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 Ceyssens, Camille Gomart, Helen Billman-Jacobe, Kathryn E. Holt, et al.

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

Jane Hawkey[1,2#‡], Simon Le Hello[3#], Benoît Doublet[4], Sophie Granier[5,6], Rene Hendriksen[7], W. Florian Fricke[8,9] Pieter-Jan Ceyssens[10], Camille Gomart[3], Helen Billman-Jacobe[11], Kathryn E. Holt[1,2,12#], François-Xavier Weill[3#‡]

[1]Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria 3010, Australia.
[2]Department of Infectious Diseases, Central Clinical School, Monash University, Melbourne, Victoria 3004, Australia.
[3]Unité des Bactéries Pathogènes Entériques, Centre National de Référence des *Escherichia coli*, *Shigella* et *Salmonella*, World Health Organisation Collaborative Center for the typing and antibiotic resistance of *Salmonella*, Institut Pasteur, 75015, Paris, France.
[4]ISP, Institut National de la Recherche Agronomique, Université François Rabelais de Tours, UMR 1282, Nouzilly, France.
[5]Université PARIS-EST, Agence Nationale de Sécurité Sanitaire de L'Alimentation, de L'Environnement et du Travail (ANSES), Laboratory for Food Safety, 94701 Maisons-Alfort, France.
[6]Agence Nationale de Sécurité Sanitaire de L'Alimentation, de L'Environnement et du Travail (ANSES), Fougères Laboratory, 35306 Fougères, France
[7]Research Group for Genomic Epidemiology, National Food Institute, Technical University of Denmark, Kongens Lyngby, Denmark.
[8]Department of Microbiome Research and Applied Bioinformatics, University of Hohenheim, Stuttgart, Germany
[9]Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, USA.
[10]Unit Bacterial Diseases, Sciensano, Brussels, Belgium
[11]Asia-Pacific Centre for Animal Health, Faculty of Veterinary and Agricultural Science, University of Melbourne, Parkville, 3010, Australia
[12]London School of Hygiene & Tropical Medicine, London, WC1E 7HT, UK.

[#]Contributed equally

[‡]Corresponding authors:

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

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

Abstract

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

indicates this clone emerged in Egypt before disseminating into Northern, Southern and Western Africa, then to the Middle East, Asia and the European Union. The MDR clone has since accumulated various substitution mutations in the quinolone resistance determining regions (QRDR) of DNA gyrase (*gyrA*) and DNA topoisomerase IV (*parC*), such that most strains carry three QRDR mutations which together confer resistance to ciprofloxacin. The majority of AMR genes in the *S.* Kentucky genomes were carried either on plasmids or SGI structures. Remarkably, each genome of the MDR clone carried a different SGI1-K derivative structure; this variation could be attributed to IS*26*-mediated insertions and deletions, which appear to have hampered previous attempts to trace the clone's evolution using sub-WGS resolution approaches. Several different AMR plasmids were also identified, encoding resistance to chloramphenicol, third-generation cephalosporins, carbapenems, and/or azithromycin. These results indicate that most MDR *S.* Kentucky circulating globally result from the clonal expansion of a single lineage that acquired chromosomal AMR genes 30 years ago, and has continued to diversify and accumulate additional resistances to last-line oral antimicrobials.

Impact Statement

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

Data Summary

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

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

The phylogeny and associated metadata can be viewed on Microreact: https://microreact.org/project/Hkl7CzEXV

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

Introduction

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

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

Four main types of SGI have so far been described in *S.* Kentucky: SGI1-K, SGI1-P, SGI1-Q (see **Figure 2**) and SGI2 (3). These SGI1 variants share a common genetic feature consisting of an insertion/deletion between *S005* and *S009* due to the insertion of IS*1359*, which was also found in a few other SGI1 variants in strains of different *S. enterica* serotypes isolated in 2000 in Egypt, and more recently in *P. mirabilis* (16). Additionally, these three SGI1 variants show a truncation the 5'-end of S044, the final ORF of the SGI backbone, through the insertion of IS*26* (17). SGI1-K contains a complex mosaic resistance region made of different segments of transposons Tn*21*, Tn*1721*, Tn*5393*, Tn*3*-like, and a In4-type

integron structure, as well as IS*26* elements (18). SGI1-P and SGI1-Q contain only the IS*26*-flanked Tn*3*-like structure carrying bla_{TEM-1} and only the rightmost IS*26* in S044, respectively (17).

