



HAL
open science

Global phylogenomics of multidrug resistant *Salmonella enterica* serotype Kentucky ST198

Jane Hawkey, Simon Le Hello, Benoît Doublet, Sophie Granier, Rene S Hendriksen, W. Florian Fricke, Pieter-Jan Ceyskens, Camille Gomart, Helen Billman-Jacobe, Kathryn E. Holt, et al.

► **To cite this version:**

Jane Hawkey, Simon Le Hello, Benoît Doublet, Sophie Granier, Rene S Hendriksen, et al.. Global phylogenomics of multidrug resistant *Salmonella enterica* serotype Kentucky ST198. 2019. pasteur-02079525

HAL Id: pasteur-02079525

<https://pasteur.hal.science/pasteur-02079525v1>

Preprint submitted on 26 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

1 Global phylogenomics of multidrug resistant *Salmonella enterica* serotype Kentucky 2 ST198

3
4
5 Jane Hawkey^{1,2#‡}, Simon Le Hello^{3#}, Benoît Doublet⁴, Sophie Granier^{5,6}, Rene Hendriksen⁷,
6 W. Florian Fricke^{8,9}, Pieter-Jan Ceysens¹⁰, Camille Gomart³, Helen Billman-Jacobe¹¹,
7 Kathryn E. Holt^{1,2,12#}, François-Xavier Weill^{3#‡}

8
9 ¹Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute,
10 University of Melbourne, Parkville, Victoria 3010, Australia.

11 ²Department of Infectious Diseases, Central Clinical School, Monash University, Melbourne, Victoria 3004,
12 Australia.

13 ³Unité des Bactéries Pathogènes Entériques, Centre National de Référence des *Escherichia coli*, *Shigella* et
14 *Salmonella*, World Health Organisation Collaborative Center for the typing and antibiotic resistance of
15 *Salmonella*, Institut Pasteur, 75015, Paris, France.

16 ⁴ISP, Institut National de la Recherche Agronomique, Université François Rabelais de Tours, UMR 1282,
17 Nouzilly, France.

18 ⁵Université PARIS-EST, Agence Nationale de Sécurité Sanitaire de L'Alimentation, de L'Environnement et du
19 Travail (ANSES), Laboratory for Food Safety, 94701 Maisons-Alfort, France.

20 ⁶Agence Nationale de Sécurité Sanitaire de L'Alimentation, de L'Environnement et du Travail (ANSES),
21 Fougères Laboratory, 35306 Fougères, France

22 ⁷Research Group for Genomic Epidemiology, National Food Institute, Technical University of Denmark,
23 Kongens Lyngby, Denmark.

24 ⁸Department of Microbiome Research and Applied Bioinformatics, University of Hohenheim, Stuttgart,
25 Germany

26 ⁹Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, USA.

27 ¹⁰Unit Bacterial Diseases, Sciensano, Brussels, Belgium

28 ¹¹Asia-Pacific Centre for Animal Health, Faculty of Veterinary and Agricultural Science, University of
29 Melbourne, Parkville, 3010, Australia

30 ¹²London School of Hygiene & Tropical Medicine, London, WC1E 7HT, UK.

31
32 #Contributed equally

33
34 ‡Corresponding authors:

35
36 Jane Hawkey: Department of Infectious Diseases, Monash Central Clinical School, 85 Commercial Road,
37 Melbourne, VIC 3004, Australia. E-mail : jane.hawkey@monash.edu

38
39 François-Xavier Weill: Centre National de Référence des *Escherichia coli*, *Shigella* et *Salmonella*, Unité des
40 Bactéries Pathogènes Entériques, Institut Pasteur, 28 rue du docteur Roux, 75015 Paris. Tel: +33-1 45 68 83 45,
41 Fax: +33-1 45 68 88 37. E-mail: francois-xavier.weill@pasteur.fr

42 43 Abstract

44 *Salmonella enterica* serotype Kentucky (*S. Kentucky*) can be a common causative
45 agent of salmonellosis, usually associated with consumption of contaminated poultry.
46 Antimicrobial resistance (AMR) to multiple drugs, including ciprofloxacin, is an emerging
47 problem within this serotype. We used whole-genome sequencing (WGS) to investigate the
48 phylogenetic structure and AMR content of 121 *S. Kentucky* ST198 isolates from five
49 continents (97 sequenced in this study and 24 from the GenomeTrackr project). Population
50 structure was inferred using phylogenomic analysis and whole genomes were compared to
51 investigate changes in gene content, with a focus on acquired AMR genes. Our analysis
52 showed that multidrug resistant (MDR) *S. Kentucky* isolates belonged to a single lineage,
53 which we estimate emerged circa 1989 following the acquisition of the AMR-associated
54 *Salmonella* genomic island 1 (variant SGI1-K) conferring resistance to ampicillin,
55 streptomycin, gentamicin, sulfamethoxazole, and tetracycline. Phylogeographic analysis

56 indicates this clone emerged in Egypt before disseminating into Northern, Southern and
57 Western Africa, then to the Middle East, Asia and the European Union. The MDR clone has
58 since accumulated various substitution mutations in the quinolone resistance determining
59 regions (QRDR) of DNA gyrase (*gyrA*) and DNA topoisomerase IV (*parC*), such that most
60 strains carry three QRDR mutations which together confer resistance to ciprofloxacin. The
61 majority of AMR genes in the *S. Kentucky* genomes were carried either on plasmids or SGI
62 structures. Remarkably, each genome of the MDR clone carried a different SGI1-K
63 derivative structure; this variation could be attributed to IS26-mediated insertions and
64 deletions, which appear to have hampered previous attempts to trace the clone's evolution
65 using sub-WGS resolution approaches. Several different AMR plasmids were also identified,
66 encoding resistance to chloramphenicol, third-generation cephalosporins, carbapenems,
67 and/or azithromycin. These results indicate that most MDR *S. Kentucky* circulating globally
68 result from the clonal expansion of a single lineage that acquired chromosomal AMR genes
69 30 years ago, and has continued to diversify and accumulate additional resistances to last-line
70 oral antimicrobials.

71 **Impact Statement**

72 Fluoroquinolone resistant *Salmonella enterica* and carbapenem resistant, extended
73 spectrum beta-lactamase (ESBL) producing Enterobacteriaceae are amongst the highest
74 priority pathogens posing a risk to human health as determined by the World Health
75 Organisation (WHO). All of these high level resistances have been detected in a single
76 serotype of *S. enterica*, *S. Kentucky*, against a background of multidrug resistance to first-line
77 antimicrobials, leaving very limited treatment options. Here, we analysed the genomes of *S.*
78 *Kentucky* from geographically diverse sources, to investigate the emergence and spread of
79 antibiotic resistance in this problem pathogen. We discovered that the multidrug resistant
80 (MDR) genomes in our collection comprised a clonal MDR lineage that we estimate arose in
81 Egypt in ~1989, before spreading across Africa, then into Europe, the Middle East and Asia.
82 Resistance to first-line antibiotics mostly arose from the chromosomal integration of a large
83 genomic island, the *Salmonella* Genomic Island 1 (SGI1), in the common ancestor of the
84 MDR lineage. Most strains were also fluoroquinolone resistant, due to acquisition of point
85 mutations in chromosomal genes *gyrA* and *parC* early in the clone's evolution. Additional
86 resistances, including to third-generation cephalosporins (such as ceftriaxone), carbapenems
87 (such as imipenem), and the last-line oral antibiotic azithromycin, emerged through
88 acquisition of diverse locally circulating MDR plasmids. Aside from antibiotic resistance, we
89 found no other genetic determinants that could explain the global success of this *S. Kentucky*
90 lineage. These data show the MDR clone of *S. Kentucky* is already widespread and is capable
91 of acquiring last-line resistances, suggesting it should be considered a high-risk global MDR
92 clone.

94 **Data Summary**

95 All sequencing reads generated in this study have been deposited in project
96 PRJNA445436. SRA accession numbers can be found in **Supplementary Table 1**.

97
98 The reference genome sequence for *S. Kentucky* strain 201001922 has been deposited
99 into GenBank under accession CP028357.

100
101 The phylogeny and associated metadata can be viewed on Microreact:

102 <https://microreact.org/project/Hkl7CzEXV>

103

104 **The authors confirm all supporting data, code and protocols have been provided within**
105 **the article or through supplementary data files.**

106 Introduction

107 Carbapenem-resistant, extended-spectrum beta-lactamase (ESBL)-producing
108 Enterobacteriaceae and fluoroquinolone-resistant *Salmonella* have been recently listed as
109 priority pathogens that pose the greatest threats to human health (critical and high threat
110 levels, respectively) by the World Health Organisation (WHO) (1). All these resistances have
111 been observed in a single serotype of *Salmonella enterica*, Kentucky (*S. Kentucky*), since the
112 2000s (1-4). Ciprofloxacin-resistant *S. Kentucky* (CIP^R *S. Kentucky*) was first observed in a
113 French traveller returning from Egypt in 2002, before being increasingly isolated globally
114 (1). Between 2007 and 2012, the European Centers for Disease Control and Prevention
115 (ECDC) reported 1301 isolations of *S. Kentucky* from 12 countries, including 955 (73.4%)
116 CIP^R *S. Kentucky* (5). These isolates were found in patients across the world, but
117 predominantly in Northern Africa, Europe, and Southern Asia. Several previous studies have
118 described the rapid spread of CIP^R *S. Kentucky* from Northern Africa to the rest of the
119 African continent, as well as the Middle East, Europe and Asia (2-4). CIP^R *S. Kentucky* is a
120 foodborne pathogen that causes gastroenteritis in humans, and domestic poultry has played an
121 important role in its global spread (most recently in South Asia and Europe). Multi-locus
122 sequencing typing (MLST) and pulsed-field gel electrophoresis (PFGE) have revealed that
123 CIP^R *S. Kentucky* is a single population belonging to sequence type (ST) 198 and not ST152,
124 which is a prevalent *S. Kentucky* ST found in poultry in the United States of America (USA)
125 but rarely reported in humans (7).

