rapid expansion around 1970.

372 **Virulence genes and SPIs distribution in the STs/CCs represented phylogeny of *S. Enteritidis***

373 We further compared the distribution of virulence genes and SPIs in the 13 STs/CCs that represent
374 the major phylogenetic lineages of *S. Enteritidis*. Based on the VFDB database, 162 genes were
375 present in ≥ 10 isolates of *S. Enteritidis*, 123 (75.9%) genes of which were present in all of the
376 STs/CCs. Sixteen genes (9.9%) were associated with one or more STs/CCs while the remainder were
377 sporadically distributed (Table 2). The *spv* locus including *spvB*, *spvC* and *spvR*, are reported to be
378 associated with non-typhoidal bacteraemia [50]. *SpvBCR* genes were absent in MGT1-ST3304,
379 ST180, ST1972 but present in all the other CCs. Pef fimbriae operon *pefABCD* genes were absent in
380 MGT1-ST3304, ST180, ST1972, ST183 and MGT3-CC10. The *ssek2* gene, which encodes a
381 secretion effector of SPI-2, was reported to significantly contribute to the pathogenicity of
382 *Salmonella* [51]. *ssek2* was present in MGT3-CC107 and CC18, MGT4-CC129 and CC1, but was
383 absent in MGT4-CC13 and other STs/CCs. On the reference genome P125109 (MGT4-CC1), *ssek2*
384 was observed in the prophage \square SE20. Moreover, 34% of the isolates in MGT3-CC107 (further
385 represented by five STs from MGT4 to MGT7 levels) were found to harbour the *Yersinia* high-
386 pathogenicity island (HPI), representing 75.5% of the HPI positive isolates (Data Set S1, Tab 8).

387 Among the 23 SPIs reported, SPI-2, 3, 4, 9,12,13,14 and 16 were found intact in all of the
388 STs/CCs, and SPI-8, 15, 18, 20 and 21 were absent in all STs/CCs (Table 2). SPI-1 was intact in
389 MGT1-ST1972 (clade C), while all the other STs/CCs had three SPI-1 genes (STM2901, STM2902
390 and STM2903) missing. SPI-5 was nearly intact in all STs/CCs with one to two genes missing in
391 MGT1-ST183 and MGT4-CC101. SPI-11 was only intact in MGT4-CC101, with four to five genes
392 missing in the other STs/CC. SPI-17 was intact in the majority of the STs/CCs except for MGT1-

393 ST3304 and ST180 (clade A) with only one gene observed. SPI-19 was nearly intact in MGT1-
394 ST3304, ST180, ST1972 and ST183, but was truncated in all the other CCs in clade B. None of the
395 STs/CCs harboured the intact SPI-6, 7, 10, 22, and 23, only a few genes of which were observed.
396 MGT1-ST183, MGT3-CC15 and MGT4-CC13 were found to harbour 44 to 49 genes of SPI-7 (149
397 in total), the majority of which were located on the Fels2-like prophage.

398 **Discussion**

399 *S. Enteritidis* is one of the most common foodborne pathogens causing large scale national and
400 international outbreaks and food recalls [3, 52]. Understanding the population structure and genomic
401 epidemiology of *S. Enteritidis* is essential for its effective control and prevention. In this study, we
402 applied the genomic typing tool MGT to *S. Enteritidis* and developed a database for international
403 applications. We used MGT to describe its local and global epidemiology, and its population
404 structure using 26,670 publicly available genomes and associated metadata. In this work, STs and
405 CCs were assigned to each of the nine MGT levels, with MGT1 refers to the legacy seven gene
406 MLST [12]. Thus, attention should be paid to the prefix MGT levels for those STs and CCs to avoid
407 confusion.

408 **The *S. Enteritidis* MGT scheme enables a scalable resolution genomic nomenclature**

409 The design of the *S. Enteritidis* MGT is based on the *S. Typhimurium* MGT scheme published
410 previously [20]. The first eight levels (MGT1 to MGT8) of *S. Enteritidis* MGT scheme used the same
411 loci as for *S. Typhimurium*. We defined the core genome of *S. Enteritidis* which had 3932 core genes
412 and 1054 core intergenic regions, which was used as MGT9. MGT9 substantially increased the
413 subtyping resolution for *S. Enteritidis* [13]. Our study suggests that the MGT levels one to eight
414 could be applied to all *Salmonella enterica* serovars as a common scheme for the species, and only
415 an additional serovar-specific MGT9 scheme needs to be designed for individual serovars that
416 require the highest resolution for outbreak investigations.

417 The online database of the *S. Enteritidis* MGT scheme offers an open platform for global
418 communication of genomic data and facilitates detection of international transmission and outbreaks
419 of *S. Enteritidis*. The STs, especially at the middle resolution MGT level, were associated with
420 different geographic regions, sources and MDR. The extension of MGT to *S. Enteritidis* was based
421 on our previous study on *S. Typhimurium* [20]. Importantly, the stable characteristics of STs, which
422 are based exact comparison, makes up for the main drawback of single-linkage clustering methods
423 that the cluster types may change, especially at higher resolutions. MGT STs enables long term
424 epidemiological communication between different laboratories. And the variable resolution of MGT
425 offers flexibility for temporal and spatial epidemiological analysis using an underlying stable
426 nomenclature of STs from the finest resolution level. Additionally, CCs at each MGT level were able
427 to cluster the STs with one allele difference and are concordant with phylogenetic lineages as shown
428 in this study.

429 **MGT for *S. Enteritidis* uncovers geographic, source and temporal epidemiological trends of *S.***
430 ***Enteritidis* within and between countries**

431 A total of 26,670 isolates were successfully typed using the MGT scheme which allowed the
432 examination of the global (or international) and local (or national) epidemiology of *S. Enteritidis*.
433 The salient feature of the flexibility of the different levels of MGT has also been illustrated through
434 this analysis. The lower resolution levels of MGT were found to be able to effectively describe the
435 geographic variation in different continents, countries or regions. Of these levels we showed that
436 MGT4-STs best described the global epidemiology of *S. Enteritidis* at continental level with some
437 STs distributed globally while others more geographically restricted. MGT4-ST15 was a global
438 epidemic type which was prevalent in almost all continents. By contrast, MGT4-ST16 (also
439 described by MGT3-ST15) was prevalent in Central/Eastern Africa and MGT4-ST11 (also described
440 by MGT3-ST10) was prevalent in West Africa, which agreed with a previous study [5].

441 In USA, MGT4-CC13 and MGT-CC1 showed remarkable difference in their epidemiology.
442 MGT4-CC13 STs including MGT4-ST99, ST136 and ST135 were the main cause of clinical
443 infections and were commonly isolated from poultry. The prevalent STs were generally similar in
444 different states. In particular, MGT4-ST99, ST136 and ST135 were the dominant types in almost all
445 states of the USA. The distribution and poultry association of these STs suggest that they may be
446 responsible for several multistate outbreaks caused by *S. Enteritidis* contaminated eggs [1, 53]. Since
447 poultry related products (i.e. eggs, chicken and turkey, especially eggs) are known to be the main
448 source of *S. Enteritidis* infections or outbreaks in the USA [4], this isn't surprising, but it
449 demonstrates the utility of the MGT.

450 On the other hand, MGT4-CC1 including MGT4-ST171, ST15 and ST163, were less prevalent in
451 the USA. Nevertheless, these STs were isolated from human infections but were very rare in poultry,
452 although sampling bias may affect this conclusion as the poultry isolates in the dataset was not from
453 systematic sampling of poultry sources. On the other hand, beef, sprouts, pork, nuts and seeds can
454 also be contaminated by *S. Enteritidis* [4]. These non-poultry related foods may be the source of the
455 MGT4-CC1 caused infections. Overall MGT4-CC13 was found to be significantly associated with
456 poultry source, which is concordant with a previous study [10]. Further studies are required to
457 explain the association of MGT4-CC13 but not MGT4-CC1 with poultry products. The lack of
458 potential source isolates within MGT4-CC1 STs highlights the need for comprehensive sequencing
459 and epidemiological efforts across the food production chain. Machine learning approaches, which
460 have been applied to the root source identification for *S. Typhimurium* outbreaks, could also
461 facilitate the source tracing of *S. Enteritidis* [17].

462 Temporal variation of UK *S. Enteritidis* was depicted clearly by MGT. From 2014 onwards, all
463 clinical *S. Enteritidis* were routinely sequenced in the UK [47]. Some MGT5-STs were found to
464 occur for a few months and then disappear and may be indicative of outbreaks. For example, MGT5-

465 ST156, which increased substantially during June and July in 2014 but became rare after 2015 and
466 describes a reported large scale outbreak [19]. Other STs persisted for long periods of time. For
467 example, MGT5-ST1 and ST29, appeared to be endemic isolates which may be associated with local
468 reservoirs. Therefore, the variation of MGT-STs across different seasons offered additional
469 epidemiological signals for suspected outbreaks and endemic infections of *S. Enteritidis*. The
470 stability of STs avoids the potential cluster merging issues of single linkage clustering based methods
471 and ensures continuity for long-term surveillance [20].