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

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

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

Methods

Bacterial isolates used in this study

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

Antimicrobial susceptibility testing

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

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

Whole-genome sequencing

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

Other studied genomes

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

Sequencing and construction of reference genome 201001922

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

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

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

Mapping and phylogenomic analysis

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

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

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

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

Assembly, annotation and pangenome analysis

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

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

Identification of resistance, virulence and phage genes

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

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

Reconstruction of SGI sequences

ISMapper v1 (47) and the assembly graph viewer Bandage (48) were used to piece together segments of the SGI in each genome. To do this, each assembly was queried with BLAST to identify which contigs contained SGI backbone and AMR genes. Each assembly was also queried for IS*26* using ISMapper's assembly improvement mode (47), identifying contigs that contained IS*26* flanking sequence. Contigs containing flanking IS*26* sequence with SGI genes or AMR genes were hypothesised to be part of the SGI. Both pieces of information (BLAST and ISMapper results) were used in conjunction with the reference SGI1-K reference sequence (accession AY463797) to determine which contigs could be joined together. In some cases, it was unclear whether IS*26*-flanked AMR genes were located within the SGI or a plasmid. In these cases, Bandage was used to examine the assembly graphs and determine the paths linking the SGI, IS*26* and AMR genes, providing additional evidence for contig connection.

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

Analysis of IncI1 and IncC plasmids

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

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

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

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

Results

Phylogenetic analysis of *S.* Kentucky ST198

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

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

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

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

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

Long-term persistence in a single patient

Three *S.* Kentucky ST198 isolates were recovered, in consecutive years (2009, 2010 and 2011) from the same patient who had been infected in Egypt (dark orange box, **Figure 1**). These isolates belonged to the MDR lineage and shared an MRCA circa 2005, suggesting persistent colonization of ~6 years duration (**Figure 1**). The 2011 isolate, 201100664, differed the most from the inferred MRCA (30 SNVs; 21 non-synonymous SNVs, 6 synonymous SNVs, 3 intergenic SNVs), yielding an estimated in vivo substitution rate of 5 SNVs per year, faster than that estimated by BEAST analysis of the whole data set. Many of the non-synonymous mutations were in genes responsible for flagella (n=7) and iron transport (n=2) (**Table S4**), although no motility changes were detected in this isolate. Eleven SNVs separated 201000305 and 09-9322 (8 non-synonymous SNVs, 2 synonymous SNVs, 1 intergenic SNV). One of these eleven SNVs was found in another iron transport gene (asmbl_3909, **Table S4**).

SGI in *S.* Kentucky

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

Almost every SGI1-positive *S.* Kentucky ST198 isolate in this study had a distinct SGI structure (**Figure 2b, Figure S4**). In addition to large deletions of the SGI backbone, some isolates had inversions of whole or part of the resistance gene segment of the island, with various deletions and rearrangements of the transposons (**Figure 2b**). There were multiple different IS*26* insertion sites within the resistance elements of the island, providing evidence that IS*26* has mediated the majority of differences found in the resistance region of the island (**Figure 2b**). We found that IS*26* was rarely present in *S.* Kentucky ST198 isolates outside of the MDR lineage (**Figure S3**). Within the MDR lineage, linear regression analysis of IS*26* copy number against year of isolation showed some evidence of IS*26* accumulation over time (0.12 IS*26* copies per year, p=0.01, R^2=0.05) (**Figure S5**).

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

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

Overall, we found that 35 isolates in the full strain set carried at least one plasmid, covering 13 different known plasmid incompatibility types (**Table S1**). Within the MDR lineage, there was extensive phenotypic and genotypic variation in antimicrobial susceptibility observed (**Figure 3**). A part of this variability could be attributed to the acquisition of plasmids carrying additional AMR genes, as 32 isolates in the MDR lineage carried genes outside the SGI that are likely plasmid-borne (**Figure 3e**). Known plasmid replicons were identified in 23 isolates, and in total we identified eight different plasmid incompatibility types across the MDR strain set (C, I1, L/M, Q1, W, X1, X4, Y). From these 23 isolates carrying known plasmid incompatibility types, we were able to determine precise plasmid-AMR gene links for 20 isolates.