126
127 Before the 1990s, *S. Kentucky* ST198 was susceptible to all antibiotics. Since then,
128 multidrug resistance has emerged (1). In the early 1990s, *S. Kentucky* ST198 acquired a
129 variant of the *Salmonella* genomic island 1 (SGI1) into the chromosome, likely in Egypt (6).
130 Initially characterised in *S. enterica* serotype Typhimurium strain DT104 (7), the SGI1 is a
131 site-specific Integrative Mobilizable Element (IME) that integrates in the 3'-end of the
132 conserved chromosomal gene *trmE* (8). SGI1 is the prototype element of a multidrug
133 resistance IME family named SGI/PGI/AGI which includes both *Proteus* genomic islands
134 (PGI) (9) and *Acinetobacter* genomic islands (AGI) (12). They consist of a 27 kbp related
135 backbone with conserved gene synteny and variable regions containing complex class 1
136 integron structures, IS, and Tn elements that are responsible for multidrug resistance. As an
137 IME, SGI1 is specifically mobilized *in trans* by conjugative IncC plasmids (10-12). The most
138 recent findings revealed complex interactions between SGI1 and IncC plasmids for transfer
139 and maintenance. Since the first description of SGI1 in *S. Typhimurium* DT104, several
140 variants of SGI/PGI/AGI have been discovered, which differ in their antimicrobial resistance
141 (AMR) gene content and AMR gene cluster structure (13,14) in species of families
142 *Enterobacteriaceae*, *Morganellaceae*, and *Acinetobacter baumannii* (12,18). These variants
143 usually differ in the composition of the integron, and each variant carries different AMR
144 genes. One variant of the SGI, known as SGI2 or SGI1-J, differs not only in the composition
145 of the integron, but also in the site at which the integron is inserted into the SGI backbone
146 (6,15).

147
148 Four main types of SGI have so far been described in *S. Kentucky*: SGI1-K, SGI1-P,
149 SGI1-Q (see **Figure 2**) and SGI2 (3). These SGI1 variants share a common genetic feature
150 consisting of an insertion/deletion between *S005* and *S009* due to the insertion of *IS1359*,
151 which was also found in a few other SGI1 variants in strains of different *S. enterica* serotypes
152 isolated in 2000 in Egypt, and more recently in *P. mirabilis* (16). Additionally, these three
153 SGI1 variants show a truncation the 5'-end of *S044*, the final ORF of the SGI backbone,
154 through the insertion of *IS26* (17). SGI1-K contains a complex mosaic resistance region
155 made of different segments of transposons *Tn21*, *Tn1721*, *Tn5393*, *Tn3*-like, and a *In4*-type

156 integron structure, as well as IS26 elements (18). SGI1-P and SGI1-Q contain only the IS26-
157 flanked Tn3-like structure carrying *bla*_{TEM-1} and only the rightmost IS26 in S044,
158 respectively (17).

159

160 After the acquisition of SGI1 by the multidrug resistant (MDR) lineage, high level-
161 resistance to fluoroquinolones emerged, conferred by a combination of three amino-acid
162 substitutions in the quinolone resistance-determining region (QRDR) of *gyrA* and *parC*.
163 Previous epidemiological studies determined that these mutations likely arose in Egypt in the
164 early 2000s (2).

165

166 Finally, additional resistance was gained through the acquisition of locally circulating
167 plasmid-borne ESBL, AmpC and/or carbapenemase genes (4). Additionally, the geographic
168 distribution of ciprofloxacin-resistant (CIP^R) *S. Kentucky* ST198 overlaps with other highly
169 drug resistant Enterobacteriaceae carrying plasmid-borne ESBL, AmpC and/or
170 carbapenemase genes, leading to predictions that highly-drug resistant *S. Kentucky* ST198
171 strains are likely to become more frequent in the near future due to novel plasmid
172 acquisitions (3,4).

173

174 To date, all previous studies have used conventional typing methods (MLST, PCR,
175 PFGE, and antimicrobial susceptibility testing (AST)) and together they suggest that the
176 recent global spread of CIP^R *S. Kentucky* may reflect the expansion of a single clone, driven
177 by the emergence of AMR. However, the precise nature, order, and timing of the
178 evolutionary events underlying this overall picture, remain unclear. Here we investigate the
179 global population structure of MDR *S. Kentucky* ST198 using whole-genome sequencing
180 (WGS) and phylogenomic analysis to interrogate a collection of 121 human and non-human
181 isolates collected from 33 countries on five continents, between 1937 and 2016. We use
182 comparative genomics to reconstruct the various steps in the acquisition of AMR
183 determinants within the emerging MDR *S. Kentucky* ST198 clone, and to investigate the
184 presence of genetic elements not related to AMR that might have conferred other selective
185 advantages to this emerging bacterial pathogen.

186

187 **Methods**

188

189 **Bacterial isolates used in this study**

190 A total of 97 *S. Kentucky* ST198 isolates were directly analysed in this study (**Table**
191 **S1**), including 68 isolates collected between 1937 and 2013 that were previously studied by
192 conventional molecular methods (2-4,19), and 29 new isolates collected between 2008 and
193 2016. These isolates originated from the French National Reference Center for *E. coli*,
194 *Shigella*, *Salmonella* (Institut Pasteur) and several other international laboratories and were
195 selected on the basis of their diversity (human or non-human source, geographic area and
196 year of isolation, PFGE types, and antimicrobial resistance phenotypes and genotypes). WGS
197 data for a further 24 *S. Kentucky* isolates was included in genomic analyses as detailed
198 below.

199

200 **Antimicrobial susceptibility testing**

201 AST was performed on all 97 *S. Kentucky* ST198 isolates using the disk diffusion
202 method with a panel of 32 antimicrobial agents (Bio-Rad, Marnes-La-Coquette, France) as
203 described previously (24). The minimum inhibitory concentrations (MICs) of ceftriaxone,
204 ceftazidime, imipenem, ertapenem, meropenem, ciprofloxacin, azithromycin, and tigecycline
205 were also determined by Etests (AB Biodisk, Solna, Sweden). Results were interpreted with

206 the Antibiogram Committee of the French Society for Microbiology/European Committee on
207 Antimicrobial Susceptibility Testing (CA-SFM/EUCAST) (www.sfm-microbiologie.org/)
208 breakpoints. In particular, we used ciprofloxacin clinical breakpoints defined for intestinal
209 *Salmonella* isolates: susceptible when MIC \leq 0.25 mg/L and resistant when MIC $>$ 0.5 mg/L.
210

211 **Whole-genome sequencing**

212 The 97 *S. Kentucky* ST198 isolates were subjected to WGS with Illumina at GATC
213 Biotech (Konstanz, Germany, Illumina HiSeq) ($n=45$), the Institut Pasteur (PF1 and P2M
214 sequencing platforms, Illumina HiSeq and NextSeq, respectively) ($n=43$), the Technical
215 University of Denmark ($n=7$, Illumina MiSeq), or at the Institute for Genome Sciences,
216 University of Maryland School of Medicine (IGS-UoM, Illumina HiSeq) ($n=2$). Paired-end
217 reads varied in read length depending on the sequencing platform/site, from 100 to 146 bp,
218 yielding a mean of 196-fold coverage per isolate (minimum 30-fold, maximum 687-fold)
219 (**Table S1**). Short-read sequences have been deposited at the European Nucleotide Archive
220 (ENA) (<http://www.ebi.ac.uk/ena>), under study accession number PRJNA445436 and the
221 genome accession numbers are provided in **Table S1**.
222

223 **Other studied genomes**

224 Additional *S. Kentucky* ST198 WGS data were obtained from the GenomeTrakr
225 project (<https://ftp-trace.ncbi.nih.gov/pathogen/Results/Salmonella>) (20,21). All 3,014 *S.*
226 *Kentucky* isolates in the *Salmonella* project were downloaded from NCBI on 2016-01-06,
227 and ST was determined using SRST2 (22). From the 73 available ST198 GenomeTrakr
228 sequences, we excluded those that were missing the source information required for our
229 analysis (source, location and year of isolation), and retained those from geographic regions
230 underrepresented in our own dataset that were non-redundant in terms of source/outbreak
231 ($n=24$; accessions in **Table S1**), bringing the total number of genomes analysed in this study
232 to 121.
233

234 **Sequencing and construction of reference genome 201001922**

235 Genomic DNA from *S. Kentucky* ST198 isolate 201001922 was also sequenced using
236 a hybrid sequencing approach at IGS-UoM, as previously described (26). Paired-end, 3-kb
237 insert libraries sequenced on the 454 GS FLX Titanium platform (Roche, Branford, CT) were
238 combined with paired-end, 300 to 400-bp insert libraries sequenced with 100-bp read length
239 on the HiSeq 2000 platform (Illumina, San Diego, CA). Hybrid assemblies were generated
240 with the Celera assembler (<http://wgs-assembler.sourceforge.net/wiki/>) based on different
241 ratios of 454 and Illumina sequence data and the outputs were compared with respect to the
242 number of resulting scaffolds and total scaffold length. For the final assembly, a 27-fold
243 genome coverage of 454 and a 30-fold coverage of Illumina sequence data were combined to
244 create a draft genome sequence consisting of eleven scaffolds and a total length of 4.86 Mbp.
245

246 Contigs and scaffolds from the draft assembly were concatenated using a linker sequence
247 (NNNNCACACACTTAATTAATTAAGTGTGTGNNNN), in order to generate continuous
248 "pseudochromosomes". The linker sequence contains START and STOP codons in each
249 frame and orientation, to allow the gene finder to call truncated genes at all contig ends.
250 Contig orders and orientations within the pseudochromosome were determined based on
251 NUCmer v3.23 (23) nucleotide sequence comparison to ST152 *S. Kentucky* strain
252 CVM29188 (SL475) as a reference genome. Protein-coding and RNA gene predictions and
253 functional annotations were carried out with CloVR-Microbe (28).
254

255 The genome sequence of *S. Kentucky* ST198 isolate 201001922 has been deposited in
256 GenBank under the accession number CP028357.

