Microevolution of Monophasic *Salmonella* Typhimurium during Epidemic, United Kingdom, 2005–2010

Liljana Petrovska, Alison E. Mather,[1] Manal AbuOun, Priscilla Branchu, Simon R. Harris, Thomas Connor, K.L. Hopkins, A. Underwood, Antonia A. Lettini, Andrew Page, Mary Bagnall, John Wain, Julian Parkhill, Gordon Dougan, Robert Davies, Robert A. Kingsley

Microevolution associated with emergence and expansion of new epidemic clones of bacterial pathogens holds the key to epidemiologic success. To determine microevolution associated with monophasic *Salmonella* Typhimurium during an epidemic, we performed comparative whole-genome sequencing and phylogenomic analysis of isolates from the United Kingdom and Italy during 2005–2012. These isolates formed a single clade distinct from recent monophasic epidemic clones previously described from North America and Spain. The UK monophasic epidemic clones showed a novel genomic island encoding resistance to heavy metals and a composite transposon encoding antimicrobial drug resistance genes not present in other *Salmonella* Typhimurium isolates, which may have contributed to epidemiologic success. A remarkable amount of genotypic variation accumulated during clonal expansion that occurred during the epidemic, including multiple independent acquisitions of a novel prophage carrying the *sopE* gene and multiple deletion events affecting the phase II flagellin locus. This high level of microevolution may affect antigenicity, pathogenicity, and transmission.

Salmonella enterica is one of the most common enteric pathogens of humans and animals. An estimated 94 million cases of nontyphoidal salmonellosis occur worldwide each year, causing considerable illness and death; in the United States, the associated economic burden estimated by the US Centers for Disease Control and Prevention is >$2 billion US per year (*1*,*2*).

S. enterica consists of >2,500 serovars, of which *S. enterica* serovar Typhimurium (*Salmonella* Typhimurium) is the most ubiquitous in zoonotic reservoirs for human infection and the environment (*3*). Over the past half century, the epidemiology of *Salmonella* Typhimurium has been characterized by successive waves of dominant multidrug-resistant clones (*4*). During 1966–2010 in Europe, where variants are distinguished by definitive (phage) type (DT), *Salmonella* Typhimurium DT9, DT204, DT104, and DT193 emerged successively as multidrug-resistant strains (*5*). Epidemic strains dominate for 4–15 years before being replaced by a new dominant phage type. The emergence and spread of *Salmonella* Typhimurium DT104 was global (*6*) and largely responsible for the increased multidrug-resistant *Salmonella* isolates in Europe and North America in the 1990s (*7*). As DT104 incidence has waned in the United Kingdom, monophasic variants of *Salmonella* Typhimurium with the antigenic formula 1,4,[5],12:i:- have emerged (*8*), although it is not clear if this current monophasic *Salmonella* Typhimurium epidemic is related to other epidemics of monophasic variants previously reported in North America (*9*), Spain (*10*), and elsewhere in Europe (*11*). Analysis of the genomic deletions in the phase II flagellum locus responsible for the monophasic phenotype suggested that multiple independent clones may be emerging in the United States and Europe (*9*).

The first description of a monophasic *Salmonella* Typhimurium epidemic in Europe was that of a "Spanish clone," which emerged rapidly during 1997 and was characterized by a deletion in the allantoin–glyoxylate operon and the fljAB operon, phage type U302, and resistance pattern ACSuGSTSxT (resistant to ampicillin, chloramphenicol, sulfonamide, gentamicin, streptomycin, tetracycline, and co-trimoxazole) (*10*). Since this time, many European countries have reported increased incidence of this serotype, particularly associated with pig herds (12–15) but later with cattle (*16*,*17*). However, in contrast to the Spanish clone, these current monophasic *Salmonella* Typhimurium epidemic strains have commonly been associated with phage types DT193 or DT120 and a predominant

Author affiliations: Animal and Plant Health Agency, Addlestone, UK (L. Petrovska, M. AbuOun, M. Bagnall, R. Davies); The Wellcome Trust Sanger Institute, Cambridge, UK (A.E. Mather, S.R. Harris, T. Connor, A. Page, J. Parkhill, G. Dougan, R.A. Kingsley); Institute of Food Research, Norwich, UK (P. Branchu, R.A. Kingsley); Public Health England, London, UK (K.L. Hopkins, A. Underwood, J. Wain); Istituto Zooprofilattico Sperimentale delle Venezie, Legnaro (PD), Italy (A.A. Lettini)

DOI: http://dx.doi.org/10.3201/eid2204.150531

[1]Current affiliation: University of Cambridge, Cambridge, UK.

ASSuT tetraresistance pattern (resistant to ampicillin, streptomycin, sulfonamide, and tetracycline), suggesting that the epidemics are distinct.

The molecular basis for the success of epidemic clones of bacterial pathogens has implications for the surveillance and management of infectious diseases. Epidemiologic success depends on selective advantage of epidemic clones, resulting from their unique genotype. The current multidrug-resistant *Salmonella* 4,[5],12:i:- epidemic in the Europe was first reported around 2005 and is mainly associated with isolates of phage types DT193 and DT120 (18).

We investigated the phylogenetic relationship of 206 strains of *Salmonella* Typhimurium (*Salmonella* 1,4,[5]:i:1,2) and monophasic *Salmonella* Typhimurium (*Salmonella* 1,4,[5],12:i:-), isolated from humans, livestock, or contaminated food from the United Kingdom or Italy from 1993 through 2010. We report the whole-genome sequence variation of *Salmonella* Typhimurium and *Salmonella* 1, 4,[5],12:i:- isolates from the United Kingdom and Italy and the application of these data to phylogenetic reconstruction of the epidemic. We address the questions of whether the monophasic *Salmonella* Typhimurium isolates in the United Kingdom are part of a single epidemic and how they are related to previously circulating biphasic and monophasic *Salmonella* Typhimurium strains.

Materials and Methods

We used bacterial isolates from strain collections held by the Animal and Plant Health Agency (Addlestone, UK); Public Health England (Colindale, London, UK); or the National Regional Laboratory for *Salmonella,* Istituto Zooprofilattico Sperimentale delle Venezie (Legnaro, Italy). The serotype and phage type were determined as previously described (*19*). The presence of the *fljB* locus and the occupancy of the *thrW* locus was initially determined by PCR amplification as previously described (*11*). Strain selection was intended to represent the diversity of *Salmonella* Typhimurium in the United Kingdom and not to be representative of the epidemiology (online Technical Appendix 1, http://wwwnc.cdc.gov/EID/article/22/4/15-0531-Techapp1.xlsx).

To determine antimicrobial drug sensitivity, we tested isolates from animals in the United Kingdom and Italy for susceptibility to antimicrobial drugs according to standard procedure (20). Resistance or susceptibility were interpreted on the basis of British Society for Antimicrobial Chemotherapy break points; we report the intermediate category as resistant. We determined antimicrobial drug sensitivity of isolates from human patients in the United Kingdom by using a modified break-point technique on Iso-Sensitest agar (Oxoid, Basingstoke, UK) (online Technical Appendix 2, http://wwwnc.cdc.gov/EID/article/22/4/15-0531-Techapp2.pdf). The MIC for copper sulfate was the concentration at which bacterial growth optical density 600 nm was >0.1 after culture (without shaking) at 37°C for 24 hours in Luria Bertani (Oxoid) broth buffered with 25 mmol/L HEPES (4-[2-hydroxyethyl]-1-piperazineethanesulfonic acid) at pH7. We then determined the whole-genome sequence by using the HiSeq Illumina (http://www.illumina.com) platform, sequence analysis, de novo assembly, annotation, and PCR amplification (online Technical Appendix 2).

Results

Salmonella 4,[5],12:i:- Strains

We determined that contemporary *Salmonella* 4,[5],12:i:- strains in the United Kingdom are part of a single clonally expanding clade. We constructed a maximum-likelihood phylogeny of all 97 monophasic and 142 *Salmonella* Typhimurium strains (online Technical Appendix 1) by using 12,793 variable sites in the genome, with reference to the whole-genome sequence of reference strain SL1344, excluding single-nucleotide polymorphisms (SNPs) in prophage, insertion sequence elements, and repetitive sequences (Figure 1). Most (77 of 97) monophasic strains were from a single distinct clade that seemed to be part of the current monophasic *Salmonella* Typhimurium epidemic because they were the most abundant and most recently isolated strains. However, older monophasic isolates were also found in at least 3 other clades within the *Salmonella* Typhimurium tree (Figure 1, indicated with *). A clade containing 8 isolates including 2 DT191a (Figure 1, indicated with†) was closely related to a *Salmonella* 1,4,[5],12:i:- isolate from the North American epidemic strain CVM23701 (*9*). Only 6 SNPs distinguished this isolate from strain H07 474 0455. In addition, a clade containing 6 *Salmonella* Typhimurium var. Copenhagen (4,12:i:1,2) strains (e.g., H070160417) and a clade containing 4 isolates (e.g., H103720606) contained monophasic strains.