472 **MGT for *S. Enteritidis* facilitates outbreak detection, source tracing and evaluation of** 473 **outbreak propensity**

474 Whole genome sequencing offered the highest resolution for outbreak detection and source tracing.
475 However, resolution of cgMLST heavily depends on the diversity of the species. Our previous study
476 showed that for *S. Typhimurium*, serovar core genome offered higher resolution than species level
477 core genome for outbreak investigation [20, 54]. We designed MGT9 for *S. Enteritidis* using 4986
478 loci, which is around 2000 more loci than *Salmonella enterica* core genes [13, 20], thus increasing
479 the resolution of subtyping.

480 To facilitate outbreak detection MGT9 STs were further grouped to ODCs. There is no agreed
481 upon single cut-off for outbreak detection and dynamic thresholds have been suggested [46]. Here,
482 we used ODC2, which has a two-allele difference cut-off, to identify potential outbreak clusters.
483 Although the data do not allow us to confirm whether any of these ODC2 clusters were actual
484 outbreaks, a number of known outbreaks from other studies fell into ODC2 clusters [46, 48, 55]. A
485 two-allele cut-off was selected because it is at the lower end of cut-offs in reported *Enteritidis*
486 outbreaks [48, 55], which should limit the number of false positive outbreak calls. However, due to
487 the variability in diversity of isolates from different outbreaks, a dynamic threshold for cut-offs
488 would be more sensitive and specific to detect outbreaks [46]. Further work for *S. Enteritidis* is

489 required to address this issue fully. Using the UK clinical isolates from 2014 to 2018, we found that
490 the European and North American prevalent lineage MGT4-CC13 was significantly correlated with
491 larger ODC2 clusters (≥ 50 isolates each) than the global epidemic lineage of MGT4-CC1. Thus,
492 MGT4-CC13 is more likely to cause large scale outbreaks than MGT4-CC1. In the past few years,
493 several large scale outbreaks in Europe were due to the contaminated eggs [48, 49, 52]; these
494 isolates all belonged to MGT4-CC13. In the USA, MGT4-CC13 was the dominant lineage in both
495 human infections and poultry product contamination, while MGT4-CC1 was relatively rare in
496 poultry. Industrialised and consolidated poultry/eggs production and marketing could have facilitated
497 the spread of MGT4-CC13 causing large scale outbreaks. Further studies are required to definitively
498 identify the biological and environmental mechanisms facilitating these large MGT4-CC13
499 outbreaks.

500 ODCs offer a means for accurate detection of outbreaks at the highest typing resolution, MGT9
501 [20]. Other studies used SNP address or cgMLST [19, 45, 48, 49] with SNP cut-offs or HierCC
502 clustering for outbreak detection. All three methods use the same single linkage clustering method
503 for outbreak detection. MGT offers further advantage that potential outbreak clusters can be
504 precisely defined by a stable MGT-ST identifier rather than a cluster number [20].

505 **MGT for *S. Enteritidis* improves precision of detection and tracking of MDR clones globally**

506 The rise of AMR in *Salmonella* is a serious public health concern [5]. The global spread of AMR can
507 be mediated by lateral transfer of resistance genes as well as clonal spread of resistant strains. This
508 study systematically evaluated the presence of antibiotic resistant genes in the 26,670 genomes of *S.*
509 *Enteritidis*, 9.4% of which harboured antibiotic resistant genes. Among the two global epidemic
510 lineages, MGT4-CC1 was found to contain significantly more antibiotic resistant isolates than
511 MGT4-CC13.

512 We further identified STs that were associated with MDR. Eleven STs (≥ 10 isolates each, 659
513 isolates in total) of different MGT levels were associated with MDR. These STs from different MGT
514 levels were mutually exclusive, emphasising the flexibility of MGT for precise identification of
515 MDR clones. MGT3-ST15 and MGT3-ST10 were representative of the previously reported Africa
516 invasive infection related lineages [5], harboured resistant genes of up to six different antibiotic
517 classes. The other STs, which were mainly from Europe and North America, harbour AR genes to up
518 to eight different classes of drugs. MGT3-ST30 and MGT4-ST718, which belonged to MGT4-CC1
519 and were observed in Europe, had been reported to be MDR phenotypically [56]. Significantly,
520 isolates from these two STs were mainly isolated from blood samples [56]. These MDR STs
521 correlated with blood stream infections should be monitored closely as they can potentially be more
522 invasive [5].

523 Because plasmids are known to play a key role in the acquisition of drug resistance genes [57, 58],
524 we identified Inc plasmid groups that are likely present in the *S. Enteritidis* population. Several of
525 these putative plasmids were MDR associated. MDR of African isolates in MGT3-ST10 and ST15
526 has been reported to be mediated by plasmids [5]. The West Africa lineage MGT3-ST10 isolates
527 were reported to have IncI1 plasmids [59]. In this study, both IncI1 and IncQ1 plasmid were present
528 in the Africa MGT3-ST10 and ST15 isolates as well as STs belonged to global epidemic lineage
529 MGT4-CC1. IncQ1 plasmids, which are highly mobile and widely transferred among different
530 genera of bacteria [60], are correlated with MDR of *Salmonella* [61, 62]. IncI1 plasmids are
531 responsible for the cephalosporin resistance of *Salmonella* and *Escherichia* in several countries [63-
532 65]. Moreover, IncN and IncX1 plasmids were common in the MDR associated STs of MGT4-CC1
533 and CC13. The use of MGT could enhance the surveillance of drug resistance plasmids.

534 **Defining the population structure of *S. Enteritidis* with increasing resolution and precision at**
535 **different MGT levels**

536 In this study, we defined the global population structure of *S. Enteritidis* using MGT-CCs. Three
537 clades of *S. Enteritidis* were previously defined with clade A and C as outgroups to the predominant
538 clade B [11], a classification supported by this study. We further identified 10 lineages within clade
539 B, which can be represented by different CCs from MGT2 to MGT4 corresponding with time of
540 divergence. Lower resolution level MGT CCs described old lineages well while higher resolution
541 level MGT CCs described more recently derived lineages (Fig. S7). Thus, the epidemiological
542 characteristics of a lineage can be identified using STs and/or CCs at appropriate MGT level. MGT2-
543 CC3 represent the Europe endemic MGT1-ST183 lineage, within which MGT2-ST3 and MGT2-
544 ST82 were able to identify previously defined phage types PT11 and PT66 respectively [44]. MGT3-
545 CC10 included all of the reported West Africa lineage isolates, and MGT3-CC15 included all of the
546 reported Central/Eastern Africa lineage isolates as well as Kenyan invasive *S. Enteritidis* isolates [5,
547 8]. MGT3-ST10 and MGT3-ST15 (represented 96.3% of MGT3-CC10 and 90.2% of MGT3-CC15
548 respectively), were MDR with similar resistance patterns. Most importantly, the two major global
549 epidemic lineages are clearly described by MGT4-CC1 and MGT4-CC13. MGT4-CC1 corresponds
550 to the global epidemic clade in the study of Feasey *et al.*, and MGT4-CC13 to the outlier clade of
551 that study [5]. MGT4-CC1 and CC13 correspond to lineage III and lineage V in Deng *et al.*'s study
552 [10]. Thus, by using STs and/or CCs of different MGT levels the population structure of *S.*
553 *Enteritidis* can be described clearly and consistently.

554 Interestingly, the two dominant lineages, MGT4-CC1 and CC13, showed different population
555 expansion time and evolutionary dynamics. Bayesian evolutionary analysis revealed a core genome
556 mutation rate for *S. Enteritidis* of 1.9×10^{-7} substitution/site/year which is similar to the genome
557 mutation rate of 2.2×10^{-7} substitution/site/year estimated by Deng *et al* [10]. In contrast, the mutation
558 rate of MGT4-CC13 was 1.5 times faster than MGT4-CC1. Population expansion of MGT4-CC1
559 occurred with two peaks in the 1950s and 1970s, while MGT4-CC13 population increased steadily
560 until the 1970s. The expansion around the 1970s may have been driven by the development of the

561 modern industrialised poultry industry, including poultry farms and/or processing plants, that *S.*
562 Enteritidis colonized. This is concordant with Deng *et al*'s speculation that the expansion of *S.*
563 Enteritidis population was correlated with the poultry farm industry [10, 18].

564 **Acquisition and degradation of virulence factors in the STs/CCs represent phylogeny of *S.***
565 **Enteritidis**

566 By screening the virulence genes and SPIs in different STs/CCs representing the three clades and
567 major lineages of *S. Enteritidis*, distributional differences were observed for some virulence genes
568 and SPIs. The main difference observed was between clade A/C and clade B, and MGT1-ST183 and
569 other lineages of clade B. *SpvBCR*, *pefABCD*, and *rck* genes were found to be located on the same
570 plasmid in the complete genomes of *S. Enteritidis* (data not shown). However, the Europe endemic
571 MGT1-ST183 and the West Africa endemic MGT3-CC10 were positive for the *spvBCR* genes, but
572 were negative for the *pefABCD* and *spvBCR* genes, suggesting that there may exist other
573 mechanisms of acquisition of *spvBCR* genes. Variation in the distribution of genes on SPIs was also
574 observed. Some virulence or SPIs genes were likely to have been lost in certain STs/CCs.
575 Degradation of virulence genes had been previously observed in *S. Enteritidis* and *S. Typhimurium*
576 [5, 66]. In summary, both acquisition and degradation of virulence factors occurred in the evolution
577 of *S. Enteritidis*.