There appeared to be no link between geography and plasmid type, with plasmids present in isolates from multiple different regions (**Figure S6**). The majority of genes encoding carbapenemases (*bla*$_{\text{OXA-48}}$ and *bla*$_{\text{NDM-1}}$), ESBLs (*bla*$_{\text{CTX-M-1}}$) and cephamycinases (*bla*$_{\text{CMY-2}}$, *bla*$_{\text{CMY-4}}$ and *bla*$_{\text{CMY-16}}$) were carried by either IncI1 or IncC (previously IncA/C$_2$) plasmids (**Figure 3d, 3e**). Two IncL/M plasmids were found to carry *bla*$_{\text{OXA-48}}$ or *bla*$_{\text{CTX-M-15}}$, and an IncW plasmid was found to carry *bla*$_{\text{VIM-2}}$ (**Figure 3d, 3e**). The eight isolates resistant to azithromycin contained the *mph(A)* gene. These isolates clustered into two groups. A plasmid location of *mph(A)* was found for four isolates. Three different Inc types were identified (IncI1, IncC, and IncL/M).

There was little evidence that any plasmids were being maintained as the MDR lineage evolved (**Figure 3**), although the group of three isolates recovered from the same patient in Egypt (09-9322, 201000305, 201100664; discussed above) all carried IncI1 plasmids. These three plasmids were identical in their core gene content, although IncI1 plasmids in 201100664 differed from those in the earlier two isolates by two intergenic SNVs (**Figure S7**). Interestingly these three isolates all lacked the SGI and any other chromosomal resistance genes, and their IncI1 plasmids differed substantially from one another in resistance gene content (**Figure 3e**). The two early isolates mostly carried resistance genes for aminoglycosides, sulfonamides, trimethoprim, phenicols and macrolides. The plasmid in the final isolate, 201100664, had lost almost all of the resistance genes found in the previous two isolates, except for *mph(A*), and had gained the carbapenemase-encoding *bla*$_{\text{OXA-48}}$ gene. IncI1 plasmids were detected in a further six *S.* Kentucky ST198 genomes, but these did not cluster in either the IncI1 plasmid tree or the chromosome tree, consistent with seven distinct introductions of IncI1 plasmids into the *S.* Kentucky ST198 MDR lineage, each associated with distinct AMR gene contents (**Figure 3**, **Figure S7**).

Two isolates of the MDR lineage carried IncC plasmids (99-2998 and 201410673). Both IncC plasmids were genotyped as pST3, which is commonly associated with *bla*$_{\text{CMY}}$ (50), and this cephamycinase-encoding gene was found in the plasmid from isolate 99-2998. Interestingly, the IncC plasmid in isolate 201401673 was carrying a carbapenemase-encoding *bla*$_{\text{NDM-1}}$ gene, which is more commonly found in pST1 IncC plasmids (50). This *bla*$_{\text{NDM-1}}$ gene was found in a different structural context to the *bla*$_{\text{NDM}}$ genes in the pST1 IncC plasmids; as usual it was downstream of IS*Aba125*, however instead of being upstream of *ble*, instead it was upstream of *qacEΔ1* and *sul1*, with a remnant of the *ble* gene left behind from the insertion of *qacEΔ1* (**Figure S8**). We found that this *bla*$_{\text{NDM-1}}$ region was entirely covered by WGS reads, with no breaks or gaps in coverage, supporting that it is the true structure in this plasmid (**Figure S8**). This configuration also appears in another pST3 IncC plasmid, pRH-1238, from *S. enterica* serotype Corvallis (GenBank accession KR091911), isolated from a wild bird in Germany (57).

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

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

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

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

Discussion

Our data show that nearly all MDR *S.* Kentucky ST198 belong to a single lineage that has accumulated AMR determinants since the early 1990s (**Figure 1**). It first acquired a variant of the SGI1, SGI1-K which conferred resistance to ampicillin, streptomycin, gentamicin, sulfamethoxazole, and tetracycline (**Figure 2**). The SGI1 structure appears to be highly susceptible to genetic rearrangements, with distinct forms found in each isolate likely due to the transpositional activity of IS*26*, which resulted in deletion of some or all genes inside SGI1. The loss of resistance genes was often made up for by acquisition of additional MDR plasmids (**Figure 3**).