257

258 **Mapping and phylogenomic analysis**

259 Short reads for all 121 *S. Kentucky* ST198 isolates were mapped to the reference
260 genome 201001922 using the mapping pipeline RedDog v1b4
261 (<https://github.com/katholt/RedDog>) to identify single nucleotide variants (SNVs) as
262 previously described (24,25). RedDog uses Bowtie2 v2.2.3 (26) with the sensitive local
263 method and a maximum insert size of 2000 to map all genomes to the reference genome.
264 SNVs were then identified using SAMtools v0.0.19 (27) with phred score ≥ 30 , and alleles at
265 each locus were determined by comparing to the consensus base in that genome, using
266 SAMtools pileup to remove low quality alleles (phred base quality ≤ 20 , read depth ≤ 5 or a
267 heterozygous base call). SNVs were filtered to exclude those present in repeat regions, phage
268 regions, or the SGI. Gubbins v1 (28) was run using default settings to identify and remove
269 SNVs in recombinant regions. The final SNV set used for phylogenetic analysis consisted of
270 2,066 SNVs.

271

272 To estimate a Bayesian phylogeny with divergence dates, an alignment of SNV alleles
273 was passed to BEAST v2.4.6 (29), in addition to isolation dates for each genome. The model
274 parameters were as follows: GTR+G substitution model, lognormal relaxed clock, constant
275 population size. As the coefficient of rate variation parameter was calculated to be 0.57 (95%
276 HPD 0.44-0.70), and the distribution was not abutting zero, a relaxed clock model was
277 favoured over a strict clock. The model with a constant population size produced higher
278 overall likelihoods compared to a Bayesian skyline model, and calculations of changes in
279 population size in the skyline model indicated that the population had been constant over
280 time, so the simpler model was favoured. Five independent BEAST runs of 100 million
281 iterations were combined, representing 450 million Markov chain Monte Carlo (MCMC)
282 generations after burn-in removed. Parameter estimates were calculated using Tracer v1.6
283 (30). A maximum clade credibility tree was generated using TreeAnnotator v1.7.5 (31). To
284 test the robustness of the molecular clock signal, ten further BEAST runs with randomised tip
285 dates were generated using the same model.

286

287 Additional testing of the molecular clock was undertaken by constructing a maximum-
288 likelihood phylogeny using RAxML v8.1.23 (32), using 100 bootstrap replicates, with the
289 final set of SNVs. To check for a molecular clock signal, a linear regression was performed
290 using the root-to-tip distances from the phylogeny with year of isolation.

291

292 Phylogeographic analysis was performed by modelling geographic region (defined by
293 the United Nations subregion geoschemes (33)) as a discrete trait on the final BEAST tree,
294 using an empirical Bayes method (34) implemented in the *make.simmap* function in *phytools*
295 v0.6.44 (35).

296

297 **Assembly, annotation and pangenome analysis**

298 All reads were filtered using FastXToolKit v0.0.14 (36) to remove all reads
299 containing bases called as 'N', and Trimmomatic v0.30 (37) was used to remove any reads
300 with an average phred quality score below 30. Each isolate genome was assembled using
301 SPAdes v3.5 (38) using a kmer range of 21, 33, 55, 65 and 75. Scaffolding was performed
302 using SSPACE v3.0 (39) and GapFiller v1.10 (40) with default settings. All assemblies were
303 ordered against the *S. Kentucky* ST198 strain 201001922 reference genome using Abacas
304 v1.3.1 (41). Prokka v1.10 (42) was used to annotate each assembly using a preferential

305 protein database made up of coding sequences from the 201001922 reference genome, the
306 ARG-Annot resistance database (43), the SGI1, SGI1-K and SGI2 references (accessions
307 AF261825, AY463797 and AY963803). Roary v3.6.0 (44) was used to determine core and
308 accessory genes for all annotated genomes. Core genes were defined as present in at least
309 95% of genomes.

310

311 **Identification of resistance, virulence and phage genes**

312 AMR gene alleles were determined by mapping short reads to the ARG-Annot
313 resistance database (43) using SRST2 (22). AMR gene locations were determined by
314 interrogating genome assemblies with BLAST v2.3.0 (45). Associations between AMR genes
315 and SGI type or geographic regions were determined using two-way contingency tables for
316 each gene. Each region was tested with Fisher's Exact Test to determine if the frequency of
317 the gene was positively associated with that specific region compared to all other regions. A
318 p-value cut-off of 0.05 was used to determine significance.

319

320 Presence or absence of *Salmonella* virulence genes defined in the VFDB database
321 (46) was determined using SRST2 to screen the short read data. All genomes were screened
322 using PHASTER (50) to detect phage regions.

323

324 **Reconstruction of SGI sequences**

325 ISMapper v1 (47) and the assembly graph viewer Bandage (48) were used to piece
326 together segments of the SGI in each genome. To do this, each assembly was queried with
327 BLAST to identify which contigs contained SGI backbone and AMR genes. Each assembly
328 was also queried for IS26 using ISMapper's assembly improvement mode (47), identifying
329 contigs that contained IS26 flanking sequence. Contigs containing flanking IS26 sequence
330 with SGI genes or AMR genes were hypothesised to be part of the SGI. Both pieces of
331 information (BLAST and ISMapper results) were used in conjunction with the reference
332 SGI1-K reference sequence (accession AY463797) to determine which contigs could be
333 joined together. In some cases, it was unclear whether IS26-flanked AMR genes were located
334 within the SGI or a plasmid. In these cases, Bandage was used to examine the assembly
335 graphs and determine the paths linking the SGI, IS26 and AMR genes, providing additional
336 evidence for contig connection.

337

338 IS26 copy number was estimated by mapping all genomes to the IS26 sequence using
339 Bowtie v2.2.9 (26), and dividing the read depth across IS26 by the average chromosomal
340 read depth. To assess whether IS26 copy number was increasing over time within the MDR
341 lineage, a linear regression analysis was performed using estimated IS26 copy number and
342 year of isolation for each isolate.

343

344 **Analysis of IncI1 and IncC plasmids**

345 All *S. Kentucky* ST198 genomes were screened for plasmid replicons using SRST2
346 v0.2.0 with the version of the PlasmidFinder database (49) that is distributed in the SRST2
347 package.

348

349 Reads from *S. Kentucky* ST198 isolates containing IncI1 plasmids as well as a set of
350 publicly available IncI1 plasmid sequences (**Table S2**) were mapped to the IncI1 plasmid
351 pNF1358 (accession DQ017661). SNVs were called using the same method as described
352 above for chromosomal SNVs. The resulting SNVs were filtered to include only those that
353 were present in core genes (defined as genes present in 100% of the IncI1 plasmid sequences,
354 see **Table S3**). The final alignment consisted of 1,380 SNVs, which was used to create a

355 maximum likelihood tree with RAxML v8.1.16 (32) using a GTR+G model with 100
356 bootstraps.

357

358 Reads from *S. Kentucky* ST198 isolates containing IncC plasmids were typed with
359 SRST2 against the cgMLST IncA/C plasmid database (50) to determine the 28-locus plasmid
360 sequence type (pST) for each plasmid.

361

362 **Results**

363

364 **Phylogenetic analysis of *S. Kentucky* ST198**

365 All 121 *S. Kentucky* ST198 genomes were mapped to the draft reference genome for
366 *S. Kentucky* ST198 strain 201001922 (see **Methods**), and 2,066 SNVs were identified in the
367 core genome. Linear regression of root-to-tip distances against year of isolation indicated
368 strong temporal structure for all isolates, as did date randomisation tests in BEAST (**Figure**
369 **S1, S2**). The alignment of these SNVs and the years of isolation were then used to construct a
370 dated phylogenetic tree using BEAST, which was further overlaid with region of origin to
371 infer routes of geographical spread (see **Methods**). The results (**Figure 1**) indicate that nearly
372 all MDR isolates belong to a single monophyletic clade of *S. Kentucky* ST198, which we
373 estimate emerged around 1989 (95% HPD 1983 - 1993) in Egypt (**Figure 1**). The BEAST
374 analysis estimated the evolutionary rate to be 4.8×10^{-7} substitutions site⁻¹ year⁻¹ (95% HPD
375 5.28×10^{-7} - 3.78×10^{-7} substitutions site⁻¹ year⁻¹; see **Figure S2**). This is equivalent to a mean
376 rate of 1.6 SNVs per year, which is similar to rates estimated for other nontyphoidal
377 *Salmonella* serotypes including Typhimurium and Agona (51-53), and faster than those
378 estimated for typhoidal serotypes Typhi and Paratyphi A (54-56).

379

380 The MDR clade includes all isolates carrying SGI1-K and derived variants, which
381 include all of the CIP^R *S. Kentucky* ST198 (**Figure 1**; more details below). In addition to the
382 SGI, the MDR lineage has accumulated amino acid mutations in the QRDR. The first
383 mutation occurred circa 1992 in *gyrA* codon 83 (TCC to TTC, Ser83Phe) (light purple,
384 **Figure 1**), and was then followed circa 1996 by a mutation in codon 80 of *parC* (AGC to
385 ATC, Ser80Ile) (pink, **Figure 1**). These mutations increased MIC to ciprofloxacin, but CIP^R
386 did not arise until additional mutations in codon 87 of *gyrA*; at least three such mutations
387 were observed in the MDR clade (GAC to GGC, AAC or TAC; Asp87Gly, Asp87Asn,
388 Asp87Tyr) (dark purple shades, **Figure 1**).

389

390 The *parC*-80 and *gyrA*-87 mutations accompanied a dramatic clonal expansion, with
391 the clone spreading from Egypt to other geographical locations (**Figure 1**). Multiple
392 independent transfers of *S. Kentucky* ST198 out of Egypt and Northern Africa are evident,
393 with two clades, carrying either Asp87Tyr (TAC) or Asp87Asn (AAC) mutations in *GyrA*
394 codon 87 emerging circa 2000. The former spread into East Africa, Middle Africa, South
395 Asia, Europe and Western Asia (dark red line, **Figure 1**); the latter spread to South-East Asia,
396 Europe and West Africa (black line, **Figure 1**).