578 Moreover, the main differences in virulence factors between the two dominant lineages MGT4-
579 CC1 and CC13 were *ssek2* and the number of SPI-7 genes. *ssek2*, located on the prophage \square SE20
580 was limited to MGT4-CC1 [5], and around 35 SPI-7 genes, located on the Fels2-like prophage were
581 limited to MGT4-CC13 [5]. \square SE20 was found to contribute to mouse infections [67]. Fels2-like
582 prophage was also present in a bloodstream infection related lineage three of *S. Typhimurium* ST313
583 in Africa, which is estimated to have significantly higher invasiveness index than the other lineages
584 of ST313 [66]. However, the detailed pathogenic role of Fels2-like prophage in *Salmonella*
585 infections remained unclear. There are potentially undiscovered virulence factors contributing to the

586 epidemiological variation (infection severity, outbreak propensity and geographic spreading) among
587 clades and lineages of *S. Enteritidis*.

588 **Limitations of this study and challenges for public database**

589 The epidemiological results generated in this study were based on the publicly submitted metadata
590 and the correctness of the results therefore depends on the accuracy of this data. In as many cases as
591 possible metadata were verified from published sources. However, the possibility exists that
592 incorrect metadata or repeated sequencing of the same isolates influence the results. We showed the
593 effect of the latter was minimal (**Supplementary Material**). However, good metadata is essential,
594 and an internationally agreed on minimal metadata would be useful for better epidemiological
595 surveillance of *S. Enteritidis*. The online MGT system offers unprecedented power for monitoring *S.*
596 *Enteritidis* across the globe. Good metadata further enhances that power. Moreover, all the genomes
597 of this work were Illumina short-read sequencing except for one reference complete genome. The
598 thresholds of identity and coverage for searching AR genes and plasmid replicon genes were \geq
599 90%. Incomplete genomes and these thresholds may result in some AR and plasmid replicon genes
600 being missed. Additionally, considering the mobile nature of plasmids and prophages, some STs may
601 lose AR or virulence genes in rare cases, timely updating of ST definitions with respect to AR and
602 virulence states will therefore be necessary.

603 **Conclusions**

604 In this study, we defined the core genome of *S. Enteritidis* and developed an MGT scheme with nine
605 levels. MGT9 offers the highest resolution and is suitable for outbreak detection whereas the lower
606 levels (MGT1 to MGT8) are suitable for progressively longer epidemiological timescales. At the
607 MGT4 level, globally prevalent and regionally restricted STs were identified, which facilitates the
608 identification of the potential geographic origin of *S. Enteritidis* isolates. Specific source associated
609 STs were identified, such as poultry associated MGT4-STs, which were common in human cases in

610 the USA. At the MGT5 level, temporal variation of STs was observed in *S. Enteritidis* infections
611 from the UK, which reveal both long lived endemic STs and the rapid expansion of new STs. Using
612 MGT3 to MGT6, we identified MDR-associated STs to facilitate tracking of MDR spread.
613 Additionally, certain plasmid types were found to be highly associated with the same MDR STs,
614 suggesting plasmid-based resistance acquisition. Furthermore, we evaluated the phylogenetic
615 relationship of the STs/CCs by defining the population structure of *S. Enteritidis*. A total of 10 main
616 lineages were defined in the globally distributed clade B of *S. Enteritidis*, which were represented
617 with 10 STs/CCs. Of these, MGT4-CC1 and CC13 were the dominant lineages, with significant
618 differences in large outbreak frequency, antimicrobial resistance profiles and mutation rates. Some
619 virulence and SPIs genes were distributed differently among the different STs/CCs represented
620 clades/lineages of *S. Enteritidis*. The online open database for *S. Enteritidis* MGT offers a unified
621 platform for the international surveillance of *S. Enteritidis*. In summary, MGT for *S. Enteritidis*
622 provides a flexible, high resolution and stable genome typing tool for long term and short term
623 surveillance of *S. Enteritidis* infections.

624 **Authors' contributions**

625 L.L., M.P., and R.L. designed the study. L.L. and M.P. set up the database. S.K and M.P. set up the
626 website, M.P., R.L., D.H. L.C. S.O., Q.W., M.T., V.S. and S.K. provided critical analysis and
627 discussions, L.L. wrote the first draft and all authors contributed to the final manuscript.

628 **Funding information**

629 This work was funded by a project grant from the National Health and Medical Research Council of
630 Australia (grant number 1129713). Lijuan Luo was supported by a UNSW scholarship (University
631 International Postgraduate Award). The funders had no role in study design, data collection and
632 interpretation, or the decision to submit the work for publication.

633 **Acknowledgements**

634 The authors thank Duncan Smith and Robin Heron from UNSW Research Technology Services for
635 high performance computing assistance.

636 **Conflicts of interest**

637 The authors declare that there are no conflicts of interest.

638

639 **References**

- 640 1. **Dewey-Mattia D, Manikonda K, Hall AJ, Wise ME, Crowe SJ.** Surveillance for
641 Foodborne Disease Outbreaks - United States, 2009-2015. *MMWR Surveill Summ*
642 2018;67(10):1-11.
- 643 2. Salmonellosis - Annual Epidemiological Report for 2017 [database on the Internet]2020.
644 Available from: <https://www.ecdc.europa.eu/en/publications-data>.
- 645 3. **Pijnacker R, Dallman TJ, Tijjsma ASL, Hawkins G, Larkin L et al.** An international
646 outbreak of *Salmonella enterica* serotype Enteritidis linked to eggs from Poland: a
647 microbiological and epidemiological study. *Lancet Infect Dis* 2019;19(7):778-786.
- 648 4. **Snyder TR, Boktor SW, M'Ikanatha N M.** Salmonellosis Outbreaks by Food Vehicle,
649 Serotype, Season, and Geographical Location, United States, 1998 to 2015. *J Food Prot*
650 2019;82(7):1191-1199.
- 651 5. **Feasey NA, Hadfield J, Keddy KH, Dallman TJ, Jacobs J et al.** Distinct *Salmonella*
652 Enteritidis lineages associated with enterocolitis in high-income settings and invasive disease
653 in low-income settings. *Nat Genet*, Article 2016;48(10):1211-1217.
- 654 6. **Mohan A, Munusamy C, Tan YC, Muthuvelu S, Hashim R et al.** Invasive *Salmonella*
655 infections among children in Bintulu, Sarawak, Malaysian Borneo: a 6-year retrospective
656 review. *BMC Infect Dis* 2019;19(1):330.
- 657 7. **Thung TY, Mahyudin NA, Basri DF, Wan Mohamed Radzi CW, Nakaguchi Y et al.**
658 Prevalence and antibiotic resistance of *Salmonella* Enteritidis and *Salmonella* Typhimurium
659 in raw chicken meat at retail markets in Malaysia. *Poult Sci* 2016;95(8):1888-1893.

- 660 8. **Akullian A, Montgomery JM, John-Stewart G, Miller SI, Hayden HS et al.** Multi-drug
661 resistant non-typhoidal *Salmonella* associated with invasive disease in western Kenya. *PLoS*
662 *Negl Trop Dis* 2018;12(1):e0006156.
- 663 9. **Parn T, Dahl V, Lienemann T, Perevoscikovs J, De Jong B.** Multi-country outbreak of
664 *Salmonella* enteritidis infection linked to the international ice hockey tournament. *Epidemiol*
665 *Infect* 2017;145(11):2221-2230.
- 666 10. **Deng X, Desai PT, den Bakker HC, Mikoleit M, Tolar B et al.** Genomic epidemiology of
667 *Salmonella enterica* serotype Enteritidis based on population structure of prevalent lineages.
668 *Emerg Infect Dis* 2014;20(9):1481-1489.
- 669 11. **Graham RMA, Hiley L, Rathnayake IU, Jennison AV.** Comparative genomics identifies
670 distinct lineages of *S. Enteritidis* from Queensland, Australia. *PLoS One*
671 2018;13(1):e0191042.
- 672 12. **Achtman M, Wain J, Weill FX, Nair S, Zhou Z et al.** Multilocus sequence typing as a
673 replacement for serotyping in *Salmonella enterica*. *PLoS Pathog* 2012;8(6):e1002776.
- 674 13. **Alikhan NF, Zhou Z, Sergeant MJ, Achtman M.** A genomic overview of the population
675 structure of *Salmonella*. *PLoS Genet* 2018;14(4):e1007261.
- 676 14. **Zhou Z, Alikhan N-F, Mohamed K, Fan Y, Agama Study G et al.** The EnteroBase user's
677 guide, with case studies on *Salmonella* transmissions, *Yersinia pestis* phylogeny, and
678 *Escherichia* core genomic diversity. *Genome Res* 2020;30(1):138-152.
- 679 15. **Ashton P, Nair S, Peters T, Tewolde R, Day M et al.** Revolutionising Public Health
680 Reference Microbiology using Whole Genome Sequencing: *Salmonella* as an exemplar.
681 *bioRxiv* 2015:033225.
- 682 16. **Zhou Z, Charlesworth J, Achtman M.** HierCC: A multi-level clustering scheme for
683 population assignments based on core genome MLST. *Bioinformatics* 2021.
- 684 17. **Zhang S, Li S, Gu W, den Bakker H, Boxrud D et al.** Zoonotic Source Attribution of
685 *Salmonella enterica* Serotype Typhimurium Using Genomic Surveillance Data, United
686 States. *Emerg Infect Dis* 2019;25(1):82-91.
- 687 18. **Deng X, Shariat N, Driebe EM, Roe CC, Tolar B et al.** Comparative analysis of subtyping
688 methods against a whole-genome-sequencing standard for *Salmonella enterica* serotype
689 Enteritidis. *J Clin Microbiol* 2015;53(1):212-218.
- 690 19. **Hormansdorfer S, Messelhauser U, Rampp A, Schonberger K, Dallman T et al.** Re-
691 evaluation of a 2014 multi-country European outbreak of *Salmonella* Enteritidis phage type
692 14b using recent epidemiological and molecular data. *Euro Surveill* 2017;22(50):17-00196.