IS*26* is 820 bp long and encodes a single transposase with 14 bp terminal repeats on each end (58). Each of the three SGI1 subtypes found in the MDR lineage carried one or more copies of IS*26*, and all genomes in the MDR lineage carried IS*26*, with no genomes outside of this clade carrying IS*26*. The recently described mechanism used by IS*26* to transpose may provide an explanation as to why the SGI variants in these isolates are so dynamic. During the transposition, IS*26* extracts itself from the donor DNA molecule, as well as DNA lying upstream of it between itself and another IS*26* element, and uses this to form a translocatable unit (59). It then finds another IS*26* element in the receiving DNA molecule, and inserts itself as well as the excised donor DNA next to it, forming a tandem array of IS*26*s in direct orientation (59). This model illustrates that IS*26* is likely the causative agent for many of the deletions, inversions and transpositions within the SGI, eventually resulting in the genesis of the different SGI1 variants (SGI1-K, SGI1-P and SGI1-Q) seen in this dataset (**Figure 2**).

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 IS*26* 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.

Acknowledgments

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

Funding information

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

Author Statements

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

Conflicts of Interest

The authors declare no conflicts of interest.

Abbreviations

AMR: Antimicrobial resistance
AST: Antimicrobial susceptibility testing
BEAST: Bayesian Evolutionary Analysis Sampling Trees
CA-SFM: Antibiogram Committee of the French Society for Microbiology
CIP: Ciprofloxacin
CIP^R: Ciprofloxacin-resistant
ENA: European Nucleotide Archive
ESBL: extended-spectrum beta-lactamase
IS: Insertion sequences
MDR: Multidrug-resistant
MLST: Multi-locus sequence typing
MIC: Minimal inhibitory concentration
PCR: Polymerase chain reaction

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

References

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

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

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

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

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

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

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

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

9. Siebor E, Neuwirth C. Emergence of Salmonella genomic island 1 (SGI1) among Proteus mirabilis clinical isolates in Dijon, France. J Antimicrob Chemother. Oxford University Press; 2013 Aug;68(8):1750–6.

10. Doublet B, Boyd D, Mulvey MR, Cloeckaert A. The *Salmonella* genomic island 1 is an integrative mobilizable element. Mol Microbiol. 2005 Mar;55(6):1911–24.

11. Carraro N, Matteau D, Luo P, Rodrigue S, Burrus V. The master activator of IncA/C conjugative plasmids stimulates genomic islands and multidrug resistance dissemination. Hughes D, editor. PLoS Genet. Public Library of Science; 2014 Oct 23;10(10):e1004714.

12. Douard G, Praud K, Cloeckaert A, Doublet B. The *Salmonella* Genomic Island 1 is specifically mobilized in trans by the IncA/C multidrug resistance plasmid family. Xu S-Y, editor. PLoS ONE. Public Library of Science; 2010 Dec 20;5(12):e15302.

13. Hall RM. *Salmonella* genomic islands and antibiotic resistance in *Salmonella enterica*. Future Microbiol. 2010 Oct;5(10):1525–38.

14. Siebor E, Neuwirth C. Proteus genomic island 1 (PGI1), a new resistance genomic island from two Proteus mirabilis French clinical isolates. J Antimicrob Chemother. Oxford University Press; 2014 Aug 11;:dku314.

15. Levings RS, Djordjevic SP, Hall RM. SGI2, a relative of *Salmonella* genomic island SGI1 with an independent origin. Antimicrob Agents Chemother. American Society for Microbiology; 2008 Jul;52(7):2529–37.

16. Doublet B, Praud K, Weill F-X, Cloeckaert A. Association of IS*26*-composite transposons and complex In4-type integrons generates novel multidrug resistance loci in *Salmonella* genomic island 1. J Antimicrob Chemother. Oxford University Press; 2009 Feb;63(2):282–9.

17. Doublet B, Praud K, Bertrand S, Collard JM, Weill FX, Cloeckaert A. 52(10):3745–54.

18. Levings RS, Partridge SR, Djordjevic SP, Hall RM. SGI1-K, a variant of the SGI1 genomic island carrying a mercury resistance region, in *Salmonella enterica* serovar Kentucky. Antimicrob Agents Chemother. American Society for Microbiology; 2007 Jan;51(1):317–23.

19. Baucheron S, Le Hello S, Doublet B, Giraud E, Weill F-X, Cloeckaert A. *ramR* mutations affecting fluoroquinolone susceptibility in epidemic multidrug-resistant *Salmonella enterica* serovar Kentucky ST198. Front Microbiol. 2013 Sep 7;4:1–6.