397

398 Interestingly, the ST198 genomes isolated from agricultural sources in the USA
399 (including 98K, isolated from poultry in 1937, see **Table S1**) lack the SGI and *gyrA/parC*
400 mutations (**Figure 1**). Notably, while these strains were isolated contemporaneously with the
401 MDR clade (2003 to 2016) they are only distantly related to it, sharing a most recent common
402 ancestor (MRCA) circa 1925 (95% HPD 1898-1938; **Figure 1**). This finding is consistent
403 with previous work indicating that ST198 isolates from livestock or poultry in the USA

404 belong to a different genomic cluster (198.1) than MDR ST198 isolates from clinical cases
405 (198.2) (7).

406

407 **Long-term persistence in a single patient**

408 Three *S. Kentucky* ST198 isolates were recovered, in consecutive years (2009, 2010
409 and 2011) from the same patient who had been infected in Egypt (dark orange box, **Figure**
410 **1**). These isolates belonged to the MDR lineage and shared an MRCA circa 2005, suggesting
411 persistent colonization of ~6 years duration (**Figure 1**). The 2011 isolate, 201100664,
412 differed the most from the inferred MRCA (30 SNVs; 21 non-synonymous SNVs, 6
413 synonymous SNVs, 3 intergenic SNVs), yielding an estimated in vivo substitution rate of 5
414 SNVs per year, faster than that estimated by BEAST analysis of the whole data set. Many of
415 the non-synonymous mutations were in genes responsible for flagella (n=7) and iron
416 transport (n=2) (**Table S4**), although no motility changes were detected in this isolate. Eleven
417 SNVs separated 201000305 and 09-9322 (8 non-synonymous SNVs, 2 synonymous SNVs, 1
418 intergenic SNV). One of these eleven SNVs was found in another iron transport gene
419 (asmb1_3909, **Table S4**).

420

421 **SGI in *S. Kentucky***

422 The presence of any SGI backbone genes was taken as evidence of SGI integration
423 (**Figure S3**). The data indicate that the SGI has been acquired by *S. Kentucky* ST198 on three
424 distinct occasions, integrating each time site-specifically in the 3'-end of the *trmE* gene. SGI2
425 (previously SGI1-J), which carries the multidrug resistance region in a different position of
426 the SGI1 backbone (**Figure 2a**) was present in a single isolate from Indonesia, and SGI1-B
427 was present in a single isolate from India; both these isolates were distantly related to the
428 main MDR lineage (**Figure 1**). The vast majority (95%) of genomes belonging to the main
429 MDR lineage carried the SGI1-K subtype or one of its derivatives (SGI1-P or SGI1-Q),
430 consistent with acquisition of SGI1-K in the MRCA circa 1989 in Egypt, shortly before the
431 expansion of the clone (**Figure 1**). Within this MDR lineage, some isolates had large
432 deletions of the SGI backbone (eg: deletions spanning from *S011* to *S026*, or from *int* to
433 *S026*), but still retained the multidrug resistance region between *trmE* and *yidY* (**Figure S3**,
434 **Figure S4**).

435

436 Almost every SGI1-positive *S. Kentucky* ST198 isolate in this study had a distinct
437 SGI structure (**Figure 2b**, **Figure S4**). In addition to large deletions of the SGI backbone,
438 some isolates had inversions of whole or part of the resistance gene segment of the island,
439 with various deletions and rearrangements of the transposons (**Figure 2b**). There were
440 multiple different IS26 insertion sites within the resistance elements of the island, providing
441 evidence that IS26 has mediated the majority of differences found in the resistance region of
442 the island (**Figure 2b**). We found that IS26 was rarely present in *S. Kentucky* ST198 isolates
443 outside of the MDR lineage (**Figure S3**). Within the MDR lineage, linear regression analysis
444 of IS26 copy number against year of isolation showed some evidence of IS26 accumulation
445 over time (0.12 IS26 copies per year, $p=0.01$, $R^2=0.05$) (**Figure S5**).

446 There was no relationship between degraded SGI1s and geographic region or country,
447 or between the loss of core SGI resistance genes (defined as *aacA5*, *bla*_{TEM-1}, *sul1* and *tetA*)
448 and region (see **Methods**). We found that *strAB*, *aphA2*, *aph3-Ia*, *catA1*, *dfrA12* and
449 *mph(A)* were present significantly more frequently in Egypt compared to all other regions
450 (**Table S5**).

451

452 **Multidrug resistance genes and plasmids in *S. Kentucky* ST198**

453 Overall, we found that 35 isolates in the full strain set carried at least one plasmid,
454 covering 13 different known plasmid incompatibility types (**Table S1**). Within the MDR
455 lineage, there was extensive phenotypic and genotypic variation in antimicrobial
456 susceptibility observed (**Figure 3**). A part of this variability could be attributed to the
457 acquisition of plasmids carrying additional AMR genes, as 32 isolates in the MDR lineage
458 carried genes outside the SGI that are likely plasmid-borne (**Figure 3e**). Known plasmid
459 replicons were identified in 23 isolates, and in total we identified eight different plasmid
460 incompatibility types across the MDR strain set (C, II, L/M, Q1, W, X1, X4, Y). From these
461 23 isolates carrying known plasmid incompatibility types, we were able to determine precise
462 plasmid-AMR gene links for 20 isolates.

463
464 There appeared to be no link between geography and plasmid type, with plasmids
465 present in isolates from multiple different regions (**Figure S6**). The majority of genes
466 encoding carbapenemases (*bla_{OXA-48}* and *bla_{NDM-1}*), ESBLs (*bla_{CTX-M-1}*) and cephamycinases
467 (*bla_{CMY-2}*, *bla_{CMY-4}* and *bla_{CMY-16}*) were carried by either IncII or IncC (previously IncA/C₂)
468 plasmids (**Figure 3d, 3e**). Two IncL/M plasmids were found to carry *bla_{OXA-48}* or *bla_{CTX-M-15}*,
469 and an IncW plasmid was found to carry *bla_{VIM-2}* (**Figure 3d, 3e**). The eight isolates resistant
470 to azithromycin contained the *mph(A)* gene. These isolates clustered into two groups. A
471 plasmid location of *mph(A)* was found for four isolates. Three different Inc types were
472 identified (IncII, IncC, and IncL/M).

473
474 There was little evidence that any plasmids were being maintained as the MDR
475 lineage evolved (**Figure 3**), although the group of three isolates recovered from the same
476 patient in Egypt (09-9322, 201000305, 201100664; discussed above) all carried IncII
477 plasmids. These three plasmids were identical in their core gene content, although IncII
478 plasmids in 201100664 differed from those in the earlier two isolates by two intergenic SNVs
479 (**Figure S7**). Interestingly these three isolates all lacked the SGI and any other chromosomal
480 resistance genes, and their IncII plasmids differed substantially from one another in
481 resistance gene content (**Figure 3e**). The two early isolates mostly carried resistance genes
482 for aminoglycosides, sulfonamides, trimethoprim, phenicols and macrolides. The plasmid in
483 the final isolate, 201100664, had lost almost all of the resistance genes found in the previous
484 two isolates, except for *mph(A)*, and had gained the carbapenemase-encoding *bla_{OXA-48}* gene.
485 IncII plasmids were detected in a further six *S. Kentucky* ST198 genomes, but these did not
486 cluster in either the IncII plasmid tree or the chromosome tree, consistent with seven distinct
487 introductions of IncII plasmids into the *S. Kentucky* ST198 MDR lineage, each associated
488 with distinct AMR gene contents (**Figure 3, Figure S7**).

489
490 Two isolates of the MDR lineage carried IncC plasmids (99-2998 and 201410673).
491 Both IncC plasmids were genotyped as pST3, which is commonly associated with *bla_{CMY}*
492 (50), and this cephamycinase-encoding gene was found in the plasmid from isolate 99-2998.
493 Interestingly, the IncC plasmid in isolate 201401673 was carrying a carbapenemase-encoding
494 *bla_{NDM-1}* gene, which is more commonly found in pST1 IncC plasmids (50). This *bla_{NDM-1}*
495 gene was found in a different structural context to the *bla_{NDM}* genes in the pST1 IncC
496 plasmids; as usual it was downstream of *ISAbal25*, however instead of being upstream of
497 *ble*, instead it was upstream of *qacEAI* and *sull1*, with a remnant of the *ble* gene left behind
498 from the insertion of *qacEAI* (**Figure S8**). We found that this *bla_{NDM-1}* region was entirely
499 covered by WGS reads, with no breaks or gaps in coverage, supporting that it is the true
500 structure in this plasmid (**Figure S8**). This configuration also appears in another pST3 IncC
501 plasmid, pRH-1238, from *S. enterica* serotype Corvallis (GenBank accession KR091911),
502 isolated from a wild bird in Germany (57).

503

504 Another source for the phenotypic diversity of *S. Kentucky* ST198 susceptibility
505 profiles was variations in the SGI1 (**Figure 3d**). Notably, plasmid carriage was significantly
506 associated in the cases where SGI1-P, SGI1-Q (containing few or no AMR genes), or no SGI
507 were detected (Fisher's Exact Test, $p=0.024$, OR=2.65 95% CI = 1.09 - 6.64) (**Figures 3b,**
508 **3d, 3e**).

509

510 **Chromosomal gene content diversity amongst *S. Kentucky* ST198 isolates**

511 There was very little gene content diversity evident amongst the *S. Kentucky* ST198
512 chromosome sequences (**Figure S9**). Three phages were detected within the reference
513 genome 201001922 and these three phage regions, in addition to the SGI1, were the only
514 regions to show large differences between genomes from the MDR lineage and those from
515 other lineages (**Figure S9**). Supporting this, within the accessory gene content identified
516 using Roary (see **Methods**), only four genes were found to be present exclusively in all but
517 one of the MDR lineage genomes. All four of these genes were located within a single phage,
518 ST160 (43 kbp, 46 genes, positions 541864 - 584944 in the 201001922 reference genome).
519 This phage was found to be inserted between *ompP* and *mfaA* in the MDR lineage. A
520 variation of this phage was also present in the oldest genome, 98K, which is outside the MDR
521 lineage, however in this genome the phage was inserted between *napB* and *hutI*.