Phylogenetics of Monophasic *Salmonella* Typhimurium

A maximum-likelihood phylogenetic tree, reconstructed by using variable sites within the whole-genome sequence with reference to the draft genome sequence of a representative strain from within the epidemic (strain SO4698-09), indicated a clonally expanding clade with a maximum root-to-tip distance of ≈70 SNPs. This finding indicated that all strains in the tree shared a common ancestor in the recent past (Figure 2). All isolates from this monophasic clade were of sequence type 34. The phage type of monophasic epidemic isolates varied according to phylogeny. Most isolates were DT193 (38 of 62 typed) or DT120 (9) and various other phage types including DT7 (3), DT191a (1), DT21 (1), DT21var (1), U311 (3), U302 (2), and RDNC (3). However, although virtually all isolates in subclades

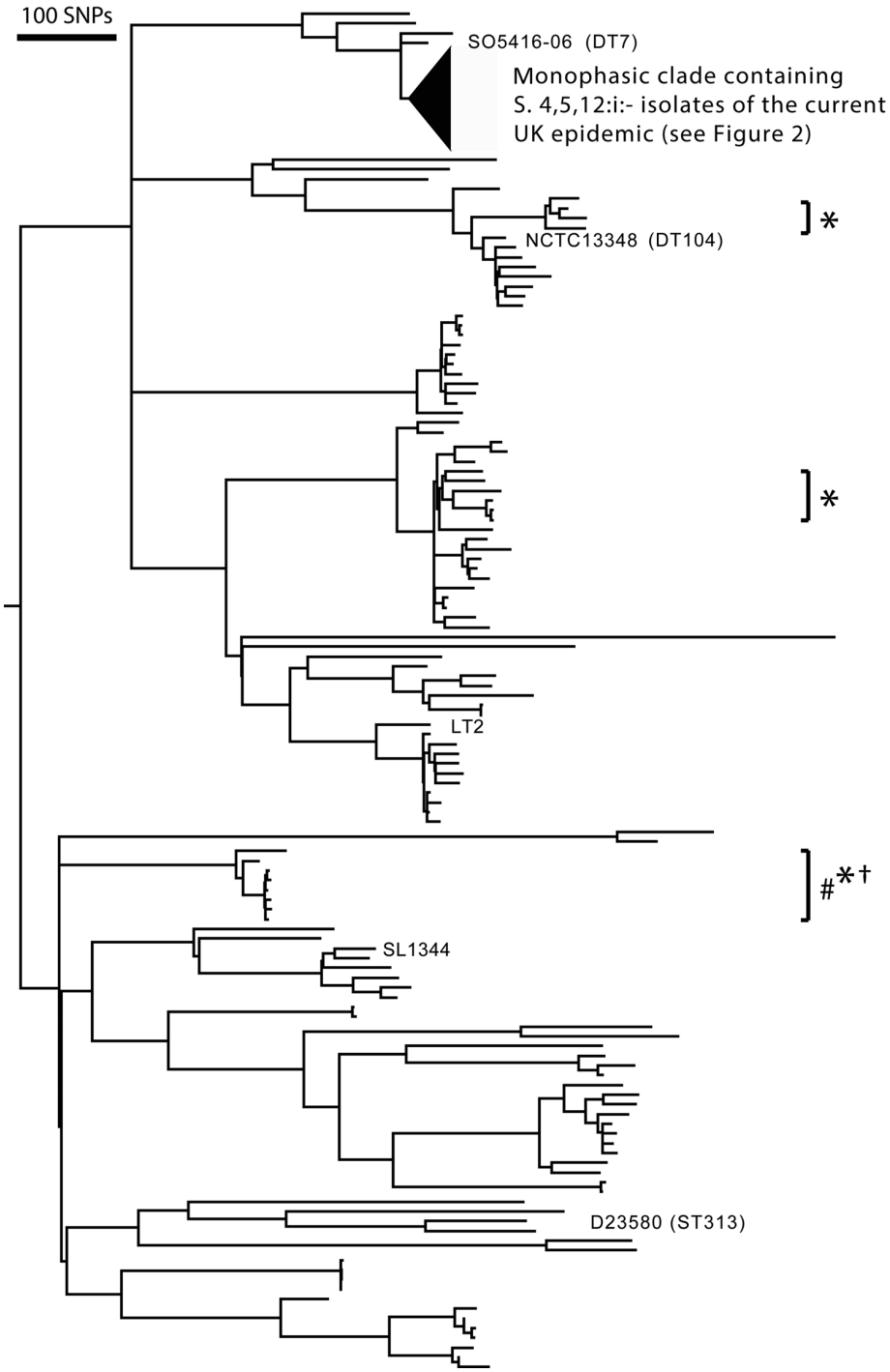


Figure 1. Phylogeny of *Salmonella enterica* serovar Typhimurium (*Salmonella* Typhimurium) and *Salmonella* 1,4,[5],12:i:- isolates from the United Kingdom and Italy, 2005–2010. Maximum-likelihood tree of 212 *Salmonella* Typhimurium and monophasic isolates was constructed by using 12,793 single-nucleotide polymorphisms (SNPs) outside of prophage elements, insertion sequence elements and sequence repeats identified by reference to the whole-genome sequence of *Salmonella* Typhimurium strain SL1344. The tree is rooted with *Salmonella* Enteritidis whole-genome sequence as an outgroup (note shown). The lineage containing the *Salmonella* 1,4,[5],12:i:- current UK epidemic group is conflated for simplicity (filled triangle). The designation of the isolates (left column) and phage type are shown (right column). *Monophasic isolates outside of the main epidemic clade. †Monophasic clade closely related to the monophasic clone CVM23701 from North America (*9*). DT, definitive (phage) type; ND, not determined. Scale bar indicates the approximate number of SNPs determined by genetic distance and the number of SNPs used to construct the tree. An expanded version of this figure is available online (http://wwwnc.cdc.gov/EID/article/22/4/15-0531-F1.htm).

A and B were DT193, the phage type was highly variable in subclade C. Biphasic DT193 strains (e.g., 4061-1997; Figure 1) isolated before 2005 were not direct ancestors of the current monophasic *Salmonella* Typhimurium epidemic because they were present on a distinct lineage. Indeed, DT193 isolates were present on 4 distinct lineages within the phylogenetic tree, highlighting the polyphyletic nature of this phage type (Figure 1). Isolates from UK animals in subclade C were relatively scarce; 1 of 21 isolates in this subclade was from a UK animal. Instead, isolates from this subclade came predominantly from humans in the United Kingdom and humans and animals in Italy. In contrast, isolates from subclade A were mostly (18 of 32) of livestock origin; only 5 were of human origin. Clade B contained an approximately equal number of human and livestock isolates. Furthermore, although isolates from UK pigs were

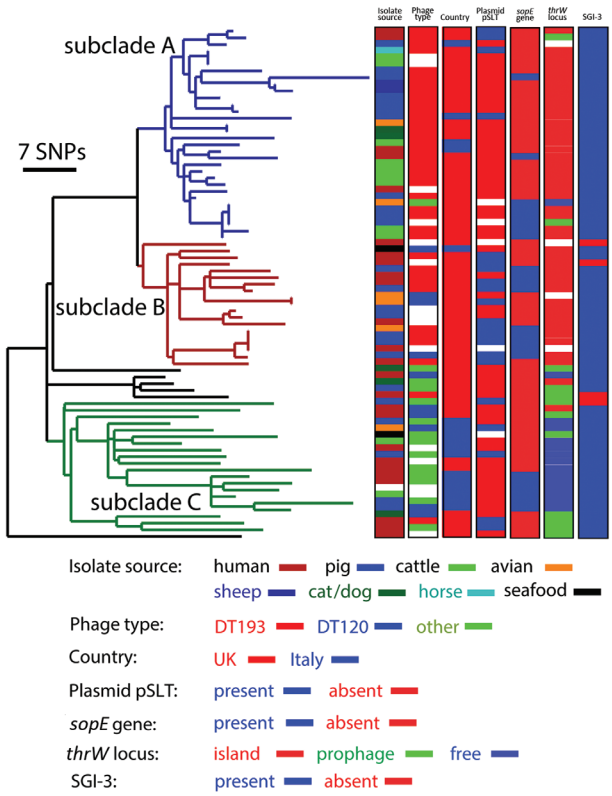


Figure 2. Phylogeny of *Salmonella* 1,4,[5],12:i:- epidemic clade isolates from the United Kingdom and Italy, 1993–2010. Maximum-likelihood tree of 77 *Salmonella* 1,4,[5],12:i:- isolates rooted with *Salmonella* Typhimurium strain SL1344 was constructed by using 1,058 single-nucleotide polymorphisms (SNPs) outside of prophage elements, insertion sequence elements, and sequence repeats identified with reference to whole-genome sequence of *Salmonella* Typhimurium strain SO4698-09. Subclades A (blue lineages), B (red lineages), and C (green lineages) are indicated. Strain designations are color coded for isolates from humans (red) and animals (blue). Epidemiologic data for the source of isolate, phage type, country of origin, presence of the virulence plasmid (pSLT), presence of the *sopE* gene, occupancy of the *thrW* locus, and presence of *Salmonella* genetic island 3 are indicated (right). Scale bar indicates the approximate number of SNPs determined by genetic distance and the number of SNPs used to construct the tree. An expanded version of this figure is available online (http://wwwnc.cdc.gov/EID/article/22/4/15-0531-F2.htm).

present in all 3 subclades, isolates from UK cattle were present only in subclade A, consistent with epidemiologic reports that the epidemic originated in pig herds and later spread to cattle herds (*17*). Despite analysis inclusion of only 6 isolates from birds, these were distributed throughout the tree, suggesting multiple transmission events into these animal populations. The distribution of isolates from humans and livestock (pigs, cattle, and sheep) within subclades of the phylogenetic tree of UK monophasic isolates was also strikingly uneven. Most (64 of 77) isolates were ASSuT tetraresistant, and the corresponding resistance genes were detected in de novo assembled sequences (online Technical Appendix 2 Figure 1), suggesting that the most recent common ancestor (MRCA) of the clade had this complement of resistance genes. However, during clonal expansion, 7 strains had lost their resistance genes entirely and another 7 had an altered complement of resistance genes.

Novel Genetic Island Encoding Resistance to Heavy Metals

A large novel genomic island (designated SGI-3) specific to the monophasic *Salmonella* Typhimurium epidemic clade is inserted at the *yjdC* locus (online Technical Appendix 2 Figure 2) in strain SO4698-09. The island contained ≈90 genes, some of which had sequences similar to those associated with plasmid transfer and conjugation, and an integrase gene, suggesting that the island may have originated by integration of a plasmid. Determination of the accessory genome indicated that the island was present in 74 of 77 isolates within the monophasic clade (Figure 2) but was absent from all strains from outside the clade. Ancestral state reconstruction performed by using ACCTRAN (*21*) (online Technical Appendix 2 Figure 3, panel A) suggested that this island was probably introduced shortly before clonal expansion of the monophasic clade. Three clusters of genes similar to genes involved in resistance to heavy metals are present on the island. Consistent with the island contributing to enhanced resistance to copper sulfate, a common animal feed additive, the MIC ($p = 0.015$) for copper sulfate was significantly greater for isolates within the monophasic *Salmonella* Typhimurium clade (24.2 ± 1.9 mmol/L) than for *Salmonella* Typhimurium isolates from outside this clade (21.2 ± 1.1 mmol/L) that did not encode the island (online Technical Appendix 2 Figure 4).