20. Allard MW, Strain E, Melka D, Bunning K, Musser SM, Brown EW, et al. Practical Value of Food Pathogen Traceability through Building a Whole-Genome Sequencing Network and Database. Kraft CS, editor. J Clin Microbiol. American Society for Microbiology Journals; 2016 Aug 1;54(8):1975–83.

21. Timme RE, Rand H, Leon MS, Hoffmann M, Strain E, Allard M, et al. GenomeTrakr proficiency testing for foodborne pathogen surveillance: an exercise from 2015. Microbial Genomics. Microbiology Society; 2018 Jul 1;4(7):289.

22. Inouye M, Dashnow H, Raven L-A, Schultz MB, Pope BJ, Tomita T, et al. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. Genome Med. 2014;6(11):90.

23. Kurtz S, Phillippy A, Delcher AL. Versatile and open software for comparing large genomes. Genome Biology. 2004;5(R12).

24. Holt KE, Parkhill J, Mazzoni CJ, Roumagnac P, Weill F-X, Goodhead I, et al. High-throughput sequencing provides insights into genome variation and evolution in *Salmonella* Typhi. Nat Genet. Nature Publishing Group; 2008 Aug 1;40(8):987–93.

25. Holt KE, Baker S, Weill F-X, Holmes EC, Kitchen A, Yu J, et al. Shigella sonnei genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe. Nat Genet. Nature Publishing Group; 2012 Sep;44(9):1056–9.

26. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Meth. 2012 Mar 4;9(4):357–9.

27. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. Oxford University Press; 2009 Aug 15;25(16):2078–9.

28. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res. Oxford University Press; 2014 Nov 20;:gku1196.

29. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, et al. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 2014 Apr;10(4):e1003537.

30. Rambaut A, Suchard MA, Xie D, Drummond AJ. Tracer v1.6 [Internet]. 2014. Available from: http://beast.bio.ed.ac.uk

31. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. BioMed Central Ltd; 2007;7(1):214.

32. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. Oxford University Press; 2014 May 1;30(9):1312–3.

33. United Nations Statistics Division. Composition of macro geographical (continental) regions, geographical sub-regions, and selected economic and other groupings [Internet]. 2013 [cited 2016 Aug 10]. Available from: http://unstats.un.org/unsd/methods/m49/m49regin.htm

34. Bollback JP. SIMMAP: stochastic character mapping of discrete traits on phylogenies. BMC Bioinformatics. BioMed Central Ltd; 2006;7(1):88.

35. Revell LJ. phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution. Blackwell Publishing Ltd; 2011 Dec 15;3(2):217–23.

36. Hannon Lab. FastXToolkit [Internet]. 2010 [cited 2016 Aug 10]. Available from: http://hannonlab.cshl.edu/fastx_toolkit/

37. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics. Oxford University Press; 2014 Apr 1;30(15).

38. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012 May;19(5):455–77.

39. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics. Oxford University Press; 2011 Feb 15;27(4):578–9.

40. Boetzer M, Pirovano W. Toward almost closed genomes with GapFiller. Genome Biology. BioMed Central; 2012 Jun 25;13(6):1.

41. Assefa S, Keane TM, Otto TD, Newbold C, Berriman M. ABACAS: algorithm-based automatic contiguation of assembled sequences. Bioinformatics. 2009 Aug 1;25(15):1968–9.

42. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. Oxford University Press; 2014 Mar 18;30(14).

43. Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother. American Society for Microbiology; 2014;58(1):212–20.

44. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015 Nov 15;31(22):3691–3.

45. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. BioMed Central Ltd; 2009 Dec 15;10(1):421.

46. Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: hierarchical and refined dataset for big data analysis - 10 years on. Nucleic Acids Res. Oxford University Press; 2016 Jan 4;44(D1):D694–7.

47. Hawkey J, Hamidian M, Wick RR, Edwards DJ, Billman-Jacobe H, Hall RM, et al. ISMapper: identifying transposase insertion sites in bacterial genomes from short read sequence data. BMC Genomics. BioMed Central Ltd; 2015;16(1):667.

48. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics. 2015 Oct 15;31(20):3350–2.

49. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother. American Society for Microbiology; 2014 Jul;58(7):3895–903.