522

523 Examination of the virulence gene content in all isolates revealed that there was no
524 difference between *S. Kentucky* ST198 isolates belonging to the MDR lineage and those
525 belonging to other lineages (**Figure S10**). Only five virulence genes were present in less than
526 95% of genomes – *gogB* (0.8%), *sipB* (7%), *sipC* (35%), *ompD* (57%) and *sciQ* (80%)
527 (**Table S6**) – however these were randomly distributed in the tree and not associated with
528 lineage (**Figure S10**).

529

530 **Discussion**

531 Our data show that nearly all MDR *S. Kentucky* ST198 belong to a single lineage that
532 has accumulated AMR determinants since the early 1990s (**Figure 1**). It first acquired a
533 variant of the SGI1, SGI1-K which conferred resistance to ampicillin, streptomycin,
534 gentamicin, sulfamethoxazole, and tetracycline (**Figure 2**). The SGI1 structure appears to be
535 highly susceptible to genetic rearrangements, with distinct forms found in each isolate likely
536 due to the transpositional activity of IS26, which resulted in deletion of some or all genes
537 inside SGI1. The loss of resistance genes was often made up for by acquisition of additional
538 MDR plasmids (**Figure 3**).

539

540 IS26 is 820 bp long and encodes a single transposase with 14 bp terminal repeats on
541 each end (58). Each of the three SGI1 subtypes found in the MDR lineage carried one or
542 more copies of IS26, and all genomes in the MDR lineage carried IS26, with no genomes
543 outside of this clade carrying IS26. The recently described mechanism used by IS26 to
544 transpose may provide an explanation as to why the SGI variants in these isolates are so
545 dynamic. During the transposition, IS26 extracts itself from the donor DNA molecule, as well
546 as DNA lying upstream of it between itself and another IS26 element, and uses this to form a
547 translocatable unit (59). It then finds another IS26 element in the receiving DNA molecule,
548 and inserts itself as well as the excised donor DNA next to it, forming a tandem array of
549 IS26s in direct orientation (59). This model illustrates that IS26 is likely the causative agent
550 for many of the deletions, inversions and transpositions within the SGI, eventually resulting
551 in the genesis of the different SGI1 variants (SGI1-K, SGI1-P and SGI1-Q) seen in this
552 dataset (**Figure 2**).

553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590

Whilst the origin of the MDR clade appears to be intimately linked with the acquisition of the SGI1 in Egypt, it is the QRDR triple-mutant CIP^R subclade that disseminated globally (**Figure 1**). Ciprofloxacin resistance is infrequent in *Salmonella* (64), and we hypothesise that this high-level resistance is linked to strong selective pressure exerted by fluoroquinolone use in poultry, *S. Kentucky*'s main reservoir (65). This resistance might also have come at no cost to the fitness of the bacterial cell, as has been shown in close relatives *S. Typhi* and *Escherichia coli* (60,61).

During its spread around the world, the *S. Kentucky* ST198 MDR lineage became more resistant by the additional acquisition of various AMR plasmids, carrying genes encoding resistance to newer drugs including third-generation cephalosporins, carbapenems and azithromycin. These genes were acquired locally around the Mediterranean basin with no subsequent clonal expansion. Interestingly, the two isolates containing IncC plasmids did not carry the SGI. This observation is supported by many studies in the literature which have described the incompatibility of the SGI and IncC plasmids, as they share the same regulatory system (11,62,63).

In this study we were unable to detect any other non-AMR related genes that could explain the clonal success of the MDR lineage. Examination of phage, pseudogenes and known virulence genes did not reveal any significant differences between the MDR lineage and other *S. Kentucky* ST198 genomes, although this does not rule out the possibility of more subtle variants contributing to virulence such as the regulatory SNV recently described for invasive *S. Typhimurium* ST313 (70).

In conclusion, WGS analysis of *S. Kentucky* ST198 has significantly expanded our knowledge of the evolution and dissemination of MDR variants of this important pathogen. Previously, as this lineage was emerging, MLST and PFGE were used in combination (1,2) for this purpose; however the diversity of PFGE types of CIP^R *S. Kentucky* ST198 isolates precluded any fine-scale or long-term analysis of *S. Kentucky* ST198 dissemination, which our data shows was likely due to noise introduced by IS26 activity. The population structure uncovered here should serve as a useful framework with which to understand and track the ongoing evolution of the MDR lineage of *S. Kentucky* ST198, which our data clarifies is a globally disseminated clone capable of rapid spread and further accumulation of last-line AMR determinants.

591 **Acknowledgments**

592

593 We thank J. Morad-Gilan (Israel), S. Bertrand (Belgium), J. Beutlich and W. Rabsch
594 (Germany), D. Wasyl (Poland), A. Baket (Egypt), B. Bouchrif (Morocco), H. Barua
595 (Bangladesh), V. Sintchenko (Australia), Noelia Antilles (Spain), Marta Cerda-Cuellar
596 (Spain) and M.S. Skovgaard (Denmark) for their support and for providing isolates; L. Fabre,
597 L. Ma, V. Enouf and C. Bouchier for sequencing the isolates. We also thank all
598 corresponding laboratories of the French National Reference Centre for *E. coli*, *Shigella*, and
599 *Salmonella*.

600

601 **Funding information**

602

603 The French National Reference Centre for *E. coli*, *Shigella*, and *Salmonella* is funded by the
604 Institut Pasteur and *Santé Publique France*. The “Unité des Bactéries Pathogènes Entériques”
605 belongs to the "Integrative Biology of Emerging Infectious Diseases" Laboratory of
606 Excellence funded by the French Government “Investissement d'Avenir” programme (grant
607 no. ANR-10-LABX-62-IBEID). C.G. was supported by a grant from Assistance Publique-
608 Hôpitaux de Paris. KEH is supported by a Senior Medical Research Fellowship from the
609 Viertel Foundation of Australia; and the Bill and Melinda Gates Foundation, Seattle, USA.

610

611 **Author Statements**

612

613

614 S.L.H. and F.-X. designed the study. S.L.H., B.D., S.G., R.H. P.-J.C. and F.-X.W. collected,
615 selected and provided characterised isolates or their genomes and their corresponding
616 epidemiological information. F.F. performed the draft genome sequencing. C.G. performed
617 phenotypic experiments. S.L.H. analysed phenotypic experiments. J.H., K.E.H. and F.-X.W.
618 analysed the genomic sequence data. J.H. wrote the manuscript, with major contributions
619 from S.L.H., K.E.H. and F.-X.W. All authors contributed to the editing of the manuscript.
620 H.B.-J. and K.E.H. supervised J.H.'s PhD.

621

622 **Conflicts of Interest**

623

624 The authors declare no conflicts of interest.

625

626 **Abbreviations**

627

628 AMR: Antimicrobial resistance

629 AST: Antimicrobial susceptibility testing

630 BEAST: Bayesian Evolutionary Analysis Sampling Trees

631 CA-SFM: Antibiogram Committee of the French Society for Microbiology

632 CIP: Ciprofloxacin

633 CIP^R: Ciprofloxacin-resistant

634 ENA: European Nucleotide Archive

635 ESBL: extended-spectrum beta-lactamase

636 IS: Insertion sequences

637 MDR: Multidrug-resistant

638 MLST: Multi-locus sequence typing

639 MIC: Minimal inhibitory concentration

640 PCR: Polymerase chain reaction

641 PFGE: Pulsed-field gel electrophoresis
642 QRDR: Quinolone resistance-determining region
643 SGI: *Salmonella* Genomic Island
644 SNV: Single-nucleotide variant
645 WHO: World Health Organisation
646 WGS: Whole-genome sequencing

647
648

649 References

650
651

652 1. Weill F-X, Bertrand S, Guesnier F, Baucheron S, Cloeckeaert A, Grimont PAD.
653 Ciprofloxacin-resistant *Salmonella* Kentucky in travelers. *Emerg Infect Dis.*
654 2006 Oct;12(10):1611–2.

655 2. Le Hello S, Hendriksen RS, Doublet B, Fisher I, Nielsen EM, Whichard JM, et
656 al. International spread of an epidemic population of *Salmonella enterica*
657 serotype Kentucky ST198 resistant to ciprofloxacin. *J Infect Dis.* Oxford
658 University Press; 2011 Sep 1;204(5):675–84.

659 3. Le Hello S, Bekhit A, Granier SA, Barua H, Beutlich J, Zając M, et al. The
660 global establishment of a highly-fluoroquinolone resistant *Salmonella enterica*
661 serotype Kentucky ST198 strain. *Front Microbiol.* Frontiers; 2013;4:395.

662 4. Le Hello S, Harrois D, Bouchrif B, Sontag L, Elhani D, Guibert V, et al. Highly
663 drug-resistant *Salmonella enterica* serotype Kentucky ST198-X1: a
664 microbiological study. *Lancet Infect Dis.* Elsevier; 2013 Aug 1;13(8):672–9.

665 5. Westrell T, Monnet DL, Gossner C, Heuer O, Takkinen J. Drug-resistant
666 *Salmonella enterica* serotype Kentucky in Europe. *Lancet Infect Dis.* Elsevier;
667 2014 Apr;14(4):270–1.

668 6. Le Hello S, Weill F-X, Guibert V, Praud K, Cloeckeaert A, Doublet B. Early
669 strains of multidrug-resistant *Salmonella enterica* serovar Kentucky sequence
670 type 198 from Southeast Asia harbor *Salmonella* genomic island 1-J variants
671 with a novel insertion sequence. *Antimicrob Agents Chemother.* American
672 Society for Microbiology; 2012 Oct;56(10):5096–102.

673 7. Boyd D, Peters GA, Cloeckeaert A, Boumedine KS, Chaslus-Dancla E,
674 Imberechts H, et al. Complete nucleotide sequence of a 43-kilobase genomic
675 island associated with the multidrug resistance region of *Salmonella enterica*
676 serovar Typhimurium DT104 and its identification in phage type DT120 and
677 serovar Agona. *J Bacteriol.* American Society for Microbiology; 2001
678 Oct;183(19):5725–32.

679 8. Boyd D, Cloeckeaert A, Chaslus-Dancla E, Mulvey MR. Characterization of
680 variant *Salmonella* genomic island 1 multidrug resistance regions from serovars
681 Typhimurium DT104 and Agona. *Antimicrobial Agents and Chemotherapy.*
682 2002;46(6):1714–22.