- 693 20. **Payne M, Kaur S, Wang Q, Hennessy D, Luo L et al.** Multilevel genome typing:
694 genomics-guided scalable resolution typing of microbial pathogens. *Euro Surveill*
695 2020;25(20):1900519.
- 696 21. **Gurevich A, Saveliev V, Vyahhi N, Tesler G.** QUASt: quality assessment tool for genome
697 assemblies. *Bioinformatics* 2013;29(8):1072-1075.
- 698 22. **Robertson J, Yoshida C, Kruczkiewicz P, Nadon C, Nichani A et al.** Comprehensive
699 assessment of the quality of *Salmonella* whole genome sequence data available in public
700 sequence databases using the *Salmonella In silico* Typing Resource (SISTR). *Microb Genom*
701 2018;4(2).
- 702 23. **Bolger AM, Lohse M, Usadel B.** Trimmomatic: a flexible trimmer for Illumina sequence
703 data. *Bioinformatics* 2014;30(15):2114-2120.
- 704 24. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M et al.** SPAdes: a new genome
705 assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*
706 2012;19(5):455-477.
- 707 25. **Souvorov A, Agarwala R, Lipman DJ.** SKESA: strategic k-mer extension for scrupulous
708 assemblies. *Genome Biol* 2018;19(1):153.
- 709 26. **Wood DE, Salzberg SL.** Kraken: ultrafast metagenomic sequence classification using exact
710 alignments. *Genome Biol* 2014;15(3):R46.
- 711 27. **Zhang S, Yin Y, Jones MB, Zhang Z, Deatherage Kaiser BL et al.** *Salmonella* serotype
712 determination utilizing high-throughput genome sequencing data. *J Clin Microbiol*,
713 10.1128/JCM.00323-15 2015;53(5):1685-1692.
- 714 28. **Yoshida CE, Kruczkiewicz P, Laing CR, Lingohr EJ, Gannon VP et al.** The *Salmonella*
715 *In silico* Typing Resource (SISTR): An Open Web-Accessible Tool for Rapidly Typing and
716 Subtyping Draft *Salmonella* Genome Assemblies. *PLoS One* 2016;11(1):e0147101.
- 717 29. **Seemann T.** Prokka: rapid prokaryotic genome annotation. *Bioinformatics*
718 2014;30(14):2068-2069.
- 719 30. **Kroger C, Dillon SC, Cameron AD, Papenfort K, Sivasankaran SK et al.** The
720 transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium.
721 *Proc Natl Acad Sci U S A* 2012;109(20):E1277-1286.
- 722 31. Tableau (version. 9.1). *J Med Libr Assoc* 2016;104(2):182-183.
- 723 32. **Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S et al.** Identification of
724 acquired antimicrobial resistance genes. *J Antimicrob Chemother* 2012;67(11):2640-2644.

- 725 33. **Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O et al.** *In silico*
726 detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence
727 typing. *Antimicrob Agents Chemother* 2014;58(7):3895-3903.
- 728 34. **Treangen TJ, Ondov BD, Koren S, Phillippy AM.** The Harvest suite for rapid core-
729 genome alignment and visualization of thousands of intraspecific microbial genomes.
730 *Genome Biol* 2014;15(11):524.
- 731 35. **Price MN, Dehal PS, Arkin AP.** FastTree 2--approximately maximum-likelihood trees for
732 large alignments. *PloS one* 2010;5(3):e9490-e9490.
- 733 36. **Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA et al.** Rapid phylogenetic
734 analysis of large samples of recombinant bacterial whole genome sequences using Gubbins.
735 *Nucleic Acids Res* 2015;43(3):e15.
- 736 37. **Hu D, Liu B, Wang L, Reeves PR.** Living Trees: High-Quality Reproducible and Reusable
737 Construction of Bacterial Phylogenetic Trees. *Molecular biology and evolution*
738 2020;37(2):563-575.
- 739 38. **Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ et al.** Bayesian phylogenetic
740 and phylodynamic data integration using BEAST 1.10. *Virus Evol* 2018;4(1):vey016.
- 741 39. **Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA.** Posterior Summarization in
742 Bayesian Phylogenetics Using Tracer 1.7. *Syst Biol* 2018;67(5):901-904.
- 743 40. **Chen L, Yang J, Yu J, Yao Z, Sun L et al.** VFDB: a reference database for bacterial
744 virulence factors. *Nucleic Acids Res* 2005;33(Database issue):D325-328.
- 745 41. **Mansour MN, Yaghi J, El Khoury A, Felten A, Mistou M-Y et al.** Prediction of
746 *Salmonella* serovars isolated from clinical and food matrices in Lebanon and genomic-based
747 investigation focusing on Enteritidis serovar. *International Journal of Food Microbiology*
748 2020;333:108831.
- 749 42. **Kendall MG.** A new measure of rank correlation. *Biometrika* 1938;30(1/2):81-93.
- 750 43. **Feil EJ.** Small change: keeping pace with microevolution. *Nat Rev Microbiol* 2004;2(6):483-
751 495.
- 752 44. **Lawson B, Franklino LHV, Rodriguez-Ramos Fernandez J, Wend-Hansen C, Nair S et**
753 **al.** *Salmonella* Enteritidis ST183: emerging and endemic biotypes affecting western
754 European hedgehogs (*Erinaceus europaeus*) and people in Great Britain. *Sci Rep*
755 2018;8(1):2449.
- 756 45. **Kanagarajah S, Waldram A, Dolan G, Jenkins C, Ashton PM et al.** Whole genome
757 sequencing reveals an outbreak of *Salmonella* Enteritidis associated with reptile feeder mice
758 in the United Kingdom, 2012-2015. *Food Microbiol* 2018;71:32-38.

- 759 46. **Payne M, Octavia S, Luu LDW, Sotomayor-Castillo C, Wang Q et al.** Enhancing
760 genomics-based outbreak detection of endemic *Salmonella enterica* serovar Typhimurium
761 using dynamic thresholds. *Microb Genom* 2019.
- 762 47. **Chattaway MA, Dallman TJ, Larkin L, Nair S, McCormick J et al.** The Transformation
763 of Reference Microbiology Methods and Surveillance for *Salmonella* With the Use of Whole
764 Genome Sequencing in England and Wales. *Frontiers in public health*, Original Research
765 2019;7(317):317.
- 766 48. **Inns T, Lane C, Peters T, Dallman T, Chatt C et al.** A multi-country *Salmonella*
767 Enteritidis phage type 14b outbreak associated with eggs from a German producer: 'near real-
768 time' application of whole genome sequencing and food chain investigations, United
769 Kingdom, May to September 2014. *Euro Surveill* 2015;20(16):21098.
- 770 49. **Inns T, Ashton P, Herrera-Leon S, Lighthill J, Foulkes S et al.** Prospective use of whole
771 genome sequencing (WGS) detected a multi-country outbreak of *Salmonella* Enteritidis.
772 *Epidemiology & Infection* 2017;145(2):289-298.
- 773 50. **Guiney DG, Fierer J.** The Role of the *spv* Genes in *Salmonella* Pathogenesis. *Front*
774 *microbiol* 2011;2:129-129.
- 775 51. **Zhang X, He L, Zhang C, Yu C, Yang Y et al.** The impact of *sseK2* deletion on *Salmonella*
776 enterica serovar Typhimurium virulence in vivo and in vitro. *BMC Microbiol*
777 2019;19(1):182.
- 778 52. **Dallman T, Inns T, Jombart T, Ashton P, Loman N et al.** Phylogenetic structure of
779 European *Salmonella* Enteritidis outbreak correlates with national and international egg
780 distribution network. *Microb Genom* 2016;2(8):e000070.
- 781 53. **Nichols M, Stevenson L, Whitlock L, Pabilonia K, Robyn M et al.** Preventing Human
782 *Salmonella* Infections Resulting from Live Poultry Contact through Interventions at Retail
783 Stores. *J Agric Saf Health* 2018;24(3):155-166.
- 784 54. **Fu S, Octavia S, Tanaka MM, Sintchenko V, Lan R.** Defining the Core Genome of
785 *Salmonella enterica* Serovar Typhimurium for Genomic Surveillance and Epidemiological
786 Typing. *J Clin Microbiol* 2015;53(8):2530-2538.
- 787 55. **Taylor AJ, Lappi V, Wolfgang WJ, Lapierre P, Palumbo MJ et al.** Characterization of
788 Foodborne Outbreaks of *Salmonella enterica* Serovar Enteritidis with Whole-Genome
789 Sequencing Single Nucleotide Polymorphism-Based Analysis for Surveillance and Outbreak
790 Detection. *J Clin Microbiol* 2015;53(10):3334-3340.