50. Hancock SJ, Phan MD, Peters KM, Forde BM, Chong TM, Yin W-F, et al. Identification of IncA/C Plasmid Replication and Maintenance Genes and Development of a Plasmid Multilocus Sequence Typing Scheme. Antimicrob Agents Chemother. 2017 Feb;61(2):AAC.01740–16.

51. Leekitcharoenphon P, Nielsen EM, Kaas RS, Lund O, Aarestrup FM. Evaluation of whole genome sequencing for outbreak detection of Salmonella enterica. PLoS ONE. 2014;9(2):e87991.

52. Okoro CK, Kingsley RA, Connor TR, Harris SR, Parry CM, Al-Mashhadani MN, et al. Intracontinental spread of human invasive Salmonella Typhimurium pathovariants in sub-Saharan Africa. Nature Publishing Group. Nature Publishing Group; 2012 Sep 30;44(11):1215–21.

53. Zhou Z, McCann A, Litrup E, Murphy R, Cormican M, Fanning S, et al. Neutral Genomic Microevolution of a Recently Emerged Pathogen, Salmonella enterica Serovar Agona. Casadesús J, editor. PLoS Genet. Public Library of Science; 2013 Apr 18;9(4):e1003471.

54. Wong VK, Baker S, Pickard DJ, Parkhill J, Page AJ, Feasey NA, et al. Phylogeographical analysis of the dominant multidrug-resistant H58 clade of *Salmonella* Typhi identifies inter- and intracontinental transmission events. Nat Genet. 2015 Jun;47(6):632–9.

55. Zhou Z, McCann A, Weill F-X, Blin C, Nair S, Wain J, et al. Transient Darwinian selection in *Salmonella enterica* serovar Paratyphi A during 450 years of global spread of enteric fever. Proc Natl Acad Sci USA. National Acad Sciences; 2014 Aug 19;111(33):12199–204.

56. Duchêne S, Holt KE, Weill F-X, Le Hello S, Hawkey J, Edwards DJ, et al. Genome-scale rates of evolutionary change in bacteria. Microbial Genomics. Microbiology Society; 2016 Nov 30;2(11).

57. Villa L, Guerra B, Schmoger S, Fischer J, Helmuth R, Zong Z, et al. IncA/C Plasmid Carrying bla(NDM-1), bla(CMY-16), and fosA3 in a Salmonella enterica Serovar Corvallis Strain Isolated from a Migratory Wild Bird in... - PubMed - NCBI. Antimicrob Agents Chemother. 2015 Sep 18;59(10):6597–600.

58. Mollet B, Iida S, Shepherd J, Arber W. Nucleotide sequence of IS*26*, a new prokaryotic mobile genetic element. Nucleic Acids Res. 1983 Sep 24;11(18):6319–30.

59. Harmer CJ, Moran RA, Hall RM. Movement of IS*26*-associated antibiotic resistance genes occurs via a translocatable unit that includes a single IS*26* and preferentially inserts adjacent to another IS*26*. mBio. 2014;5(5):e01801–14.

60. Baker S, Duy PT, Nga TVT, Dung TTN, Phat VV, Chau TT, et al. Fitness benefits in fluoroquinolone-resistant Salmonella Typhi in the absence of antimicrobial pressure. eLife. 2013 Dec 10;2:452.

61. Webber MA, Ricci V, Whitehead R, Patel M, Fookes M, Ivens A, et al. Clinically Relevant Mutant DNA Gyrase Alters Supercoiling, Changes the Transcriptome, and Confers Multidrug Resistance. mBio. American Society for Microbiology; 2013 Aug 30;4(4):e00273–13–e00273–13.

62. Harmer CJ, Hamidian M, Ambrose SJ, Hall RM. Destabilization of IncA and IncC plasmids by SGI1 and SGI2 type Salmonella genomic islands. Plasmid [Internet]. 2016 Nov;87-88:51–7.

63. Huguet KT, Gonnet M, Doublet B, Cloeckaert A. A toxin antitoxin system promotes the maintenance of the IncA/C-mobilizable Salmonella Genomic Island 1. Scientific Reports. 2016;6(1):32285.