- 683 9. Siebor E, Neuwirth C. Emergence of *Salmonella* genomic island 1 (SGI1) among
684 *Proteus mirabilis* clinical isolates in Dijon, France. *J Antimicrob Chemother.*
685 Oxford University Press; 2013 Aug;68(8):1750–6.
- 686 10. Doublet B, Boyd D, Mulvey MR, Cloeckaert A. The *Salmonella* genomic island
687 1 is an integrative mobilizable element. *Mol Microbiol.* 2005 Mar;55(6):1911–
688 24.
- 689 11. Carraro N, Matteau D, Luo P, Rodrigue S, Burrus V. The master activator of
690 IncA/C conjugative plasmids stimulates genomic islands and multidrug
691 resistance dissemination. Hughes D, editor. *PLoS Genet.* Public Library of
692 Science; 2014 Oct 23;10(10):e1004714.
- 693 12. Douard G, Praud K, Cloeckaert A, Doublet B. The *Salmonella* Genomic Island 1
694 is specifically mobilized in trans by the IncA/C multidrug resistance plasmid
695 family. Xu S-Y, editor. *PLoS ONE.* Public Library of Science; 2010 Dec
696 20;5(12):e15302.
- 697 13. Hall RM. *Salmonella* genomic islands and antibiotic resistance in *Salmonella*
698 *enterica*. *Future Microbiol.* 2010 Oct;5(10):1525–38.
- 699 14. Siebor E, Neuwirth C. *Proteus* genomic island 1 (PGI1), a new resistance
700 genomic island from two *Proteus mirabilis* French clinical isolates. *J Antimicrob*
701 *Chemother.* Oxford University Press; 2014 Aug 11;:dku314.
- 702 15. Levings RS, Djordjevic SP, Hall RM. SGI2, a relative of *Salmonella* genomic
703 island SGI1 with an independent origin. *Antimicrob Agents Chemother.*
704 American Society for Microbiology; 2008 Jul;52(7):2529–37.
- 705 16. Doublet B, Praud K, Weill F-X, Cloeckaert A. Association of IS26-composite
706 transposons and complex In4-type integrons generates novel multidrug
707 resistance loci in *Salmonella* genomic island 1. *J Antimicrob Chemother.* Oxford
708 University Press; 2009 Feb;63(2):282–9.
- 709 17. Doublet B, Praud K, Bertrand S, Collard JM, Weill FX, Cloeckaert A.
710 52(10):3745–54.
- 711 18. Levings RS, Partridge SR, Djordjevic SP, Hall RM. SGI1-K, a variant of the
712 SGI1 genomic island carrying a mercury resistance region, in *Salmonella*
713 *enterica* serovar Kentucky. *Antimicrob Agents Chemother.* American Society
714 for Microbiology; 2007 Jan;51(1):317–23.
- 715 19. Baucheron S, Le Hello S, Doublet B, Giraud E, Weill F-X, Cloeckaert A. *ramR*
716 mutations affecting fluoroquinolone susceptibility in epidemic multidrug-
717 resistant *Salmonella enterica* serovar Kentucky ST198. *Front Microbiol.* 2013
718 Sep 7;4:1–6.
- 719 20. Allard MW, Strain E, Melka D, Bunning K, Musser SM, Brown EW, et al.
720 Practical Value of Food Pathogen Traceability through Building a Whole-
721 Genome Sequencing Network and Database. Kraft CS, editor. *J Clin Microbiol.*
722 American Society for Microbiology Journals; 2016 Aug 1;54(8):1975–83.

- 723 21. Timme RE, Rand H, Leon MS, Hoffmann M, Strain E, Allard M, et al.
724 GenomeTrakr proficiency testing for foodborne pathogen surveillance: an
725 exercise from 2015. *Microbial Genomics*. Microbiology Society; 2018 Jul
726 1;4(7):289.
- 727 22. Inouye M, Dashnow H, Raven L-A, Schultz MB, Pope BJ, Tomita T, et al.
728 SRST2: Rapid genomic surveillance for public health and hospital microbiology
729 labs. *Genome Med*. 2014;6(11):90.
- 730 23. Kurtz S, Phillippy A, Delcher AL. Versatile and open software for comparing
731 large genomes. *Genome Biology*. 2004;5(R12).
- 732 24. Holt KE, Parkhill J, Mazzoni CJ, Roumagnac P, Weill F-X, Goodhead I, et al.
733 High-throughput sequencing provides insights into genome variation and
734 evolution in *Salmonella* Typhi. *Nat Genet*. Nature Publishing Group; 2008 Aug
735 1;40(8):987–93.
- 736 25. Holt KE, Baker S, Weill F-X, Holmes EC, Kitchen A, Yu J, et al. *Shigella*
737 *sonnei* genome sequencing and phylogenetic analysis indicate recent
738 global dissemination from Europe. *Nat Genet*. Nature Publishing Group; 2012
739 Sep;44(9):1056–9.
- 740 26. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat*
741 *Meth*. 2012 Mar 4;9(4):357–9.
- 742 27. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The
743 Sequence Alignment/Map format and SAMtools. *Bioinformatics*. Oxford
744 University Press; 2009 Aug 15;25(16):2078–9.
- 745 28. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al.
746 Rapid phylogenetic analysis of large samples of recombinant bacterial whole
747 genome sequences using Gubbins. *Nucleic Acids Res*. Oxford University Press;
748 2014 Nov 20;:gku1196.
- 749 29. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, et al. BEAST 2:
750 a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol*.
751 2014 Apr;10(4):e1003537.
- 752 30. Rambaut A, Suchard MA, Xie D, Drummond AJ. Tracer v1.6 [Internet]. 2014.
753 Available from: <http://beast.bio.ed.ac.uk>
- 754 31. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by
755 sampling trees. *BMC Evol Biol*. BioMed Central Ltd; 2007;7(1):214.
- 756 32. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-
757 analysis of large phylogenies. *Bioinformatics*. Oxford University Press; 2014
758 May 1;30(9):1312–3.
- 759 33. United Nations Statistics Division. Composition of macro geographical
760 (continental) regions, geographical sub-regions, and selected economic and other
761 groupings [Internet]. 2013 [cited 2016 Aug 10]. Available from:
762 <http://unstats.un.org/unsd/methods/m49/m49regin.htm>

- 763 34. Bollback JP. SIMMAP: stochastic character mapping of discrete traits on
764 phylogenies. *BMC Bioinformatics*. BioMed Central Ltd; 2006;7(1):88.
- 765 35. Revell LJ. phytools: an R package for phylogenetic comparative biology (and
766 other things). *Methods in Ecology and Evolution*. Blackwell Publishing Ltd;
767 2011 Dec 15;3(2):217–23.
- 768 36. Hannon Lab. FastXToolkit [Internet]. 2010 [cited 2016 Aug 10]. Available from:
769 http://hannonlab.cshl.edu/fastx_toolkit/
- 770 37. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina
771 Sequence Data. *Bioinformatics*. Oxford University Press; 2014 Apr 1;30(15).
- 772 38. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al.
773 SPAdes: a new genome assembly algorithm and its applications to single-cell
774 sequencing. *J Comput Biol*. 2012 May;19(5):455–77.
- 775 39. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-
776 assembled contigs using SSPACE. *Bioinformatics*. Oxford University Press;
777 2011 Feb 15;27(4):578–9.
- 778 40. Boetzer M, Pirovano W. Toward almost closed genomes with GapFiller.
779 *Genome Biology*. BioMed Central; 2012 Jun 25;13(6):1.
- 780 41. Assefa S, Keane TM, Otto TD, Newbold C, Berriman M. ABACAS: algorithm-
781 based automatic contiguation of assembled sequences. *Bioinformatics*. 2009 Aug
782 1;25(15):1968–9.
- 783 42. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*.
784 Oxford University Press; 2014 Mar 18;30(14).
- 785 43. Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud
786 L, et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance
787 genes in bacterial genomes. *Antimicrob Agents Chemother*. American Society
788 for Microbiology; 2014;58(1):212–20.
- 789 44. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al.
790 Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*. 2015
791 Nov 15;31(22):3691–3.
- 792 45. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al.
793 BLAST+: architecture and applications. *BMC Bioinformatics*. BioMed Central
794 Ltd; 2009 Dec 15;10(1):421.
- 795 46. Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: hierarchical and refined
796 dataset for big data analysis - 10 years on. *Nucleic Acids Res*. Oxford University
797 Press; 2016 Jan 4;44(D1):D694–7.
- 798 47. Hawkey J, Hamidian M, Wick RR, Edwards DJ, Billman-Jacobe H, Hall RM, et
799 al. ISMapper: identifying transposase insertion sites in bacterial genomes from
800 short read sequence data. *BMC Genomics*. BioMed Central Ltd; 2015;16(1):667.