Genotypic Variation in the *fljBA* and *thrW* Loci and Loss of the Virulence Plasmid

The monophasic phenotype results from the absence of phase-2 flagellin monomer FljB. The presence of the *fljBA* genes and the neighboring genome sequence of *Salmonella* Typhimurium and monophasic variants, determined by mapping raw sequence read data to the *fljB* locus region of the SL1344 whole-genome sequence (online Technical Appendix 2 Figure 5, panel A), indicated that the UK epidemic strains are monophasic because of multiple independent deletion events that occurred during clonal expansion. Four *Salmonella* Typhimurium isolates (2 DT7 isolates [SO5416–06 and H09164 0090], 1 DT135 isolate [SO6221–07], and 1 DT177 isolate [H08390 0191]) that were closely related and shared a common ancestor with the monophasic epidemic strains (Figure 1) encoded the entire

fljBA locus, indicating that the MRCA with these strains and the epidemic strains was biphasic. In contrast, 67 of 77 monophasic *Salmonella* Typhimurium strains from the epidemic clade lacked at least part of the *fljBA* locus, resulting from deletions ranging in size and with a distribution that was consistent with the phylogenetic relationship of the strains (online Technical Appendix 2 Figure 5, panel A). The 8 epidemic strains that did not have a deletion in the *fljB* locus were deeply rooted in the tree, consistent with multiple deletion events (1–36 kb) occurring since clonal expansion of the clade. Most deletions shared a common junction in the intergenic region of *fljB* and *iroB*. Because it was not possible to assemble short read sequence data across the *fljB* locus deletion region, to investigate the nature of the deletion, we generated long read sequence data for a representative isolate SO4698–09 by using the PacBio sequencing platform (Pacific Biosciences, Menlo Park, CA, USA). A single contig assembly of these data revealed a 15,726-bp deletion of the genome relative to SL1344 and a 27,473-bp insertion of a novel sequence (online Technical Appendix 2 Figure 5, panel B). The inserted sequence was similar to sequences of several genes from transposon Tn21, mercury resistance genes (*merTABCDE* and *merR*), and antimicrobial drug resistance genes, consistent with the resistance profile of this strain (*strA*, *strB*, *sul2*, *tet[B]*, and $bla_{\text{TEM-1}}$). The composite transposon insertion was not present in closely related isolates (e.g., SO5416-06) (Figure 1) that were outside of the monophasic clade, suggesting that it was acquired by the MRCA of the monophasic clade and not before clonal expansion. The deletions in the *fljB* locus of monophasic strains from outside the main clade from the United Kingdom were distinct from that in the UK monophasic clade but identical to those described for strains from epidemics in North America (e.g., CVM23701) (*9*) and Spain (e.g., 1115/25) (*10*) (online Technical Appendix 2 Figure 5, panel A).

In addition to hypervariability at the *fljB* locus, isolates from the epidemic group exhibited sporadic loss of the virulence plasmid pSLT. The pattern of plasmid loss within the clade could be most parsimoniously explained by loss during clonal expansion. Of note, the loss of pSLT was not uniform across the monophasic tree. Although only 13% and 20% of isolates tested contained pSLT in subclades A and C, respectively, in contrast, >70% of isolates in subclade B contained the plasmid (Figure 2).

sopE Virulence Gene

The *sopE* virulence gene was acquired on a novel prophage, mTmV (monophasic *Salmonella* Typhimurium V), by multiple independent events during clonal expansion of the epidemic clade. The *thrW* locus of contemporary monophasic *Salmonella* Typhimurium isolates has been reported to harbor either a prophage, a novel genetic island, or neither (*11*). In strain SO4698-09, the *thrW* locus contains the novel genetic island described previously but also an additional prophage element encoding the *sopE* gene that together total 55 kb. Determination of the accessory genome by using the Roary pan genome pipeline (*22*) indicated that 23 of 77 monophasic isolates from the epidemic clade contained the *sopE* gene (Figure 2). SopE is a guanine exchange factor involved in subversion of the host enterocyte cytoskeleton, a key component of the infection process (*11*,*23*,*24*). The *sopE* gene was present in 6 distinct clusters of the monophasic clade, and ancestral state reconstruction indicated that multiple independent acquisitions followed by clonal expansion of the *sopE* -positive variant was the most likely explanation for their distribution (online Technical Appendix 2 Figure 3, panel B). The *sopE* gene of strain SO4698-09 is present on a 55-kb region, designated mTmV phage, which was absent from strain SL1344 and shared the greatest similarity with the *Shigella flexneri* V prophage (online Technical Appendix 2 Figure 6) (*25*). The mTmV phage from SO4698-09 was not related to the FELS-2 prophage of *Salmonella* Typhimurium strain SL1344, which also encodes the *sopE* gene, except in a 2,443-bp region that encoded the *sopE* gene and flanking sequence. Examination of partial assemblies of other monophasic strains encoding *sopE* revealed that the gene was associated with the same prophage and inserted between the genome region corresponding to the *thrW* locus. These data indicated that a novel *sopE* phage entered the genome on at least 6 occasions during the clonal expansion of the epidemic clade. Because the *sopE* gene was present in phylogenetic clusters toward the terminal branches of the monophasic clade tree and subsequently exhibited clonal expansion, we addressed the question of whether the proportion of strains that encoded the *sopE* gene in our strain collection each year changed during 2005–2010. The frequency distribution for each year was determined from collated data from 59 strains for which date of isolation and sequence data were available and an additional 41 randomly selected monophasic strains from the United Kingdom for which the presence of the *sopE* gene was determined by PCR (Figure 3; online Technical Appendix 2 Table). Increased frequency, ranging from none in 2005 and 2006 to 40% in 2010, suggested that acquisition of this gene may have conferred a competitive advantage.

Discussion

We identified a remarkable level of microevolution during clonal expansion of the epidemic. Such expansion may affect the antigenicity, pathogenicity, and transmission of monophasic *Salmonella* Typhimurium.

The phylogenetic relationships of *Salmonella* 1,4,[5],12:i:- isolated from the United States and Europe

since the late 1990s is unclear from reports to date. Our analyses suggest that at least 3 distinct epidemics have been associated with *Salmonella* 1,4,[5],12:i:- and that most of the monophasic isolates from livestock and humans in the United Kingdom since 2006 are not directly related to isolates from either the epidemic in Spain around 1997 (*10*) or the epidemic in the United States around 2004 and 2007 (*9*). Instead, the UK epidemic is related to that reported in Germany and elsewhere since around 2005 (*11*). The US clone is characterized by a large deletion in the *fljB* locus and acquisition of a prophage, neither of which were present in the UK monophasic clone. Furthermore, the whole-genome sequence for a single isolate from the US epidemic (CVM23701) placed this isolate in a small clade of monophasic isolates from the United Kingdom isolated around 1995, distinct from the current UK clade. The clone from Spain is characterized by variable size deletions in the fljB locus, all distinct from deletions observed in the UK isolates, and a deletion in the allantoin metabolism locus, also absent from the main UK clade. The MRCA of the UK *Salmonella* 1,4,[5],12:i:- epidemic in our strain collection was shared with a biphasic *Salmonella* Typhimurium isolate with DT7 (strain H091640090), a relatively rare phage type that has not been associated with epidemics in the epidemiologic record. The common ancestor with strain H091640090 probably existed in the recent past (≈20 years) because only ≈10 SNPs have accumulated in the genome since the lineages diverged, according to the short-term substitution rate (1–2 SNPs/genome/y) previously reported for Salmonella epidemics (*26,27*).

Because virtually all monophasic strains from the current epidemic clade encoded SGI-3 but isolates from outside the clade did not, initiation of clonal expansion was probably accompanied by the acquisition of this genomic island. SGI-3 encodes resistance to heavy metals, including copper and zinc, which is potentially relevant because these are supplements commonly added to pig feed as micronutrients and general antimicrobials (*28*). Indeed, in the European Union, heavy metals have been used increasingly in response to the ban on nonspecific use of antimicrobial drugs in animal feed for growth promotion (*29*). Concentration of heavy metals in pig intestines may represent substantial selective pressure contributing to the success of this clone. Indeed, a recent study reported that an enhanced MIC (20–24 mmol/L) compared with the baseline MIC (16 mmol/L) for copper sulfate was significantly more likely to be found in isolates from pig feces (*30*).

A remarkable feature of the monophasic *Salmonella* Typhimurium epidemic in the United Kingdom is the considerable number of polymorphisms that affect coding capacity that occurred during the short period (≈10–15 years) of clonal expansion of the epidemic clade. These include a complex pattern of deletions in the *fljB* locus and surrounding genome sequence, insertions in the thrW locus, and acquisition of a novel phage carrying the sopE gene. These polymorphisms seem to be stable and not deleterious because they all appear in parts of the tree that have subsequently undergone further clonal expansion. Deletions in the *fljB* locus that occurred subsequent to the initial clonal expansion of the epidemic clade accounted for the monophasic phenotype exhibited by most of these isolates. The high frequency of deletions in this locus may be the result of a composite Tn21-like transposable element that is inserted in the hin–iroB intergenic region, a well-known characteristic of such insertions (*31*).