- 791 56. **Vidovic S, An R, Rendahl A.** Molecular and Physiological Characterization of
792 Fluoroquinolone-Highly Resistant *Salmonella* Enteritidis Strains. *Front Microbiol*
793 2019;10:729.
- 794 57. **Kaldhone PR, Carlton A, Aljahdali N, Khajanchi BK, Sanad YM et al.** Evaluation of
795 Incompatibility Group I1 (IncI1) Plasmid-Containing *Salmonella enterica* and Assessment of
796 the Plasmids in Bacteriocin Production and Biofilm Development. *Front Vet Sci*
797 2019;6(298):298.
- 798 58. **Zhou X, Li M, Xu L, Shi C, Shi X.** Characterization of Antibiotic Resistance Genes,
799 Plasmids, Biofilm Formation, and In Vitro Invasion Capacity of *Salmonella* Enteritidis
800 Isolates from Children with Gastroenteritis. *Microb Drug Resist* 2019;25(8):1191-1198.
- 801 59. **Aldrich C, Hartman H, Feasey N, Chattaway MA, Dekker D et al.** Emergence of
802 phylogenetically diverse and fluoroquinolone resistant *Salmonella* Enteritidis as a cause of
803 invasive nontyphoidal *Salmonella* disease in Ghana. *PLoS Negl Trop Dis*
804 2019;13(6):e0007485.
- 805 60. **Smalla K, Heuer H, Gotz A, Niemeyer D, Krogerrecklenfort E et al.** Exogenous isolation
806 of antibiotic resistance plasmids from piggery manure slurries reveals a high prevalence and
807 diversity of IncQ-like plasmids. *Appl Environ Microbiol* 2000;66(11):4854-4862.
- 808 61. **Castellanos LR, van der Graaf-van Bloois L, Donado-Godoy P, Leon M, Clavijo V et al.**
809 Genomic Characterization of Extended-Spectrum Cephalosporin-Resistant *Salmonella*
810 *enterica* in the Colombian Poultry Chain. *Front Microbiol* 2018;9:2431.
- 811 62. **Mastrorilli E, Pietrucci D, Barco L, Ammendola S, Petrin S et al.** A Comparative
812 Genomic Analysis Provides Novel Insights Into the Ecological Success of the Monophasic
813 *Salmonella* Serovar 4,[5],12:i. *Front Microbiol* 2018;9:715.
- 814 63. **Wong MH, Kan B, Chan EW, Yan M, Chen S.** IncI1 Plasmids Carrying Various blaCTX-
815 M Genes Contribute to Ceftriaxone Resistance in *Salmonella enterica* Serovar Enteritidis in
816 China. *Antimicrob Agents Chemother* 2016;60(2):982-989.
- 817 64. **Garcia-Fernandez A, Chiaretto G, Bertini A, Villa L, Fortini D et al.** Multilocus
818 sequence typing of IncI1 plasmids carrying extended-spectrum beta-lactamases in
819 *Escherichia coli* and *Salmonella* of human and animal origin. *J Antimicrob Chemother*
820 2008;61(6):1229-1233.
- 821 65. **Kameyama M, Chuma T, Yokoi T, Yabata J, Tominaga K et al.** Emergence of
822 *Salmonella enterica* serovar infantis harboring IncI1 plasmid with bla(CTX-M-14) in a
823 broiler farm in Japan. *J Vet Med Sci* 2012;74(9):1213-1216.

- 824 66. **Pulford CV, Perez-Sepulveda BM, Canals R, Bevington JA, Bengtsson RJ et al.**
825 Stepwise evolution of *Salmonella* Typhimurium ST313 causing bloodstream infection in
826 Africa. *Nat Microbiol* 2021;6(3):327-338.
- 827 67. **Silva CA, Blondel CJ, Quezada CP, Porwollik S, Andrews-Polymeris HL et al.** Infection
828 of mice by *Salmonella enterica* serovar Enteritidis involves additional genes that are absent
829 in the genome of serovar Typhimurium. *Infect Immun* 2012;80(2):839-849.

830

831

832

833

834

835

836

837

838

839

1 **Table 1. Antibimicrobial drug class and plasmid class of the MDR associated STs from different MGT levels.**

MGT-ST ^a	MGT4-CCs	Total No. of isolates	No. of isolates with MDR (%)	% of the MDR isolates ^d							% of isolates with plasmid ^d							Continent ^e
				Aminoglycoside	Beta-lactam	Sulphonamide	Tetracycline	Phenicol	Trimethoprim	Macrolide	Col156_1	ColRNAI_1	ColpVC_1	IncII_1_Alpha	IncI2_1_Delta	IncN_1	IncQ1_1	
MGT3-ST15	CC16 ^c	261	259 (99%)	98%	74%	97%	89%	95%	96%			1%	14%	1%		95%		Africa/EU/NA/OC
MGT3-ST10	CC11 ^c	141	122 (87%)	74%	84%	86%	82%	77%	82%			1%	95%	1%	4%	71%		Africa/EU/NA/OC
MGT3-ST30	CC1	10	10 (100%)	100%		100%	100%	100%	100%							80%		EU/NA
MGT3-ST161 ^b	CC1	79	29 (37%)	30%	84%	35%	84%	32%	3%	3%	1%	100%	86%				1%	EU
MGT3-ST9	CC1	70	66 (94%)	93%	80%	91%	80%	1%	1%	1%	1%		3%				99%	EU/NA/OC
MGT3-ST301	CC1	11	11 (100%)	100%	100%				100%							100%		NA/EU
MGT4-ST354	CC1	39	39 (100%)	100%	100%	100%	62%				3%		5%				100%	EU/NA/OC
MGT4-ST54	CC1	29	28 (97%)	90%	41%	90%	86%										97%	EU/NA/OC
MGT5-ST86	CC1	12	12 (100%)	100%	100%	100%							8%				100%	NA/EU/OC
MGT5-ST682	CC1	10	10 (100%)	100%		100%	100%	90%					30%		100%		100%	NA
MGT6-ST2698	CC13	39	35 (90%)	90%	90%	90%	90%								90%		100%	NA
Total		701	621 (89%)															

2 ^aMGT-STs with more than 80% of the isolated harbouring AR genes to three or more drug types were identified. Isolates sets in MDR associated
 3 MGT-STs were mutually exclusive. Note *aac(6')-Iaa_1* gene for aminoglycoside resistance, which was present in almost all isolates, was
 4 excluded for the AR analysis.

5 ^bMGT3-CC161 include 29 isolates with MDR genes (to five drug types) and 38 isolates with AR genes to two different drug types.

6 ^cMGT4-CC16 and CC11 belonged to MGT3-CC15 and CC10, which were representative of the two African lineages.

7 ^dThe gradient green to gray colours reflect the proportion of isolates from 100% to 1%.

8 ^eContinents in which the MDR related STs were observed: EU refers to Europe, NA to North America and OC to Oceania.

9

1 **Table 2. Virulence and SPI genes that distributed differently in the STs/CCs represented phylogeny of *S. Enteritidis*.**

	Clade	No. of isolates	The virulence gene present in >80% of the isolates in each ST/CC											SPI (No. of genes present in >= 80% of isolates) ^a																			
			<i>shdA</i>	<i>sodCI</i>	<i>flhC</i>	<i>sseI/srfH</i>	<i>spvB, spvC, spvR</i>	<i>pefA, pefB</i>	<i>pefC, pefD</i>	<i>rck</i>	<i>sseK2</i>	<i>gogB</i>	<i>grvA</i>	<i>hsiCI/vipB</i>	SPI-1 (45)	SPI-5 (10)	SPI-6 (59)	SPI-7 (149)	SPI-10 (29)	SPI-11 (17)	SPI-17 (6)	6I-IDS (31)	SPI-22 (16)	SPI-23 (11)									
MGT1-ST3304	A	47	+																					42	10 ^b	35	10	9	12	1	31 ^b	2	5
MGT1-ST180	A	37	+																					42	10 ^b	35	10	9	12	1	31 ^b	2	2
MGT1-ST1972	C	33		+	+																			45 ^b	10 ^b	37	8	10	13	6 ^b	30	2	6
MGT1-ST183	B	410		+	+	+	+							+	+	+								42	8	35	44	10	13	6 ^b	31 ^b	2	5
MGT3-CC10	B	164		+	+	+	+																	42	10 ^b	19	10	10	13	6 ^b	15	2	2
MGT3-CC15	B	274		+	+	+	+	+	+															42	10 ^b	19	49	10	13	6 ^b	15	2	2
MGT3-CC107	B	364		+	+	+	+	+	+				+											42	10 ^b	18	11	10	13	6 ^b	15	2	3
MGT3-CC18	B	83		+	+	+	+	+	+				+	+										42	10 ^b	18	11	10	13	6 ^b	14	2	2
MGT4-CC101	B	102		+	+	+	+	+	+	+														42	9	19	11	10	17 ^b	6 ^b	15	2	5
MGT4-CC129	B	199	+	+	+	+	+	+	+	+														42	10 ^b	19	10	10	13	6 ^b	15	2	5
MGT4-CC30	B	135		+	+	+	+	+	+	+														42	10 ^b	19	10	10	13	6 ^b	15	2	5
MGT4-CC1	B	8539		+	+	+	+	+	+	+														42	10 ^b	19	11	9	13	6 ^b	15	2	5
MGT4-CC13	B	14927		+	+	+	+	+	+	+														42	10 ^b	19	45	10	13	6 ^b	15	2	5

2

3 ^aSPI-2, 3, 4, 9,12,13,14 and 16 were found intact in all of the STs/CCs; SPI-8, 15, 18, 20 and 21 were absent in all STs/CCs.