- 801 48. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of
802 de novo genome assemblies. *Bioinformatics*. 2015 Oct 15;31(20):3350–2.
- 803 49. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L,
804 et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid
805 multilocus sequence typing. *Antimicrob Agents Chemother*. American Society
806 for Microbiology; 2014 Jul;58(7):3895–903.
- 807 50. Hancock SJ, Phan MD, Peters KM, Forde BM, Chong TM, Yin W-F, et al.
808 Identification of IncA/C Plasmid Replication and Maintenance Genes and
809 Development of a Plasmid Multilocus Sequence Typing Scheme. *Antimicrob*
810 *Agents Chemother*. 2017 Feb;61(2):AAC.01740–16.
- 811 51. Leekitcharoenphon P, Nielsen EM, Kaas RS, Lund O, Aarestrup FM. Evaluation
812 of whole genome sequencing for outbreak detection of *Salmonella enterica*.
813 *PLoS ONE*. 2014;9(2):e87991.
- 814 52. Okoro CK, Kingsley RA, Connor TR, Harris SR, Parry CM, Al-Mashhadani
815 MN, et al. Intracontinental spread of human invasive *Salmonella* Typhimurium
816 pathovariants in sub-Saharan Africa. *Nature Publishing Group*. *Nature*
817 *Publishing Group*; 2012 Sep 30;44(11):1215–21.
- 818 53. Zhou Z, McCann A, Litrup E, Murphy R, Cormican M, Fanning S, et al. Neutral
819 Genomic Microevolution of a Recently Emerged Pathogen, *Salmonella enterica*
820 Serovar Agona. Casadesús J, editor. *PLoS Genet*. Public Library of Science;
821 2013 Apr 18;9(4):e1003471.
- 822 54. Wong VK, Baker S, Pickard DJ, Parkhill J, Page AJ, Feasey NA, et al.
823 Phylogeographical analysis of the dominant multidrug-resistant H58 clade of
824 *Salmonella* Typhi identifies inter- and intracontinental transmission events. *Nat*
825 *Genet*. 2015 Jun;47(6):632–9.
- 826 55. Zhou Z, McCann A, Weill F-X, Blin C, Nair S, Wain J, et al. Transient
827 Darwinian selection in *Salmonella enterica* serovar Paratyphi A during 450 years
828 of global spread of enteric fever. *Proc Natl Acad Sci USA*. *National Acad*
829 *Sciences*; 2014 Aug 19;111(33):12199–204.
- 830 56. Duchêne S, Holt KE, Weill F-X, Le Hello S, Hawkey J, Edwards DJ, et al.
831 Genome-scale rates of evolutionary change in bacteria. *Microbial Genomics*.
832 *Microbiology Society*; 2016 Nov 30;2(11).
- 833 57. Villa L, Guerra B, Schmogger S, Fischer J, Helmuth R, Zong Z, et al. IncA/C
834 Plasmid Carrying bla(NDM-1), bla(CMY-16), and fosA3 in a *Salmonella*
835 *enterica* Serovar Corvallis Strain Isolated from a Migratory Wild Bird in... -
836 PubMed - NCBI. *Antimicrob Agents Chemother*. 2015 Sep 18;59(10):6597–600.
- 837 58. Mollet B, Iida S, Shepherd J, Arber W. Nucleotide sequence of IS26, a new
838 prokaryotic mobile genetic element. *Nucleic Acids Res*. 1983 Sep
839 24;11(18):6319–30.

- 840 59. Harmer CJ, Moran RA, Hall RM. Movement of IS26-associated antibiotic
841 resistance genes occurs via a translocatable unit that includes a single IS26 and
842 preferentially inserts adjacent to another IS26. *mBio*. 2014;5(5):e01801–14.
- 843 60. Baker S, Duy PT, Nga TVT, Dung TTN, Phat VV, Chau TT, et al. Fitness
844 benefits in fluoroquinolone-resistant *Salmonella* Typhi in the absence of
845 antimicrobial pressure. *eLife*. 2013 Dec 10;2:452.
- 846 61. Webber MA, Ricci V, Whitehead R, Patel M, Fookes M, Ivens A, et al.
847 Clinically Relevant Mutant DNA Gyrase Alters Supercoiling, Changes the
848 Transcriptome, and Confers Multidrug Resistance. *mBio*. American Society for
849 Microbiology; 2013 Aug 30;4(4):e00273–13–e00273–13.
- 850 62. Harmer CJ, Hamidian M, Ambrose SJ, Hall RM. Destabilization of IncA and
851 IncC plasmids by SGI1 and SGI2 type *Salmonella* genomic islands. *Plasmid*
852 [Internet]. 2016 Nov;87-88:51–7.
- 853 63. Huguet KT, Gonnet M, Doublet B, Cloeckeaert A. A toxin antitoxin system
854 promotes the maintenance of the IncA/C-mobilizable *Salmonella* Genomic
855 Island 1. *Scientific Reports*. 2016;6(1):32285.

856 **Figures and Tables**

857 **Figure 1: Phylogeographic analysis of *S. Kentucky* ST198 based on whole genome SNV**
858 **data.** Bayesian maximum clade credibility tree inferred using BEAST, with MDR lineage
859 shaded orange. Dark orange box indicates three isolates from the same patient. Major internal
860 nodes are labeled with circles indicating branch support (black, $\geq 95\%$ posterior support; red,
861 $>70\%$ posterior support; hollow, $>30\%$ posterior support); divergence date estimates (95%
862 higher posterior density values) are provided for key points in the evolution of the MDR
863 lineage. Leaf nodes are coloured by region of origin (see inset map). Coloured branches
864 indicate inferred geographical distribution of internal branches, inferred using maximum
865 likelihood ancestral trait reconstruction. Data columns indicate country of origin; source of
866 isolate (H for human, N for non-human, ? for unknown); SGI type (see inset legend);
867 quinolone resistance-related codons, with resistance-associated alleles highlighted. Reference
868 genome 201001922 is marked with red arrow. Red and black vertical lines indicate clades
869 that are mentioned in-text.

870
871

872 **Figure 2: SGI variation in *S. Kentucky* ST198. a,** Backbone of SGI, with arrows pointing
873 to the different insertion sites of the resistance region in SGI1 and SGI2. **b,** Different
874 examples of SGI1 types in *S. Kentucky* ST198. Arrows show open reading frames (ORF) of
875 the SGI backbone and MDR region with arrowheads indicating direction of transcription;
876 colour indicates gene class. Coloured blocks indicate regions of homology between
877 sequences in the same orientation; green, same orientation; orange, inverse orientation.

878
879

880 **Figure 3: Horizontally acquired antimicrobial resistance genes in the *S. Kentucky***
881 **ST198 MDR lineage. a,** Dated Bayesian (BEAST) phylogeny for the MDR lineage,
882 extracted from the tree shown in **Figure 1**. Leaf nodes are coloured by region of origin (see
883 legend); orange box highlights three isolates recovered from the same patient over three
884 years. **b-e** shows AMR features of each isolate in the tree. **b,** SGI type (see legend, dash

885 indicates no SGI detected). **c**, AMR phenotypes, indicated as boxes coloured by antimicrobial
886 class (see legend, I in box denotes intermediate resistance). **d**, AMR genes located within the
887 SGI1 are indicated with boxes coloured by antimicrobial class (* in box indicates gene is
888 interrupted). **e**, plasmid incompatibility group(s) identified in each genome; antimicrobial
889 resistance genes located within these plasmids are printed, coloured by antimicrobial class;
890 genes in brackets are genes whose location was unable to be determined.

891
892

893 **Supplementary Figure 1: Temporal signal in the maximum likelihood phylogeny.** Grey
894 dots, isolates not within the MDR lineage; red dots, isolates within the MDR lineage. Grey
895 line, linear regression of all isolates in tree; red line, linear regression of only isolates within
896 the MDR lineage. Values reported are the correlation coefficients for these regressions.

897
898

899 **Supplementary Figure 2: Mutation rate estimates for real and randomised tip dates in**
900 ***S. Kentucky ST198*.** First column, real mutation rate, in substitutions per site per year.
901 Subsequent columns show mutation rate when tip dates are randomised. Black circles are the
902 mean rate estimated by BEAST, with error bars showing 95% highest posterior density
903 (HPD).

904
905

906 **Supplementary Figure 3: Presence of SGI backbone genes in each *S. Kentucky ST198***
907 **isolate, with an estimate of overall IS26 copy number.** Left, dated Bayesian phylogeny,
908 with tips and branches coloured by region (as per map inset and **Figure 1**). Red blocks
909 indicate the presence at least one resistance gene in the SGI resistance region. Light blue
910 blocks indicate presence of chromosomal genes flanking the SGI (*trmE* and *gidY*). Dark blue
911 blocks indicate presence of SGI backbone genes. Yellow bars show estimated IS26 copy
912 number in each isolate.

913
914

915 **Supplementary Figure 4: Additional SGI1 examples.** Coding regions are represented as
916 arrows and are coloured as per legend. SGI sequences are grouped by type - SGI1-K, SGI1-P
917 and SGI1-Q. Contig breaks are shown by thick black vertical lines, with IS26 flanking
918 regions detected with ISMapper marked with purple vertical lines.

919
920

921 **Supplementary Figure 5: Relationship between year of isolation of *S. Kentucky ST198***
922 **and estimated IS26 copy number.** Points are coloured by the SGI type found in that strain,
923 as per legend.

924
925

926 **Supplementary Figure 6: Heatmap showing proportion of each plasmid replicon within**
927 **each region found in the *S. Kentucky ST198* MDR lineage.**

928
929

930 **Supplementary Figure 7: Maximum-likelihood phylogeny of IncI1 plasmids.** Phylogeny
931 is midpoint rooted, with tips coloured by genus (as per legend) and *S. Kentucky ST198*
932 isolates from this study coloured pink. Orange box indicates three *S. Kentucky* isolates from
933 a single patient, all carrying IncI1 plasmids. Country of isolation is listed next to the *S.*
934 *Kentucky* isolates used in this study.

935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975

Supplementary Figure 8: Coverage across NDM-1 region in isolate #201410673.

Positions of genes depicted as arrows, with arrows pointing in the direction of transcription. Light blue, other plasmid genes; red, resistance genes; purple, transposable elements. Genes without names are hypothetical proteins. Dark orange box shows boundaries of exact match to the same *bla*_{NDM-1} configuration in *S. enterica* serotype Corvallis (GenBank accession no. KR091911). Depth is number of reads at each base position on the *x*-axis.

Supplementary Figure 9: Comparison of a subset of S. Kentucky ST198 genomes to reference genome 201001922. Inner black ring indicates 201001922 genome position, followed by GC content. Blue rings indicate *S. Kentucky* ST198 isolates belonging to the MDR lineage, grey rings other *S. Kentucky* ST198 isolates. The location of each isolate is shown on the phylogeny - numbers indicate which ring is shown by the isolate.

Supplementary Figure 10: Heatmap of virulence gene content in all S. Kentucky ST198 isolates. Tips of the phylogeny are coloured by region, as per legend. Each column of the heatmap is a virulence gene, black indicates presence, white indicates absence. Virulence genes are clustered to reveal patterns.

Supplementary Table 1: Metadata associated with each S. Kentucky ST198 isolate used in this study.