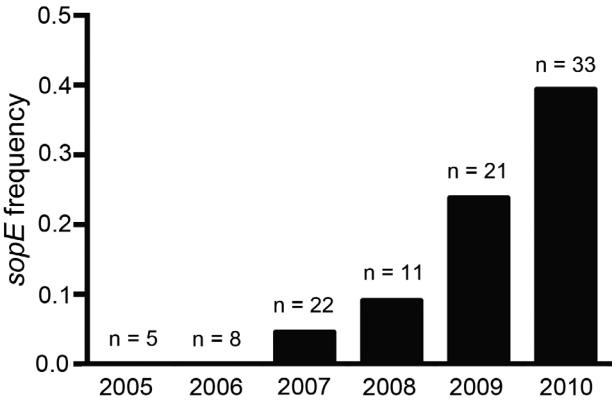


Figure 3. Frequency (proportion) of carriage of the *sopE* gene in *Salmonella* 1,4,[5],12:i:- epidemic isolates from the United Kingdom and Italy for each year during 2005–2010. The presence of the *sopE* gene was detected in draft genome assemblies by sequence comparison or by PCR amplification of genomic DNA by using primers specific for the *sopE* gene of randomly selected monophasic isolates from each year. The number of isolates investigated for each year is indicated above the bar.

The acquisition of the sopE gene on a novel prophage element that occurred through multiple recent independent events may strongly affect the pathogenesis and epidemiology of the current epidemic. Lysogeny by phages carrying the sopE gene has been associated with epidemic strains of *Salmonella* Typhimurium and of other *Salmonella* serotypes (*32*). The expression of SopE may increase the fitness of the pathogen, a possibility consistent with the observation that recent acquisition of the sopE gene by monophasic epidemic isolates has been followed by an increase in the frequency of sopE-positive isolates. The ability to induce inflammatory diarrhea is a main strategy for the transmission of *Salmonella* Typhimurium. SopE is a guanine exchange factor that activates both cdc42 and rac1; sopE2 activates only cdc42 (*33*). All *Salmonella* Typhimurium strains sequenced to date encode the sopE2 gene that exhibits 59% identity with SopE. The additional activity of SopE has a marked effect on the outcome of the interaction of *Salmonella* Typhimurium with the intestinal mucosa, resulting in

increased amounts of salmonellae in the intestinal lumen and shedding in the feces. SopE expression results in increased production of host nitrate, a valuable electron acceptor used by *Salmonella* Typhimurium for respiration (34).

In conclusion, our findings indicate that the current monophasic *Salmonella* Typhimurium clone associated with many animal species and human clinical infections in the United Kingdom arose recently. Subsequent microevolution in a short time has resulted in considerable genotypic variation affecting antigens, virulence factors, and resistance loci. Some genomic features, such as resistance to heavy metals, may have resulted in initial selection for the current clone, while more recent horizontal gene transfer or deletions and plasmid loss may have generated variation selected during the epidemic.

Addendum

It has come to the authors' attention that the designation "Salmonella Genetic Island 3 (SGI-3)" has been previously assigned to a 31-kb genomic island in a strain of Salmonella Mississippi (http://dx.doi.org/10.1371/journal.pone.0041247). To avoid confusion in the literature, we propose that the SGI-3 referred in our manuscript be designated SGI-4 in future reference.

This work was funded in part by Biotechnology and Biological Sciences Research Council grant BB/M025489/1 awarded to R.A.K., by Wellcome Trust core funding to the Pathogen Genomics team at the Sanger Institute, and by a grant from Defra OZO338. A.E.M. was supported by Wellcome Trust grant 098051 while at the Wellcome Trust Sanger Institute and Biotechnology and Biological Sciences Research Council grant BB/M014088/1 at the University of Cambridge.

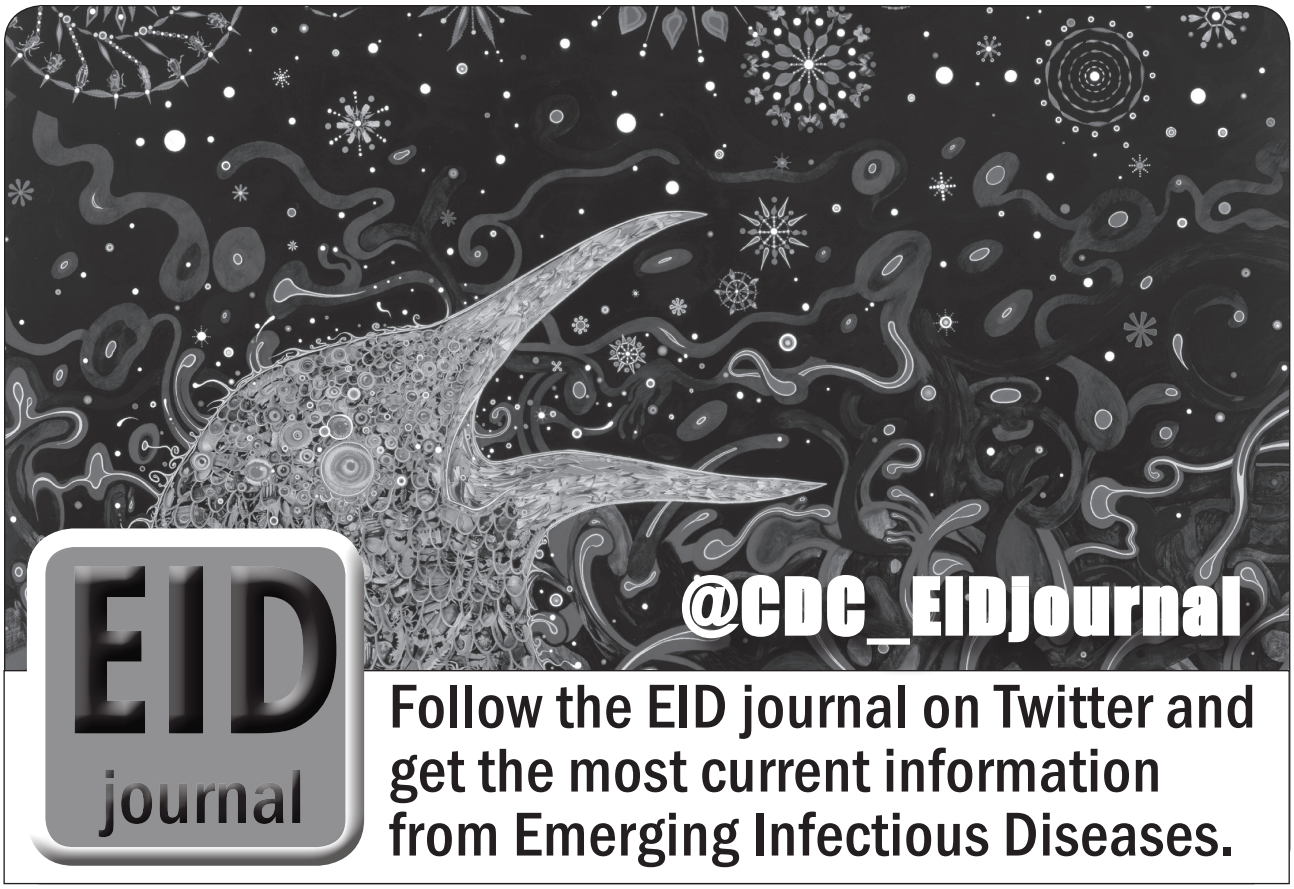
Dr. Petrovska is a research scientist at the Animal and Plant Health Agency, Weybridge, UK, and leads the Bacterial Pathogen Genomics Group. Her main research interest is comparative genomics, using high-throughput sequencing methods and functional analyses to understand the emergence and spread of infectious diseases.