4 ^bThe SPI was predicted to be intact in the ST/CC with all genes observed.

5 **Figure legends:**

6 **Fig. 1. Makeup and assignments of each *S. Enteritidis* MGT level. a.** Number of loci included in
7 each MGT level. The first eight levels are composed of *Salmonella enterica* core genes, which are
8 orthologous to those of the *S. Typhimurium* MGT1 to MGT8, except for one gene at MGT8 which
9 was excluded due to duplication in *S. Enteritidis*. The MGT9 scheme includes core genes of *S.*
10 *Enteritidis* (those not belonging to the *Salmonella* core coloured in yellow) and core intergenic
11 regions (coloured in green). The number behind or within each bar refers to the number of loci
12 included. **b.** The number of sequence types (ST) and clonal complexes (CC) assigned at each MGT
13 level among the 26,670 genomes analysed. CC includes STs of no more than one allele difference.
14 The numbers refer to the number of STs or CCs types assigned at each level, which were also
15 indicated by colour. As MGT9 included 4986 loci with highest subtyping resolution, the 26,670
16 isolates were subtyped into 20,153 different ST types and 14,441 CC clustering types at MGT9.

17

18 **Fig. 2. MGT4 STs offer a useful description of continent specific and global clades. a.** At each
19 MGT level from MGT2 to MGT8, we identified continent restricted STs where more than 70% of
20 the isolates belonged to the same continent. MGT4 was found to have the highest number of isolates
21 belonged to these continental limited STs. **b.** There were 45 MGT4-STs of more than 50 isolates
22 representing 73.5% of the total 26670 isolates, and the majority of these STs belonged to MGT4-
23 CC1 and CC13. The number, size and continental makeup of these STs is shown with the left Y axis
24 showing MGT4 STs, the right Y axis showing MGT4 CCs and the X axis showing number of
25 isolates, continental distribution of each ST is shown by different colours in each row.

26

27 **Fig. 3. USA and UK *S. Enteritidis* isolates can be distinguished by STs at MGT4 and MGT5. a.**
28 At MGT4, there were 39 STs (50 or more isolates each) representing 75.1% of the UK and USA
29 isolates (22191 in total). MGT4-ST99, ST136 and ST135 were prevalent in the USA, MGT4-ST15,

30 ST25 and ST13 were prevalent in the UK, while MGT4-ST15 and ST25 were mixed. **b.** MGT4-
31 ST15 and ST25 can be subtyped into multiple STs at MGT5, the majority of which were prevalent in
32 the UK, except for MGT5-ST412 and ST387 in the USA.

33

34 **Fig. 4. MGT4 ST proportions across states in the USA is similar but only a subset of STs are**
35 **associated with avian hosts.** For the USA *S. Enteritidis* isolates with metadata (4,383 isolates), a
36 total of 3935 genomes were from either avian or human source, 95.0% (3738/3935) of which
37 belonged to MGT4-CC1 and CC13. The MGT4-STs belonged MGT4-CC1 and CC13 of either avian
38 or human source were shown in different states of the USA. Each pie chart illustrated the MGT-ST
39 types and the size of the main STs represented with different colours. The size of the pie indicated
40 the total number of isolates in each state. While CC13 appears to originate in birds before moving to
41 humans, the reservoir of CC1 is unknown. Origins for CC1 STs were also not observed in any other
42 source type. The maps were created with Tableau v2019.2.

43

44 **Fig. 5. MGT5 STs allow identification of temporal patterns in strain diversity in the UK.** There
45 were 13 MGT5-STs of more than 100 isolates shown with different colours. Temporal patterns
46 indicate that some STs are endemic while others likely cause outbreaks. MGT5-ST156 in red was
47 observed predominantly from June to August of 2014, and disappeared after 2014. MGT5-ST423
48 was a dominant type from March to September of 2015, and became dominant again during the
49 September to December of 2016.

50

51 **Fig. 6. Global phylogenetic structure of *S. Enteritidis*.** A total of 1508 genomes were randomly
52 sampled from each ST in MGT6. A phylogenetic tree was constructed by ParSNP v1.2, which called
53 core-genome SNPs extracted from alleles. The tree scales indicated the 0.01 substitutions per locus.
54 The number on internal branches represented the percentage of bootstrap support. **a.** Three clades

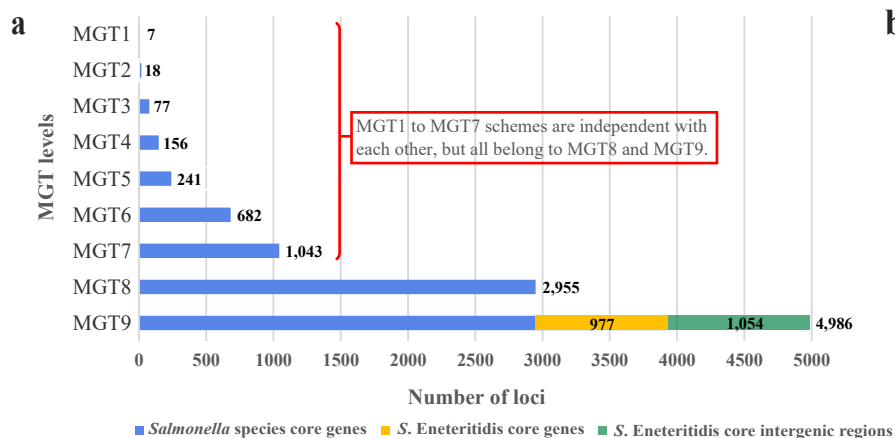
55 were identified which were concordant with previous findings: The predominant clade, highlighted
56 in yellow, is the global epidemic clade which was phylogenetically closer to *S. Gallinarum* than the
57 other two clades in purple and green background which are limited in Oceania 9. **b.** The main,
58 yellow highlighted clade was expanded into a more detailed phylogeny with *S. Gallinarium* as an
59 outgroup. 10 main lineages were defined where each lineage had more than 10 isolates. STs of the
60 seven gene MLST (or MGT1) and CCs from MGT2 to MGT4 are shown in different colours for each
61 of the 10 lineages. The pie charts for each lineage represent the proportion of the isolates belonged to
62 the same MGT4-CC shown on the right side.

63

64 **Fig. 7. Skyline population size estimation of MGT4-CC1 and CC13.** Bayesian skyline model and
65 strict clock were estimated to be the optimal combination. The vertical axis column indicates the
66 predicted log scale effective population size and the horizontal axis shows the predicted time. The
67 blue line represents the median posterior estimate of the effective population size, and the blue area
68 shows the upper and lower bounds of the 95% highest posterior density interval. **a.** Skyline
69 population expansion of MGT4-CC1. Around 1950 and 1970, there were two large population
70 expansions within MGT4-CC1. **b.** Skyline population expansion of MGT4-CC13. The population of
71 MGT4-CC13 gradually expanding until around 1970, when there was also an accelerated expansion.