Figures and Tables

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

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

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

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

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

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

Supplementary Figure 3: Presence of SGI backbone genes in each *S.* Kentucky ST198 isolate, with an estimate of overall IS*26* copy number. Left, dated Bayesian phylogeny, with tips and branches coloured by region (as per map inset and **Figure 1**). Red blocks indicate the presence at least one resistance gene in the SGI resistance region. Light blue blocks indicate presence of chromosomal genes flanking the SGI (*trmE* and *yidY*). Dark blue blocks indicate presence of SGI backbone genes. Yellow bars show estimated IS*26* copy number in each isolate.

Supplementary Figure 4: Additional SGI1 examples. Coding regions are represented as arrows and are coloured as per legend. SGI sequences are grouped by type - SGI1-K, SGI1-P and SGI1-Q. Contig breaks are shown by thick black vertical lines, with IS*26* flanking regions detected with ISMapper marked with purple vertical lines.

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

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

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

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 $bla_{\text{NDM-1}}$ 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.

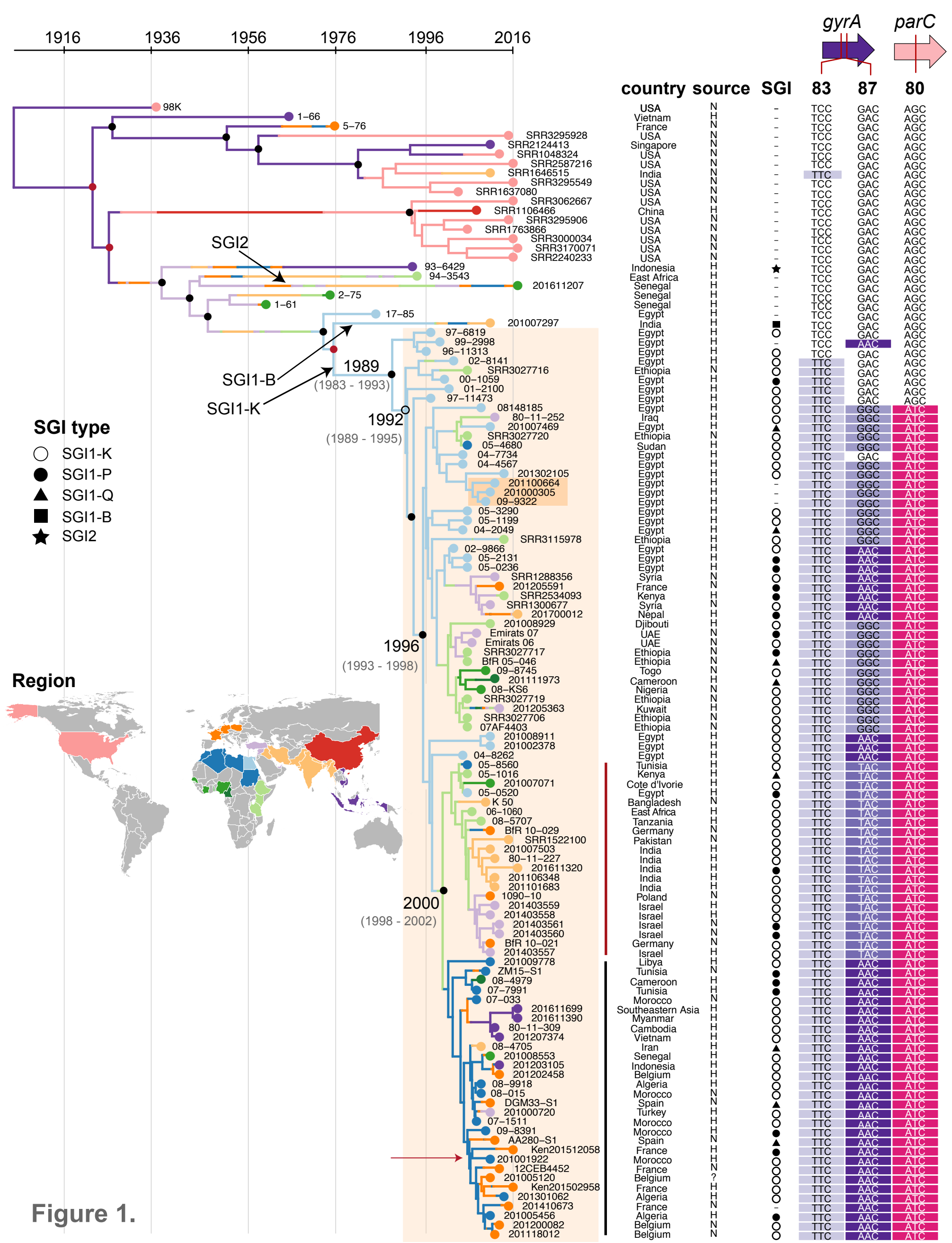
1916
1936
1956
1976
1996
2016
gyrA
parC
country source SGI 83 87 80
SGI2
SGI1-B
SGI1-K
1989
(1983 - 1993)
1992
(1989 - 1995)
1996
(1993 - 1998)
2000
(1998 - 2002)
SGI type
SGI1-K
SGI1-P
SGI1-Q
SGI1-B
SGI2
Region