References

1. Herikstad H, Motarjemi Y, Tauxe RV. *Salmonella* surveillance: a global survey of public health serotyping. http://dx.doi.org/10.1017/S0950268802006842. Epidemiol Infect. 2002;129:1–8.
2. Majowicz SE, Musto J, Scallan E, Angulo FJ, Kirk M, O'Brien SJ, et al. The global burden of nontyphoidal *Salmonella* gastroenteritis. Clin Infect Dis. 2010;50:882–9. http://dx.doi.org/10.1086/650733
3. Wales A, Davies RH. Environmental aspects of Salmonella. In: Barrow PA, Methner U, editors. Salmonella in domestic animals. 2nd ed. Wallingford (UK): CAB International; 2013. p. 399–425.
4. Rabsch W, Truepschuch S, Windhorst D, Gerlach RG. Typing phages and prophages of *Salmonella*. In: Porwollik S, editor. *Salmonella*, from genome to function. Norfolk (UK): Caister Academic Press; 2011. p. 25–48.
5. Rabsch W, Tschape H, Baumler AJ. Non-typhoidal salmonellosis: emerging problems. Microbes Infect. 2001;3:237–47. http://dx.doi.org/10.1016/S1286-4579(01)01375-2
6. Threlfall EJ. Epidemic *Salmonella typhimurium* DT 104—a truly international multiresistant clone. J Antimicrob Chemother. 2000;46:7–10. http://dx.doi.org/10.1093/jac/46.1.7
7. Threlfall EJ, Frost JA, Ward LR, Rowe B. Epidemic in cattle and humans of *Salmonella typhimurium* DT104 with chromosomally integrated multiple drug resistance. Vet Rec. 1994;134:577. http://dx.doi.org/10.1136/vr.134.22.577
8. UK Government. *Salmonella* in livestock production in GB—2014 [cited 2015 Jan 15]. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/468403/pub-salm14-intro.pdf
9. Soyer Y, Switt AM, Davis MA, Maurer J, McDonough PL, Schoonmaker-Bopp DJ, et al. *Salmonella enterica* serotype 4,5,12:i:-, an emerging *Salmonella* serotype that represents multiple distinct clones. J Clin Microbiol. 2009;47:3546–56. http://dx.doi.org/10.1128/JCM.00546-09
10. Laorden L, Herrera-Leon S, Martinez I, Sanchez A, Kromidas L, Bikandi J, et al. Genetic evolution of the Spanish multidrug-resistant *Salmonella enterica* 4,5,12:i:- monophasic variant. J Clin Microbiol. 2010;48:4563–6. http://dx.doi.org/10.1128/JCM.00337-10
11. Trüpschuch S, Laverde Gomez JA, Ediberidze I, Flieger A, Rabsch W. Characterisation of multidrug-resistant *Salmonella* Typhimurium 4,[5],12:i:- DT193 strains carrying a novel genomic island adjacent to the *thrW* tRNA locus. Int J Med Microbiol. 2010;300:279–88. http://dx.doi.org/10.1016/j.ijmm.2010.02.001
12. de la Torre E, Zapata D, Tello M, Mejia W, Frias N, Garcia Pena FJ, et al. Several *Salmonella enterica* subsp. *enterica* serotype 4,5,12:i:- phage types isolated from swine samples originate from serotype Typhimurium DT U302. J Clin Microbiol. 2003;41:2395–400. http://dx.doi.org/10.1128/JCM.41.6.2395-2400.2003
13. Mossong J, Marques P, Ragimbeau C, Huberty-Krau P, Losch S, Meyer G, et al. Outbreaks of monophasic *Salmonella enterica* serovar 4,[5],12:i:- in Luxembourg, 2006. Euro Surveill. 2007;12:719.
14. Hauser E, Tietze E, Helmuth R, Junker E, Blank K, Prager R, et al. Pork contaminated with *Salmonella enterica* serovar 4,[5],12:i:-, an emerging health risk for humans. Appl Environ Microbiol. 2010;76:4601–10. http://dx.doi.org/10.1128/AEM.02991-09
15. Barone L, Dal VA, Pellissier N, Vigano A, Romani C, Pontello M. Emergence of *Salmonella* Typhimurium monophasic serovar: determinants of antimicrobial resistance in porcine and human strains [in Italian]. Ann Ig. 2008;20:199–209.
16. Department for Environment. Food & Rural Affairs. *Salmonella* in livestock production in Great Britain. London: The Department; 2013.
17. Switt AIM, Soyer Y, Warnick LD, Wiedmann M. Emergence, distribution, and molecular and phenotypic characteristics *of Salmonella enterica* serotype 4,5,12:i:-. Foodborne Pathog Dis. 2009;6:407–15. http://dx.doi.org/10.1089/fpd.2008.0213
18. Anderson ES, Ward LR, Saxe MJ, de Sa JD. Bacteriophage-typing designations of *Salmonella typhimurium.* J Hyg (Lond). 1977;78:297–300. http://dx.doi.org/10.1017/S0022172400056187
19. British Society of Antimicrobial Chemotherapy. EUCAST disk diffusion method. 2010 [cited 2015 Jan 15]. http://bsac.org.uk/eucastbsac-disc-diffusion-method/
20. European Food Safety Authority. Scientific opinion on monitoring and assessment of the public health risk of "*Salmonella* Typhimurium-like" strains. EFSA Journal. 2010;8:1826.
21. Swofford DL, Maddison WP. Parsimony, character-state reconstructions, and evolutionary inferences. In: Mayden RL, editor. Systematics, historical ecology, and North American freshwater fishes. Stanford (CA): Stanford University Press; 1992. p. 187–223.
22. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, et al. Roary: rapid large-scale prokaryote pan genome analysis.

Bioinformatics. 2015;31:3691–3. http://dx.doi.org/10.1093/bioinformatics/btv421
23. Wood MW, Rosqvist R, Mullan PB, Edwards MH, Galyov EE. SopE, a secreted protein of *Salmonella dublin*, is translocated into the target eukaryotic cell via a *sip*-dependent mechanism and promotes bacterial entry. Mol Microbiol. 1996;22:327–38. http://dx.doi.org/10.1046/j.1365-2958.1996.00116.x
24. Hardt WD, Chen LM, Schuebel KE, Bustelo XR, Galan JE. *S. typhimurium* encodes an activator of Rho GTPases that induces membrane ruffling and nuclear responses in host cells. Cell. 1998;93:815–26. http://dx.doi.org/10.1016/S0092-8674(00)81442-7
25. Allison GE, Angeles D, Tran-Dinh N, Verma NK. Complete genomic sequence of SfV, a serotype-converting temperate bacteriophage of *Shigella flexneri.* J Bacteriol. 2002;184:1974–87. http://dx.doi.org/10.1128/JB.184.7.1974-1987.2002
26. 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. Nat Genet. 2012;44:1215–21. http://dx.doi.org/10.1038/ng.2423
27. Mather AE, Reid SW, Maskell DJ, Parkhill J, Fookes MC, Harris SR, et al. Distinguishable epidemics of multidrug-resistant *Salmonella* Typhimurium DT104 in different hosts. Science. 2013;341:1514–7. http://dx.doi.org/10.1126/science.1240578
28. Nicholson FA, Chambers BJ, Williams JR, Unwin RJ. Heavy metal contents of livestock feeds and animal manures in England and Wales. Bioresour Technol. 1999;70:23–31. http://dx.doi.org/10.1016/S0960-8524(99)00017-6
29. Slade RD, Kyriazakis I, Carroll SM, Reynolds FH, Wellock IJ, Broom LJ, et al. Effect of rearing environment and dietary zinc oxide on the response of group-housed weaned pigs to enterotoxigenic Escherichia coli O149 challenge. Animal. 2011;5:1170–8. http://dx.doi:10.1017/S1751731111000188
30. Medardus JJ, Molla BZ, Nicol M, Morrow WM, Rajala-Schultz PJ, Kazwala R, et al. In-feed use of heavy metal micronutrients in U.S. swine production systems and its role in persistence of multidrug-resistant salmonellae. Appl Environ Microbiol. 2014;80:2317–25. http://dx.doi.org/10.1128/AEM.04283-13
31. Hughes KT, Roth JR. Directed formation of deletions and duplications using Mud(Ap, lac). Genetics. 1985;109:263–82.
32. Hopkins KL, Threlfall EJ. Frequency and polymorphism of *sopE* in isolates of *Salmonella enterica* belonging to the ten most prevalent serotypes in England and Wales. J Med Microbiol. 2004;53:539–43. http://dx.doi.org/10.1099/jmm.0.05510-0
33. Friebel A, Ilchmann H, Aepfelbacher M, Ehrbar K, Machleidt W, Hardt WD. SopE and SopE2 from *Salmonella typhimurium* activate different sets of RhoGTPases of the host cell. J Biol Chem. 2001;276:34035–40. http://dx.doi.org/10.1074/jbc.M100609200
34. Lopez CA, Winter SE, Rivera-Chavez F, Xavier MN, Poon V, Nuccio SP, et al. Phage-mediated acquisition of a type III secreted effector protein boosts growth of *Salmonella* by nitrate respiration. MBio. 2012;3:pii:e00143-12. http://dx.doi.org/10.1128/mBio.00143-12

Address for correspondence: Robert A. Kingsley, Institute of Food Research, Gut Health and Food Safety, Colney, Norwich NR4 7UA, UK; email: rob.kingsley@ifr.ac.uk, Liljana.Petrovska@apha.gsi.gov.uk



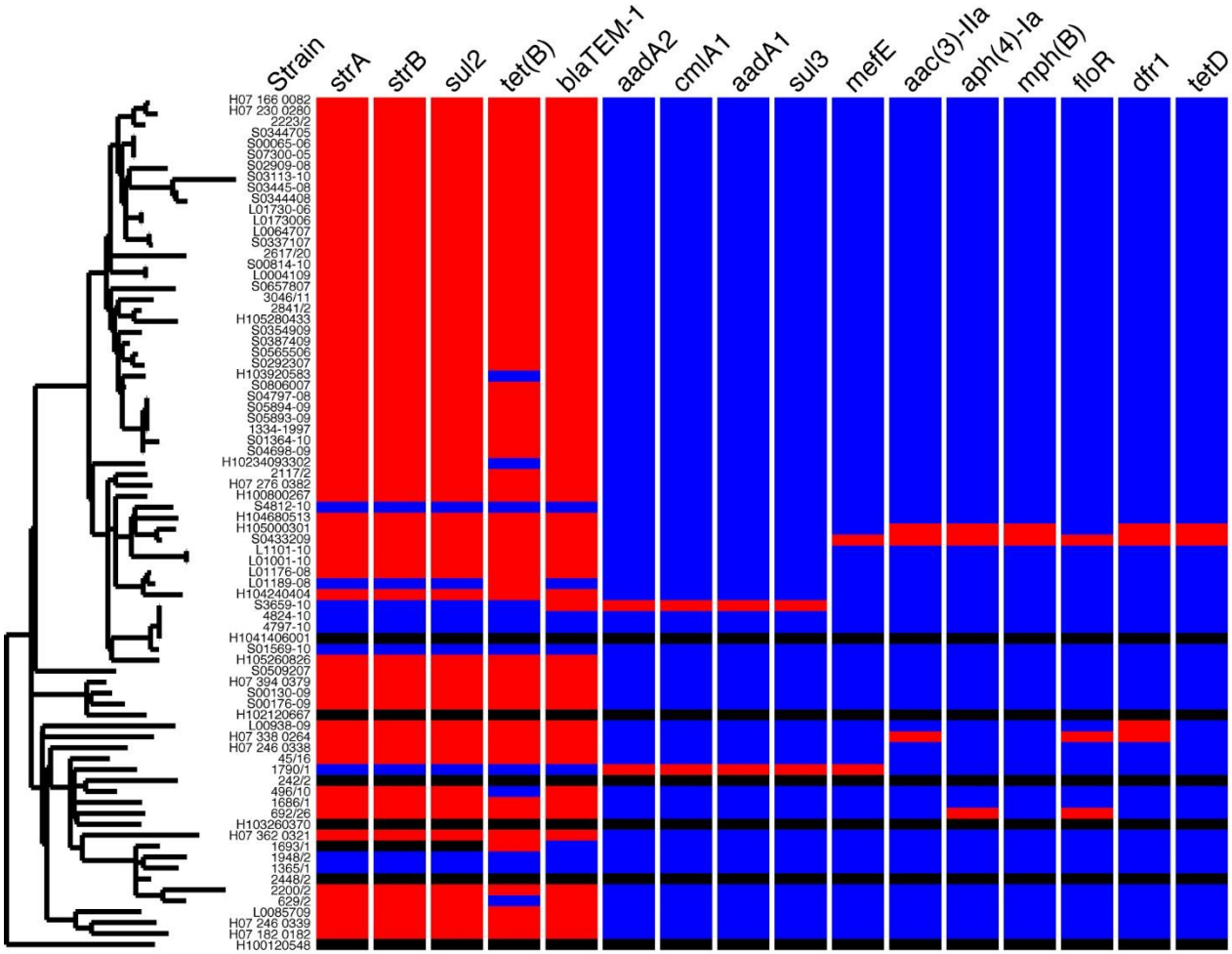
Article DOI: http://dx.doi.org/10.3201/eid2204.150531