72

73



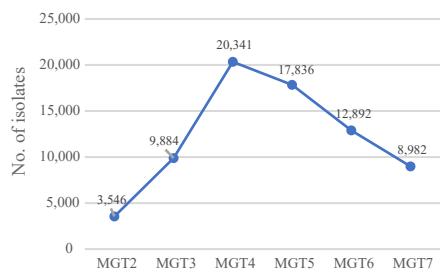
b

MGT	No. of STs	No. of CCs
MGT2	253	23
MGT3	1,222	179
MGT4	2,236	423
MGT5	4,128	1,017
MGT6	7,262	2,838
MGT7	11,107	5,680
MGT8	17,036	11,223
MGT9	20,153	14,441

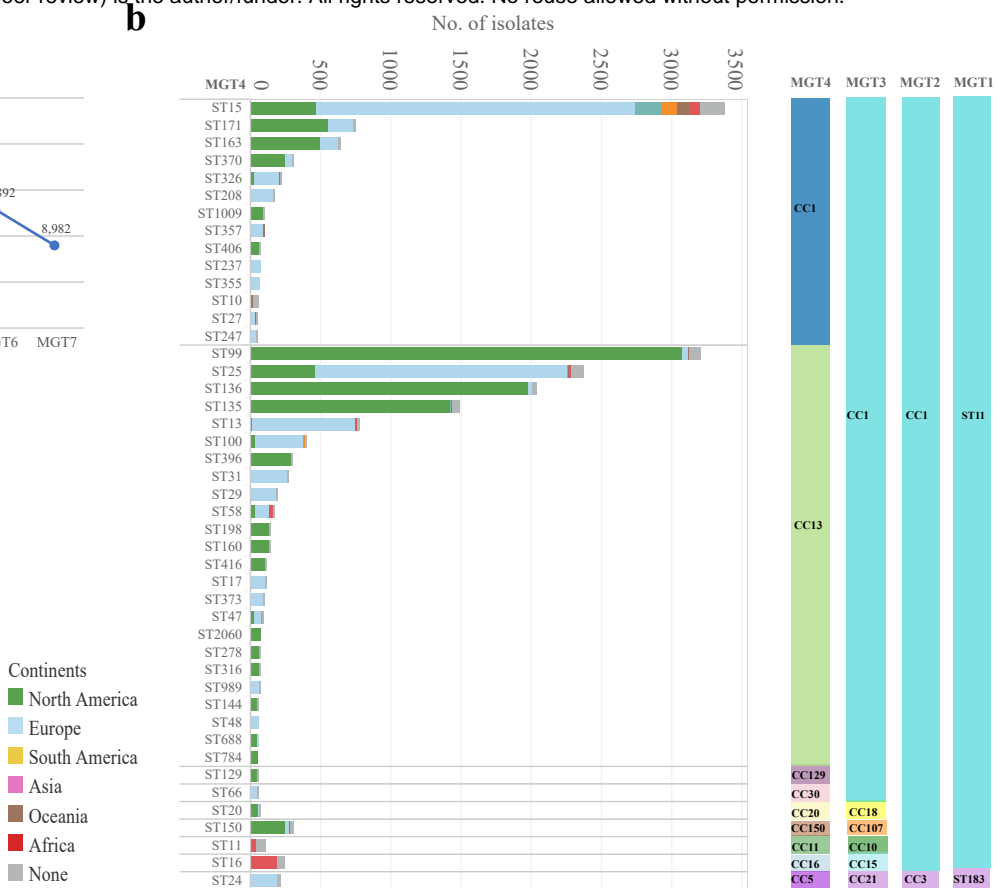
No. of types

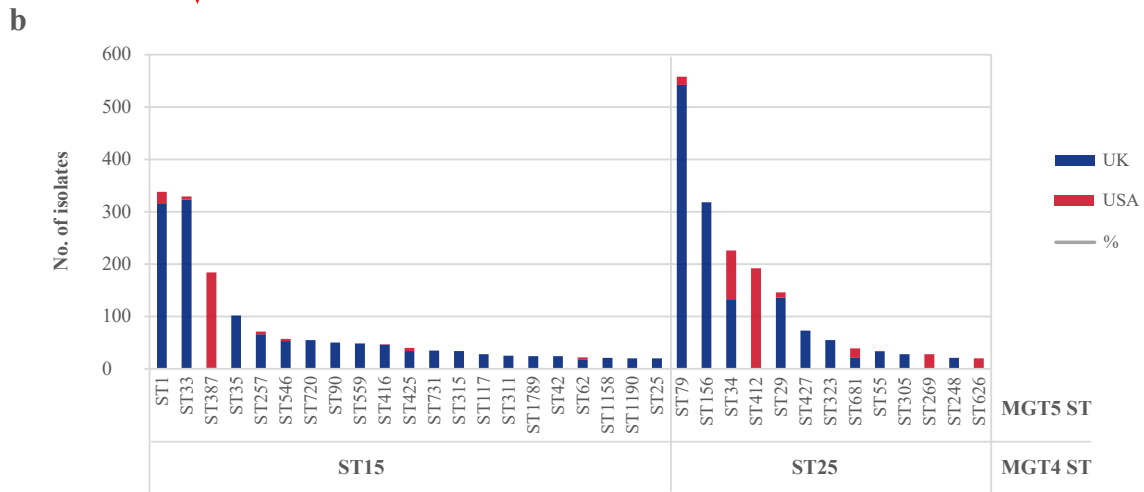
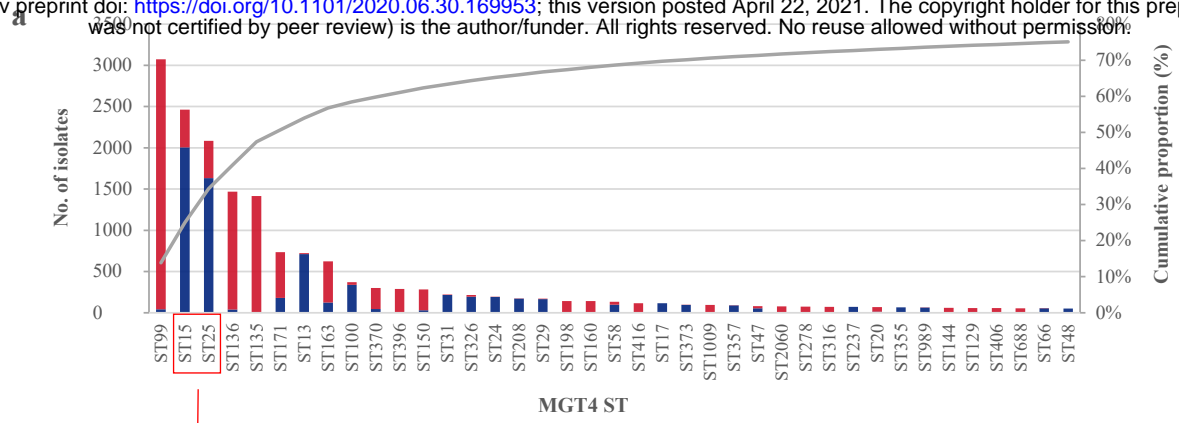
 23 20,153