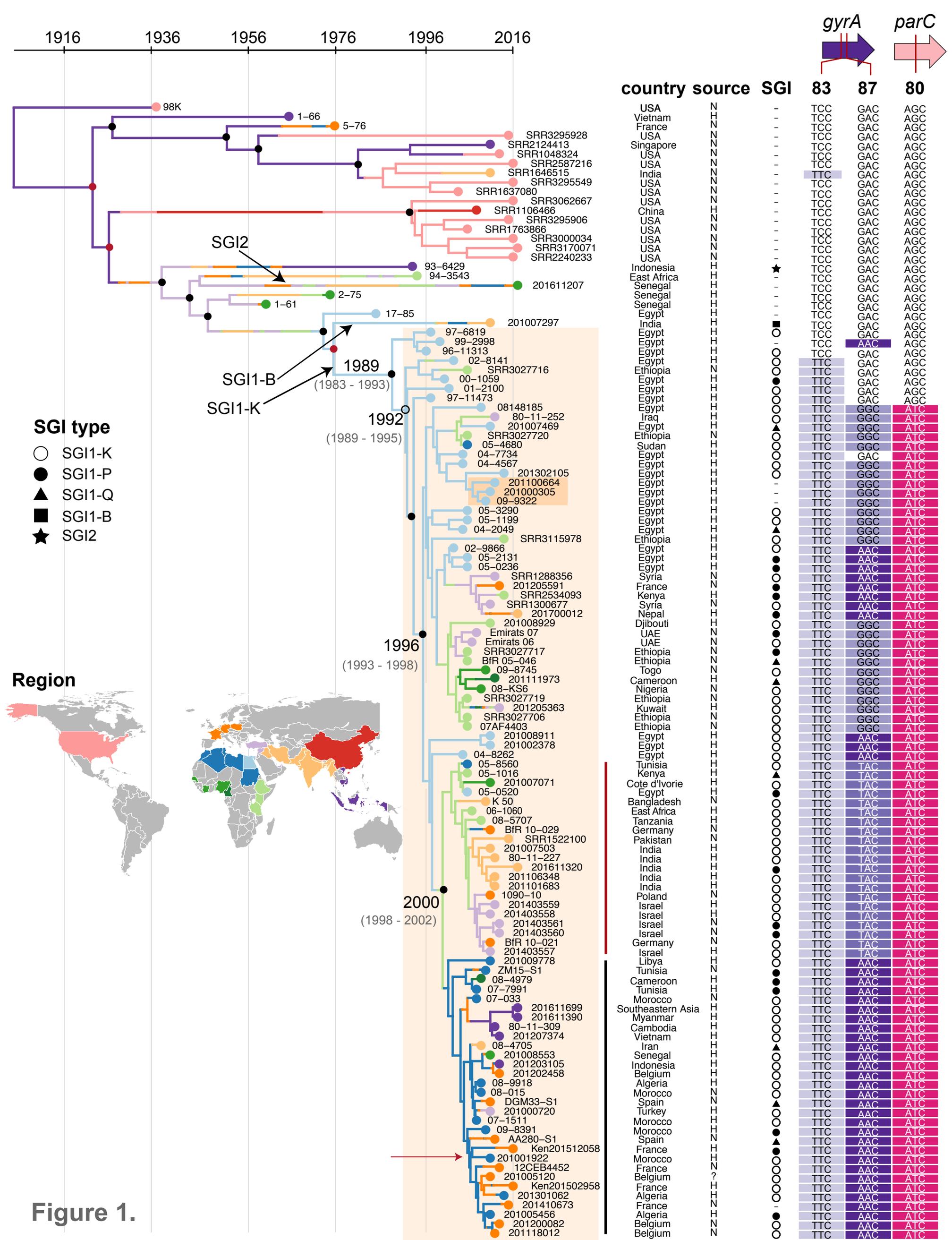
Supplementary Table 2: Plasmid accessions for IncI1 plasmids used in this study.

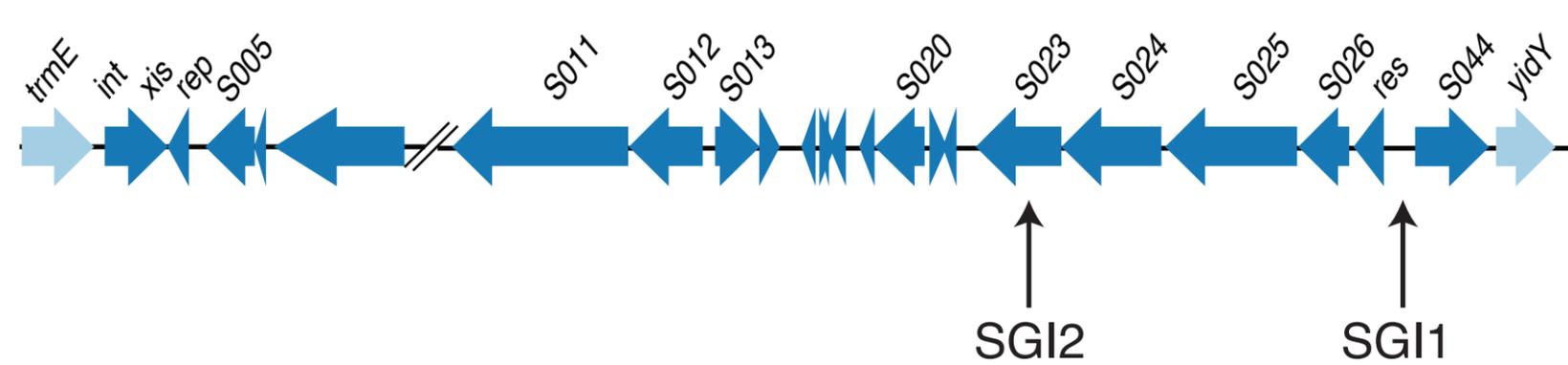
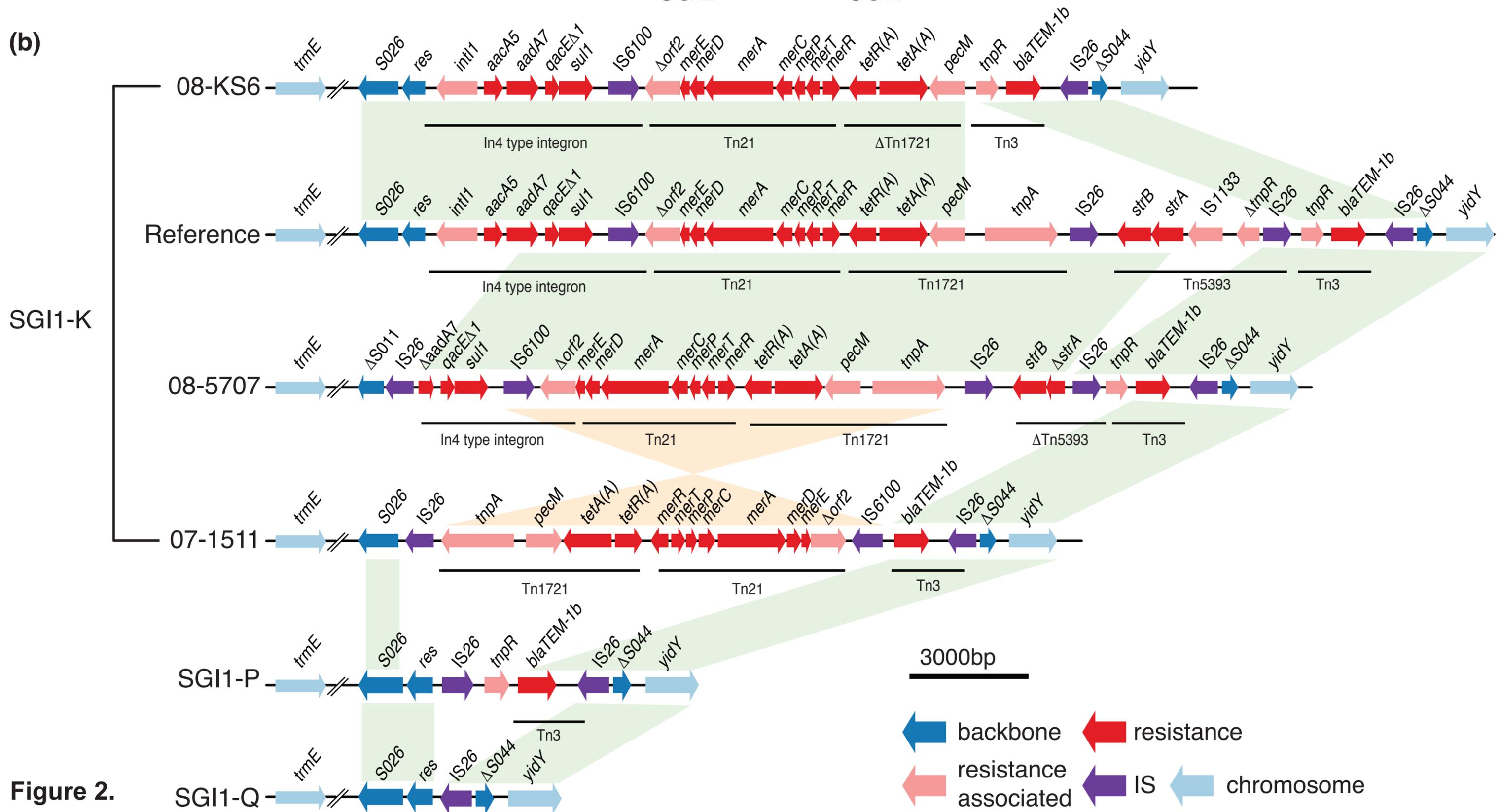
Supplementary Table 3: Genes defined as core in the reference IncI1 plasmid pNF1358 in this study.

Supplementary Table 4: Mutations found amongst the three isolates taken from the same patient.

Supplementary Table 5: P-values (calculated using Fisher's Exact Test) between resistance genes and geographic regions for each resistance gene found across all isolates. P-values < 0.05 indicate a positive association with that resistance gene and that specific geographic region.

Supplementary Table 6: Presence or absence of virulence genes from VFDB in each isolate.



(a)**(b)****Figure 2.**