Microevolution of Monophasic *Salmonella* Typhimurium during Epidemic, United Kingdom

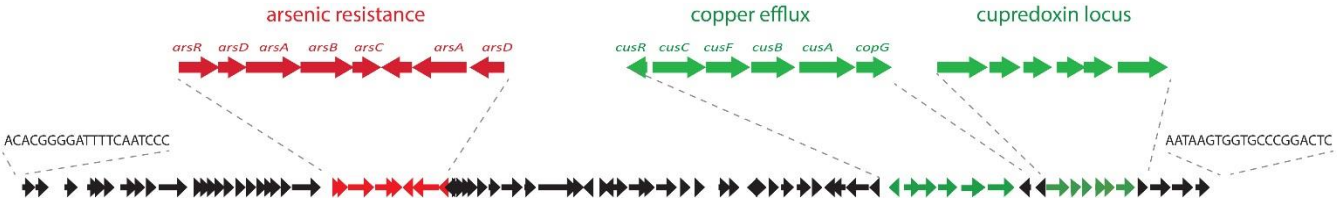
Technical Appendix 2

Technical Appendix 2 Table. Additional strains and metadata, strains used to determine *sopE* frequency

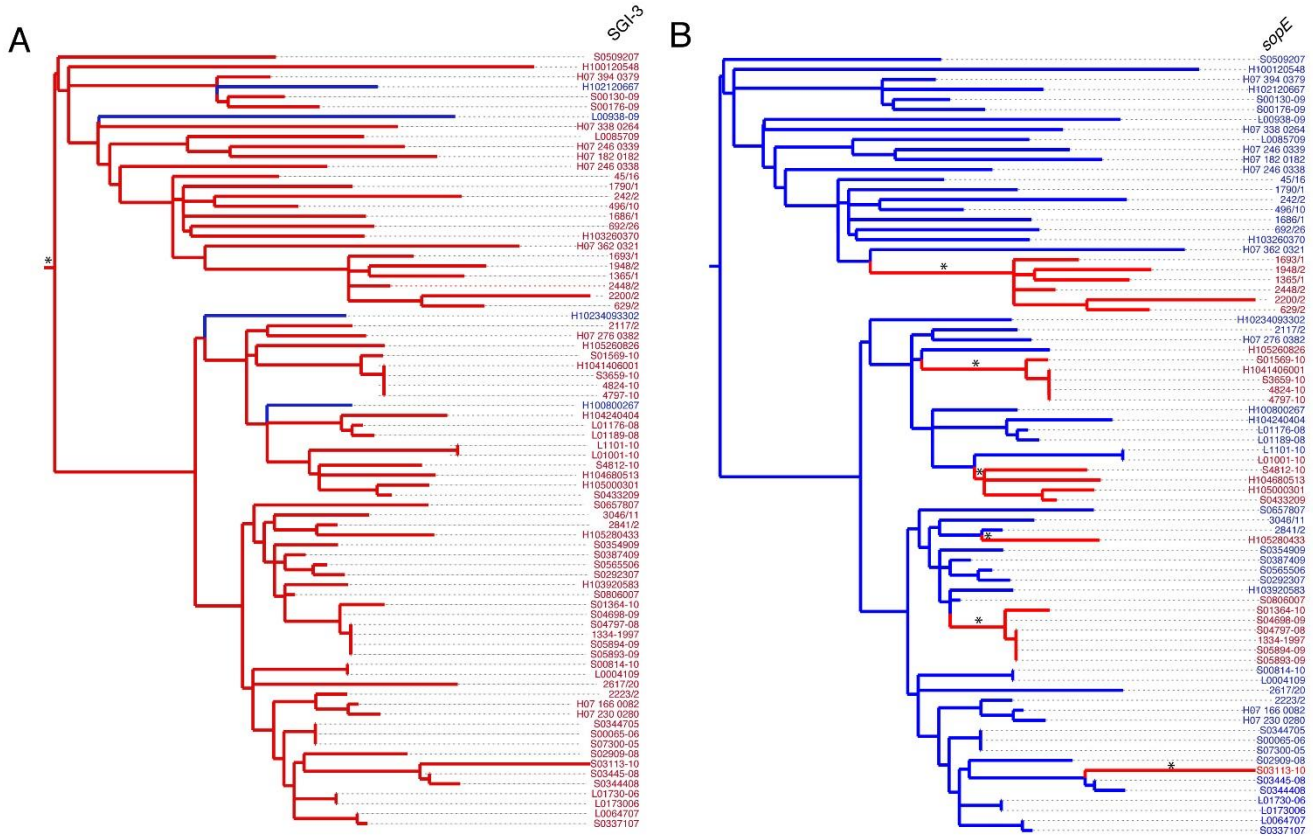
Year of isolation	Sample name	Species	Serotype	Phage type	Resistance pattern	*SopE*, by PCR
2005	S04612-05	ENVIRONMENTAL_FARM	4,12:I:-	NOPT	ASSuT	Neg
2005	S07300-05	CATTLE	4,12:I:-	NOPT	ND	Neg
2006	S06813-06	PIG	TYPHIMURIUM	DT 193	ASSuT	Neg
2006	S00065-06	CATTLE	4,12:I:-	NOPT	ND	Neg
2007	L00745-07	PIG	4,5,12:I:-	NOPT	TASxtSSU	Pos
2007	S05720-07	CATTLE	TYPHIMURIUM	DT 193	ASSuT	Neg
2007	L00653-07	PIG	4,12:I:-	NOPT	ASSuT	Neg
2007	S00250-07	PIG	4,12:I:-	ND	Sensitive	Neg
2007	S04962-07	CATTLE	4,12:I:-	NOPT	ASSuT	Neg
2007	L00650-07	PIG	4,5,12:I:-	NOPT	T	Neg
2007	S06676-07	PIG	4,5,12:I:-	NOPT	ASSuT	Neg
2008	L00555-08	DUCK	4,12:I:-	ND	T	Neg
2008	L01189-08	PIG	4,12:I:-	ND	T	Neg
2008	S05635-08	OTHER_VEG_MINERAL	4,12:I:-	ND	ASSuT	Neg
2008	S06669-08	DOG	4,12:I:-	ND	ASSuT	Neg
2008	S06718-08	CATTLE	4,12:I:-	ND	ASSuT	Neg
2009	S04409-09	PIG	4,12:I:-	DT 193	ASSuT	Pos
2009	LO506-09	PIG	4,5,12:i:-	DT 193	ASSuNaCnApr	Pos
2009	L00028-09	PIG	4,12:I:-	DT 193	ASSuT	Neg
2009	L00041-09	DOG	4,12:I:-	DT 193	ASSuT	Neg
2009	L00300-09	DOG	4,12:I:-	DT 193	TA	Neg
2009	L00663-09	MOUSE	4,12:I:-	DT 191a	ASSuT	Neg
2009	S00176-09	PIG	4,12:I:-	RDNC	ASSuT	Neg
2009	S00428-09	DOG	4,12:I:-	DT 193	ASSuT	Neg
2009	S04117-09	SHEEP	4,12:I:-	DT 193	ASSu	Neg
2009	S04700-09	CAT	4,12:I:-	DT 193	TASxtSSu	Neg
2009	S04711-09	DOG	4,12:I:-	DT 193	ASSuT	Neg
2010	S02497-10	EQ_HORSE	4,12:I:-	DT 193	ASSuT	Pos
2010	S03660-10	CHICKEN	4,5,12:I:-	ND	Sensitive	Pos
2010	S00028-10	CATTLE	4,12:I:-	DT 193	ASSuT	Neg
2010	S00474-10	CAT	4,12:I:-	DT 193	ASSuT	Neg
2010	S00814-10	CHICKEN	4,12:I:-	DT 193	ASSuT	Neg
2010	S01299-10	PIG	4,12:I:-	DT 193	ASSuT	Neg
2010	S01332-10	CAT	4,12:I:-	DT 193	ASSuT	Neg
2010	S01585-10	OTHER_VEG_MINERAL	4,12:I:-	DT 193	TNASxtSSu	Neg
2010	S01764-10	CHICKEN	4,12:I:-	DT 193	ASSuT	Neg
2010	S03060-10	CATTLE	4,12:I:-	DT 193	T	Neg
2010	L00809-10	PIG	4,5,12:I:-	DT 193	ASSuT	Neg
2010	L01001-10	CHICKEN	4,5,12:I:-	DT 120	ASSuT	Neg
2010	S00771-10	PIG	4,5,12:I:-	DT 193	TNASxtCCnSSuApr	Neg



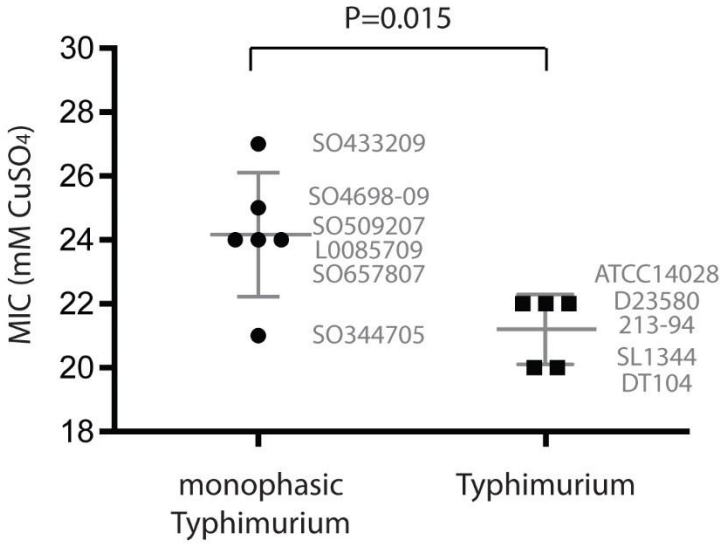
Technical Appendix 2 Figure 1. Presence of antibiotic resistance genes in the monophasic *Salmonella* Typhimurium epidemic strains from the UK. The presence (red) or absence (blue) of antibiotic resistance genes are shown in the context of the maximum likelihood tree described in Figure 2 in the main text. Some data were unavailable due to poor quality sequence assembly (black).