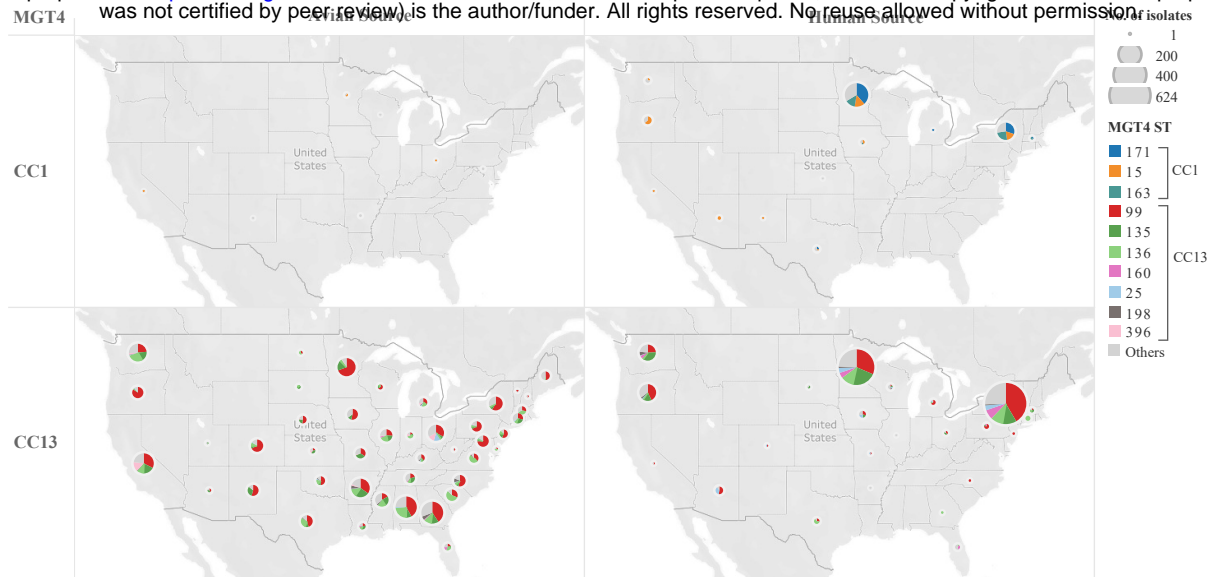
a

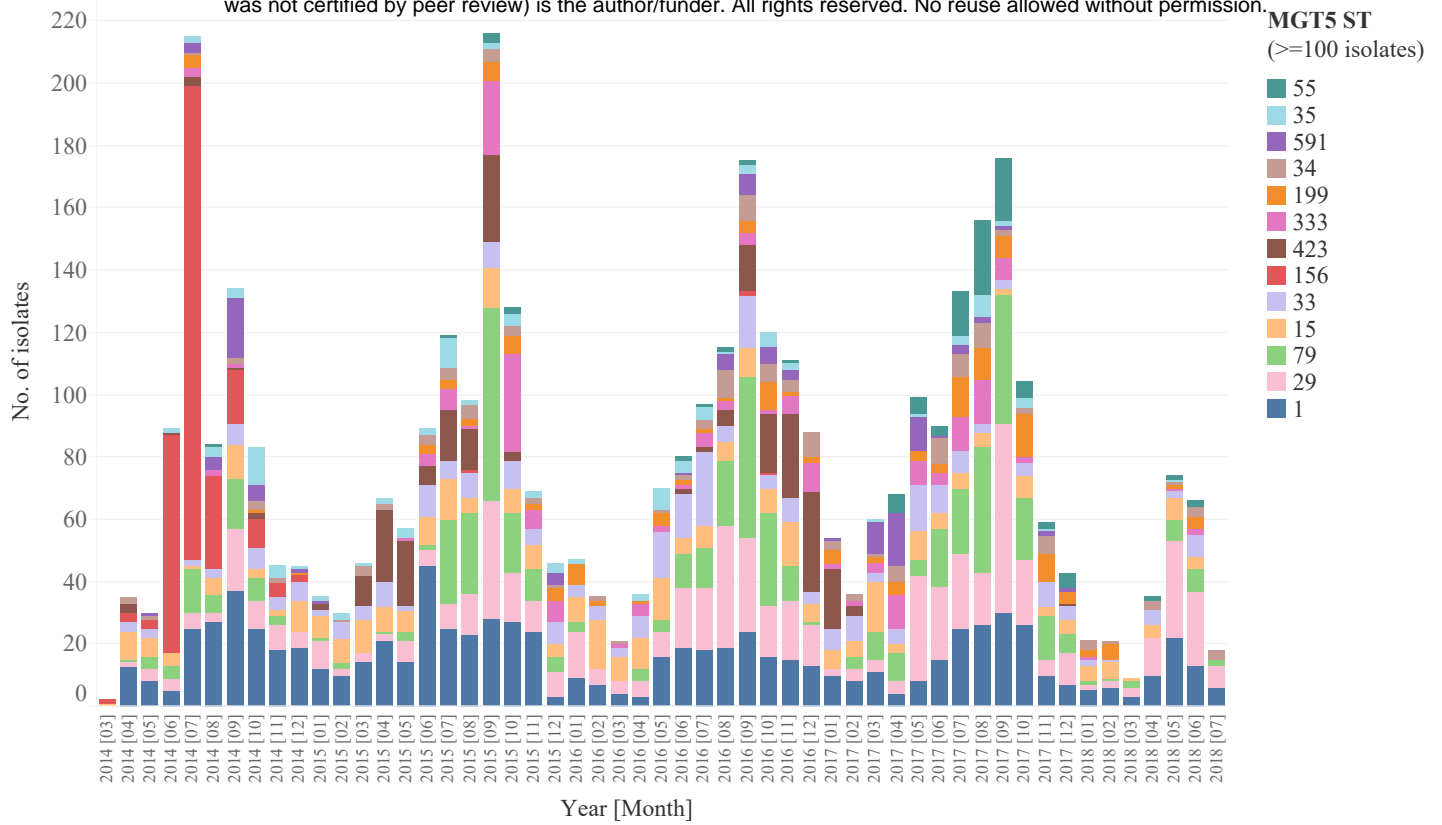


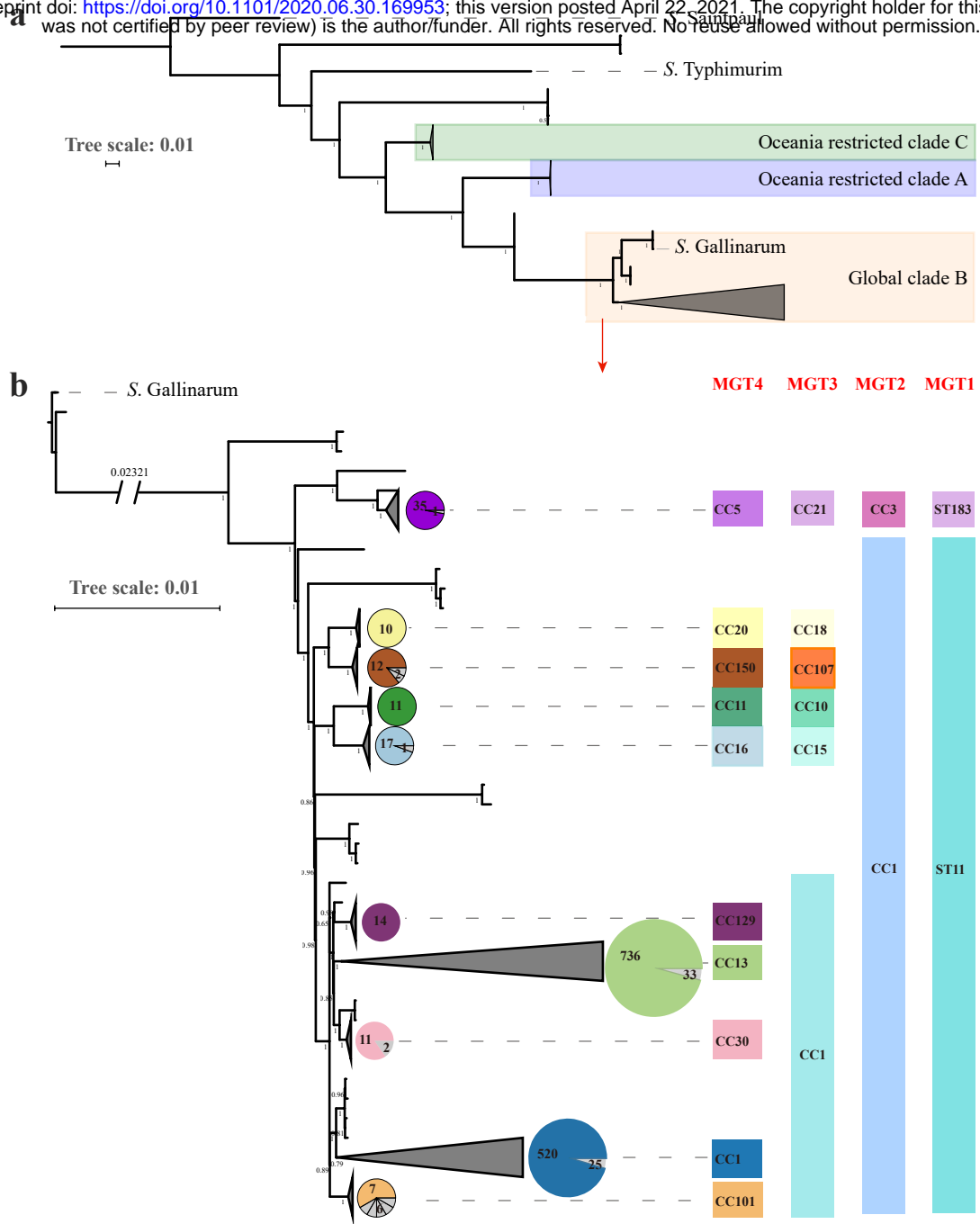
b





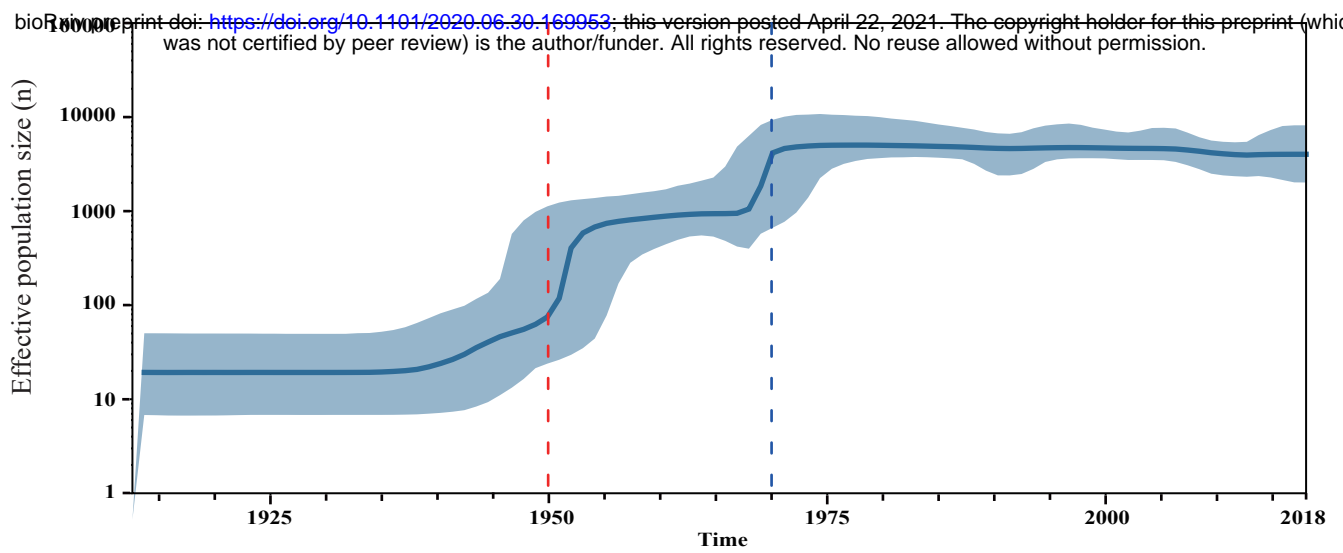






a**MGT4 CC1**

bioRxiv preprint doi: <https://doi.org/10.1101/2020.06.30.169953>; this version posted April 22, 2021. 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.

**b****MGT4 CC13**