Figure 1.

Figure 2.

Figure 3.

(a) (b) (c) (d) (e)

Spectinomycin Streptomycin Gentamicin Kanamycin Tobramycin Netilmicin Amikacine Sulfonamides Trimethoprim Tetracycline Chloramphenicol Azithromycin Ampicillin Ceftriaxone Ceftazidime Cefoxitin Imipenem

aacA5 dfrA17 aacA5 aadA7 sul1 tetR tetA strB strA blaTEM

97–6819
99–2998
96–11313
02–8141
SRR3027716
00–1059
01–2100
97–11473
08148185
80–11–252
201007469
SRR3027720
05–4680
04–7734
04–4567
201302105
201100664
201000305
09–9322
05–3290
05–1199
04–2049
SRR3115978
02–9866
05–2131
05–0236
SRR1288356
201205591
SRR2534093
SRR1300677
201700012
201008929
Emirats 07
Emirats 06
SRR3027717
BfR 05–046
09–8745
201111973
08–KS6
SRR3027719
201205363
SRR3027706
07AF4403
201008911
201002378
04–8262
05–8560
05–1016
201007071
05–0520
K 50
06–1060
08–5707
BfR 10–029
SRR1522100
201007503
80–11–227
201611320
201106348
201101683
1090–10
201403559
201403558
201403561
201403560
BfR 10–021
201403557
201009778
ZM15–S1
08–4979
07–7991
07–033
201611699
201611390
80–11–309
201207374
08–4705
201008553
201203105
201202458
08–9918
08–015
DGM33–S1
201000720
07–1511
09–8391
AA280–S1
Ken201512058
201001922
12CEB4452
201005120
Ken201502958
201301062
201410673
201005456
201200082
201118012

C strAB sul1 sul2 tetAR floR blaCMY-4
X1 aadA1 sul3 cmlA floR
(tetAR)
(aadA2 dfrA12)
(aph3-Ia aadA2 catA1 dfrA12 mphA)
(aph3-Ia aadA2 catA1 dfrA12)
(aadA2 aph3-1a catA1 dfrA12 sul1 mphA)
I1 blaOXA-48 mphA
I1 aadA7 aph3-Ia sul1 dfrA12 catA1 mphA
I1 aadA2 aph3-Ia sul1 dfrA12 catA1 mphA (blaCMY-2)
(aph3-Ia)
I1 aac3-IId aadA22
I1 (sul3)
(aph3-Ia)
Q1 aph3-Ia sul2
(aadA5 tetB)
I1 (aadA1 dfrA12 blaOXA-10 arr2)
X1 floR sul3
I1 aadA17 sul3 cmlA
Y aadA12 aac3-IId dfrA12
(aacA5)
(aacA5)
X1 strB floR
I1 blaCTX-M-1
I1 blaCMY-2
X1 strAB sul2
W aac3-I blaVIM-2
L/M, Q1 strAB blaOXA-48 (sul3 dfrA14 mphA)
C aadA5 aphA6 strAB sul1 sul2 dfrA17 tetAR floR mphA msrE ermB blaCMY-16 blaNDM-1 fosA
L/M aac3-IId aadA2 armA dfrA12 mphA msrE blaCTX-M-15
X4 aadA22
X4 aadA22

Region
- Egypt
- Northern Africa
- Eastern Africa
- Western Africa
- Middle Africa
- South-Eastern Asia
- Western Asia
- Southern Asia
- Europe

SGI type
- SGI1-K
- SGI1-P
- SGI1-Q

Antibiotic class
- aminoglycosides
- sulfonamides
- trimethoprim
- tetracyclines
- phenicols
- macrolides
- penicillins
- cephalosporins
- carbapenems
- other
- no information