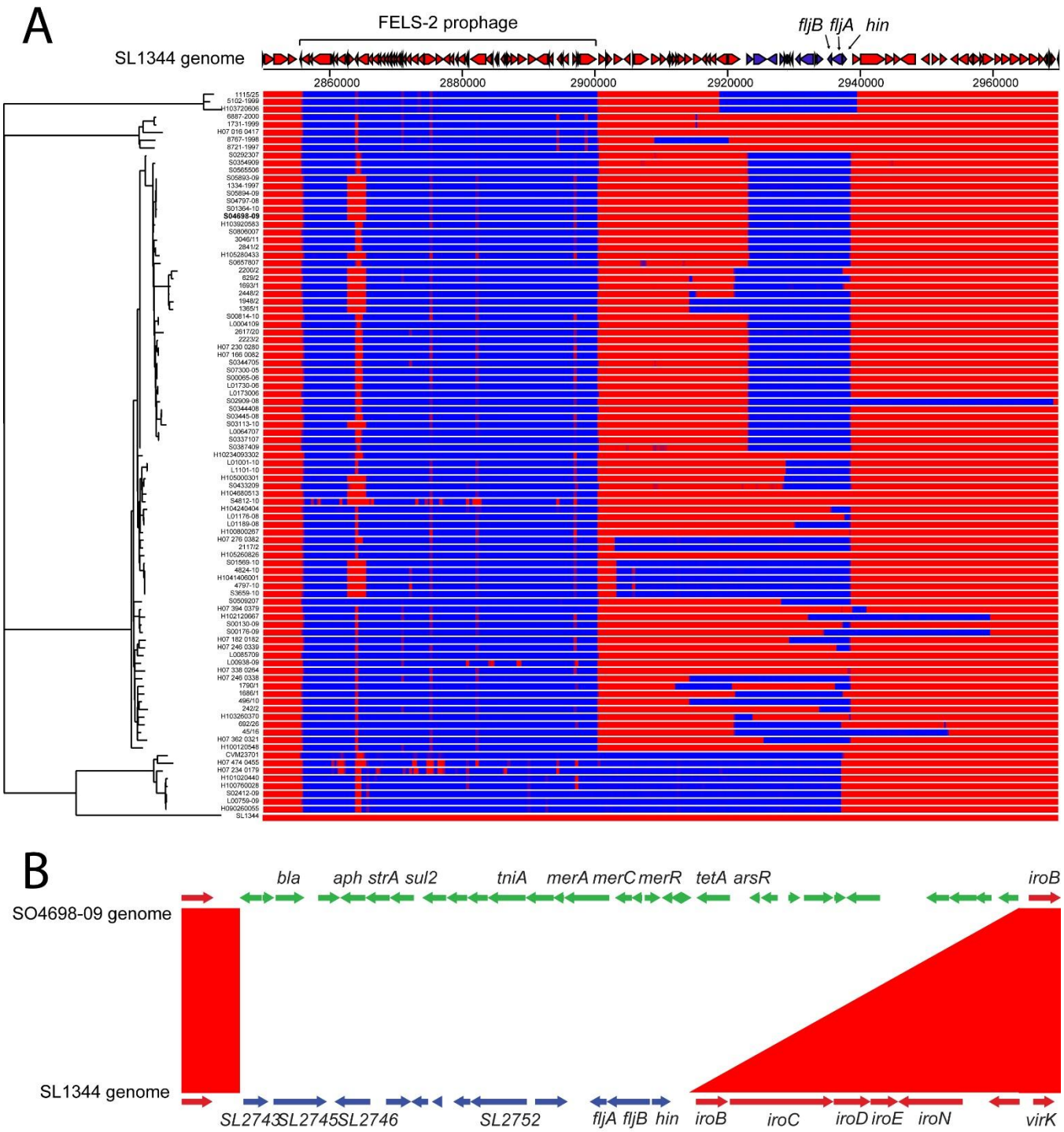
Technical Appendix 2 Figure 2. Gene arrangement of the novel genomic island of *Salmonella* 1,4,[5],12:i:- strain SO4698–09. Arrows indicate predicted genes within the island. The position of genes with predicted functions by sequence comparison are indicated for arsenic resistance (red), cadmium, zinc and copper resistance (green). The nucleotide sequence flanking the insertion in the whole genome sequence of SO4698–09 (PRJEB10340) is indicated.



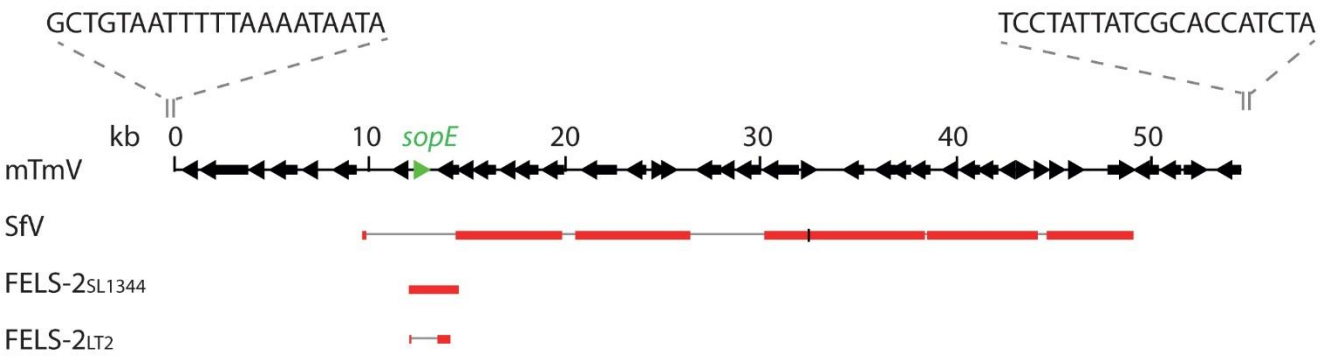
Technical Appendix 2 Figure 3. Ancestral state reconstruction of SGI-3 and *sopE* gene within the monophasic epidemic clade. Maximum likelihood trees for 77 UK and Italy monophasic isolates as previously described in Figure 2 in the main text. Ancestral state for presence (red edges) or absence (blue edges) of SGI-3 (A) or *sopE* (B) were reconstructed based on maximum parsimony using ACCTRAN. * indicate the inferred acquisition of the genetic element.



Technical Appendix 2 Figure 4. MIC of monophasic *Salmonella* Typhimurium and *Salmonella* Typhimurium isolates to copper sulfate in rich broth culture. The ability of monophasic *Salmonella* Typhimurium (filled circles) or *Salmonella* Typhimurium (filled squares) isolates to grow in Luria Bertani broth in the presence of copper sulfate (pH7) were monitored by the optical density of culture. The MIC was defined as the concentration at which cultures attained at least OD_{600nm} of 0.1. The mean for each phylogenetic group (gray bar) +/− standard deviation are indicated. Student's t test was used to test significance.



Technical Appendix 2 Figure 5. Heat map showing deletions around the *fljB* locus of the *Salmonella* 1,4,[5],12:i:- epidemic clade isolates. The heat map (A) indicating mapped sequence read coverage for *Salmonella* 1,4,[5],12:i:- epidemic clade isolates to the *fljB* locus and flanking sequence of the whole genome sequence of *Salmonella* Typhimurium strain SL1344. Color indicates 0 mapped reads (blue) to ≥20 bases (red). Filled arrows indicate genes in the SL1344 genome sequence as described previously (*1*). A maximum likelihood tree of phenotypically monophasic isolates from the strain collection is shown.



Technical Appendix 2 Figure 6. Prophage element mTmV from strain SO4698–09 and BLAST results with SfV and FELS-2 prophage. Predicted open reading frames in the 55 kb mTmV prophage of strain SO4698 are shown with flanking nucleotide sequence for orientation. Regions with significant BLAST results (red bar) in the related prophage SfV prophage and FELS-2 prophages are indicated below.

Supplementary Materials and Methods

Selection of *S*. Typhimurium Isolates.

Animal, human and environmental isolates were from the Animal and Plant Health Agency (APHA) *Salmonella* archive, the Public Health England (PHE) strain collection or the Italian Reference Laboratory for *Salmonella* (NRL-IZSVe, Legnaro, Italy). The isolates were selected in order to represent a) *S.* 1, 4,[5],12:i:- strains from the UK from the initial period of the current epidemic (2005 to 2010) from pigs, cattle, poultry, wild avian, sheep, horse and companion animals, b) *S.* 1, 4,[5],12:i:- strains from the UK isolated from before the current epidemic (1994-2005), c) *S.* Typhimurium strains with phage types DT193 and DT120, d) *S.* 1, 4,[5],12:i:- isolates from Italy, and e) *S.* Typhimurium from of diverse phage types isolated from animals and humans over the past 20 years. This strain collection was therefore captures the diversity of *S.* 1, 4,[5],12:i:- and *S.* Typhimurium isolates and is not representative of the epidemiology from the UK and Italy. These strains are described in online Technical Appendix 1 (http://wwwnc.cdc.gov/EID/article/22/4/15-0531-Techapp1.xlsx), and the whole genome sequence of these strains was determined by Illumina HiSeq. An additional 40 *S.* 1, 4,[5],12:i:- strains from the UK were randomly selected from the APHA strain collection (described in online Technical Appendix 2 Table) in order to further investigate the frequency of the presence of the *sopE* gene using PCR amplification (Figure 3 in the main text).

Concentrations for determination of minimal inhibitory concentrations.

Determination of antimicrobial sensitivity of animals isolates from the UK (APHA) and isolates from Italy (NRL-IZSVe) were tested for susceptibility to the following Antimicrobials amikacin 30 µg, amoxicillin/clavulanate 30 µg, ampicillin 10 µg, apramycin 15 µg, cefotaxime 30 µg, ceftazidime 30 µg,

chloramphenicol 30 µg, ciprofloxacin 1 µg, furazolidone 15 µg, gentamicin 10 µg, nalidixic acid 30 µg, neomycin 10 µg, streptomycin 10 µg, sulphonamide compounds 300 µg, trimethoprim/sulfamethoxazole 25 µg, and tetracycline 10 µg. Isolates were interpreted as resistant or susceptible on the basis of BSAC breakpoints. Where there is an intermediate BSAC category this is reported here as resistant. Antimicrobial sensitivity of human isolates from the UK (PHE) was determined using a modified breakpoint technique on Isosensitest agar (Oxoid, Basingstoke, UK). The final plate concentrations (µg/mL) used routinely by the HPA were: ampicillin (A; 8), chloramphenicol (C; 8), gentamicin (G; 4), kanamycin (K; 16), neomycin (Ne; 8), streptomycin (S; 16), sulphonamides (Su; 64), tetracycline (T; 8), trimethoprim (Tm; 2), furazolidone (Fu; 8), nalidixic acid (Nx; 16), ciprofloxacin (low-level (Cpl); 0.125); (high-level (Cp); 1)), amikacin (Ak; 4), cephalexin (Cx; 16), cephradine (Cr; 16), cefuroxime (Cf; 16), ceftriaxone (Cn; 1) and cefotaxime (Ct; 1).

Whole genome sequencing and analysis.

For Illumina HiSeq, libraries with 300 base pair (bp) insert size were constructed and 100 bp paired-end reads were generated using standard methodologies. Sequence data were submitted to the European Nucleotide Archive (http://www.ebi.ac.uk/ena) (online Technical Appendix 1). For sequence-read alignment and single nucleotide polymorphism (SNP) detection, paired-end Illumina sequence data were mapped to the reference genome *S*. Typhimurium strain SL1344 (*1*), or draft assembly of monophasic Typhimurium strain SO4698-09, using SMALT (ftp://ftp.sanger.ac.uk/pub4/resources/software/smalt/smalt-manual-0.7.4.pdf). SNPs were identified using samtools mpileup and filtered with a minimum mapping quality of 30 and quality ratio cut-off of 0.75. SNPs identified in prophage elements and repetitive sequence regions of the *S*. Typhimurium reference were excluded from subsequent phylogenetic analysis as previously described (*2*). For phylogenetic analyses, a maximum-likelihood phylogenetic tree was constructed from the SNP alignment with RAxML v7.0.4 (*3*) using a general time reversible (GTR) substitution model with gamma correction for among-site rate variation. Support for nodes on the trees were assessed using 100 bootstrap replicates. For parsimonious ancestral state inference, accelerated transformation (ACCTRAN) was performed on ML trees (*4*). The raw Illumina HiSeq data were used to generate a *de novo* draft assembly of the genome using VELVET v0.7.03 (*5*). Annotation of draft genomes was performed using RAST (*6*). In order to determine the presence or absence of sequence in multiple isolates without the need for assembly and genome alignment we mapped raw Illumina HiSeq reads to a region of the genome of interest using SMALT and generated heat maps that represented the relative mapped sequence read depth at each nucleotide locus of the reference sequence. The presence of the *sopE* gene using whole genome sequence de novo assembly of monophasic strains (Supplementary

Information) or by PCR amplification of genomic DNA using primers CAGTTGGAATTGCTGTGG and GCTTCAAACGCTCAATGATATAG prepared from additional isolates for which whole genome sequence was not determined.

Pangenome analysis.

The presence or absence of every gene was determined by constructing a pan genome (http://sanger-pathogens.github.io/Bio-PanGenome). From this the sequences of the *sopE* gene were extracted from each isolate (if present). The whole genome de novo assemblies were first annotated with Prokka (*7*). Predicted coding regions were extracted and converted to protein sequences. CD-hit (v4.6.1) was used to iteratively perform a first pass clustering to identify near perfect sequence matches (*8*). Beginning with a sequence identity of 100% and a matching length of 100%, the protein sequences were clustered. If a sequence was found in every isolate, it was considered a conserved gene and the cluster added to the final results. All of these sequences were then removed and not considered for blast analysis. CD-hit analysis was repeated again with a lower threshold, reducing by 0.5% down to 98%, with conserved clusters removed at each stage. One final clustering step was performed with CD-hit, with a sequence identity of 100% leaving one representative sequence for each cluster in a protein FASTA file. A blast database was created from this FASTA file. Low complexity regions were masked out with SegMasker [ncbi_blast_plus], and a protein blast database was created with makeblastdb [ncbi_blast_plus]. Segments of the FASTA file was compared to the blast database to perform an all against all blast (v2.2.28). The combined blast results were then processed by MCL (v263) that clustered the input sequences (*9*). A normalised bit score was used (bit scores normalized by length of the HSP). The clusters were then re-inflated with the final CD-hit clustering results and with the iterative CD-hit conserved clusters. The clusters were then labeled with the gene names transferred from the input annotation. The functional annotation was also recorded for each cluster. A multi-FASTA file of nucleotides was created for each gene with the corresponding sequence from each isolate in which the gene was present. The sequences were translated into proteins using fastatranslate (v2.2.0) (*9*), aligned using muscle (v3.8.31) (*10*), then reverse translated to nucleotide sequences using RevTrans (v1.4), to give an aligned multi-FASTA file (*11*).

For determination of whole genome sequence of S. Typhimurium SO4698-09 for *de novo* assembly using SMRT PacBio®, Template Prep Kit (PacBio, Menlo Park, CA, USA) and BluePippin™ Size Selection System protocol were employed to prepare size-selected libraries (20kb) from 5 µg of sheared and concentrated DNA. Sequencing was performed using the magnetic bead collection protocol, a 20,000 bp insert size, stage start, and 180-minute movies. A *de novo* assembly was generated from

these reads using the Hierarchical Genome Assembly Process (HGAP) software version 3.0, with the genome size parameter set to 5 Mb. The raw read data and assemblies are submitted to the ENA database with accession number PRJEB10340.

Determination of genotypic variation by PCR amplification and plasmid profiling.

The occupation of the *thrW* locus (Figure 2 in the main text) was determined by PCR amplification of genomic DNA and determination of amplicon size by gel electrophoresis as previously described (*12*). Determination of the presence of the sopE gene in isolates for which whole genome sequence was not available was determined by PCR amplification of genomic DNA using primers 5' CAGTTGGAATTGCTGTGG 3' and 5' GCTTCAAACGCTCAATGATATAG 3' that amplifies a 417 bp region within the gene. The identity of the PCR product was determined by agarose gel electrophoresis. Determination of the presence of the pSLT plasmid was determined by preparation of large molecular weight plasmid DNA from isolates cultured in LB broth and separation by agarose gel electrophoresis as previously described (*13*).

References

1. Kröger C, Dillon SC, Cameron AD, Papenfort K, Sivasankaran SK, Hokamp K, et al. The transcriptional landscape and small RNAs of Salmonella enterica serovar Typhimurium. Proc Natl Acad Sci U S A. 2012;109:E1277–86. PubMed http://dx.doi.org/10.1073/pnas.1201061109

2. 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. Nat Genet. 2012;44:1215–21. PubMed http://dx.doi.org/10.1038/ng.2423

3. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006;22:2688–90. PubMed http://dx.doi.org/10.1093/bioinformatics/btl446

4. Swofford DL, Maddison WP. Reconstructing Ancestral Character States under Wagner Parsimony. Math Biosci. 1987;87:199–229. http://dx.doi.org/10.1016/0025-5564(87)90074-5

5. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008;18:821–9. PubMed http://dx.doi.org/10.1101/gr.074492.107

6. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 2014;42:D206–14. PubMed http://dx.doi.org/10.1093/nar/gkt1226

7. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–9. PubMed http://dx.doi.org/10.1093/bioinformatics/btu153

8. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28:3150–2. PubMed http://dx.doi.org/10.1093/bioinformatics/bts565

9. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 2002;30:1575–84. PubMed http://dx.doi.org/10.1093/nar/30.7.1575

10. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004;5:113. PubMed http://dx.doi.org/10.1186/1471-2105-5-113

11. Wernersson R, Pedersen AG. RevTrans: Multiple alignment of coding DNA from aligned amino acid sequences. Nucleic Acids Res. 2003;31:3537–9. PubMed http://dx.doi.org/10.1093/nar/gkg609

12. Trüpschuch S, Laverde Gomez JA, Ediberidze I, Flieger A, Rabsch W. Characterisation of multidrug-resistant Salmonella Typhimurium 4,[5],12:i:- DT193 strains carrying a novel genomic island adjacent to the thrW tRNA locus. International journal of medical microbiology. Int J Med Microbiol. 2010;300:279–88. PubMed http://dx.doi.org/10.1016/j.ijmm.2010.02.001

13. Kado CI, Liu ST. Rapid procedure for detection and isolation of large and small plasmids. J Bacteriol. 1981;145:1365–73. PubMed