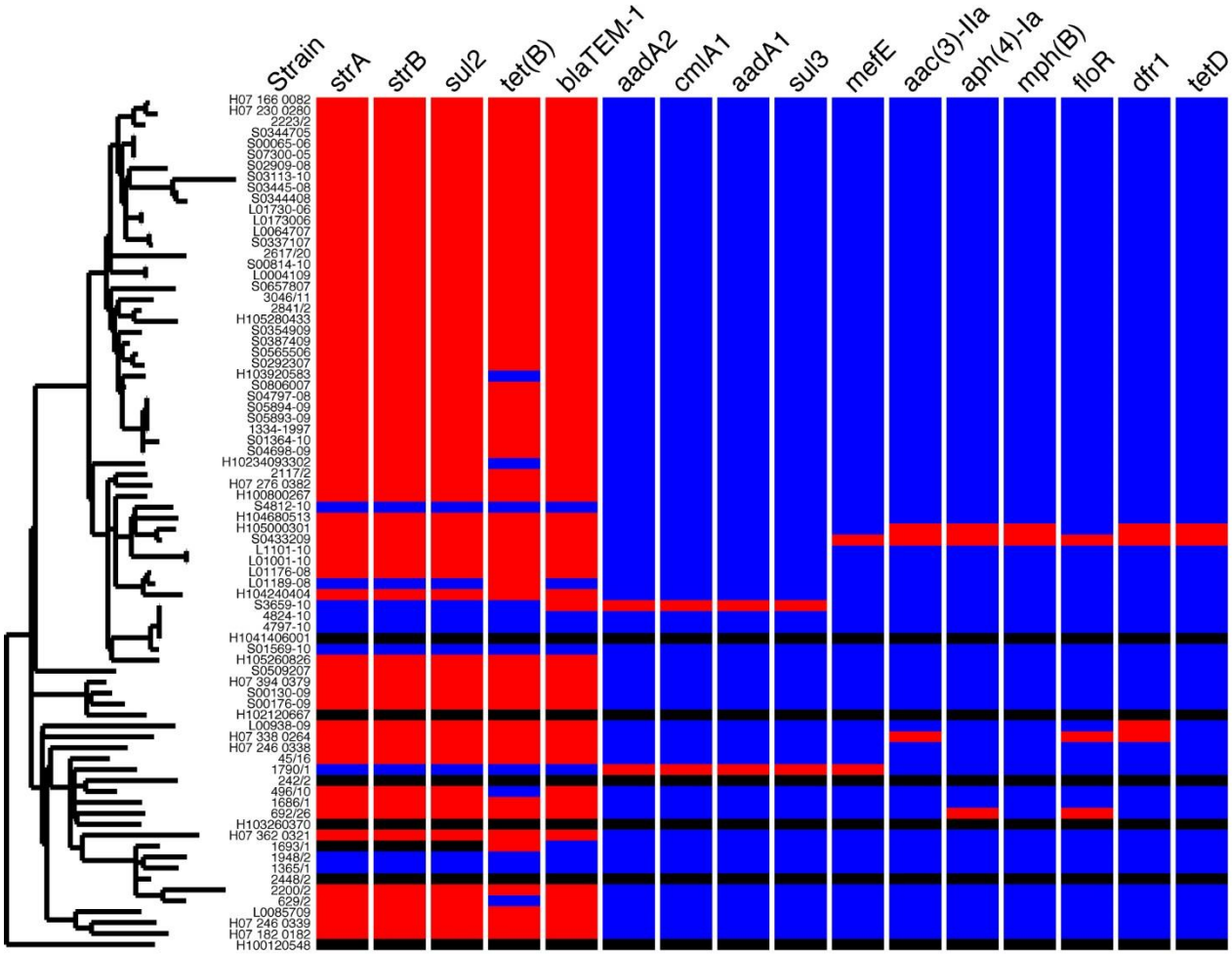
Article DOI: http://dx.doi.org/10.3201/eid2204.150531

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

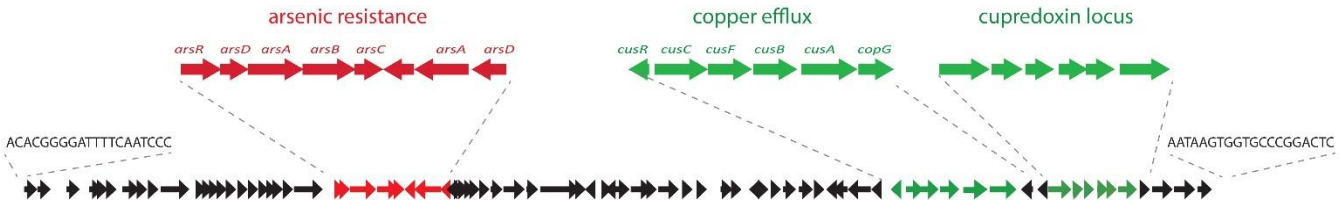
Technical Appendix 2

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

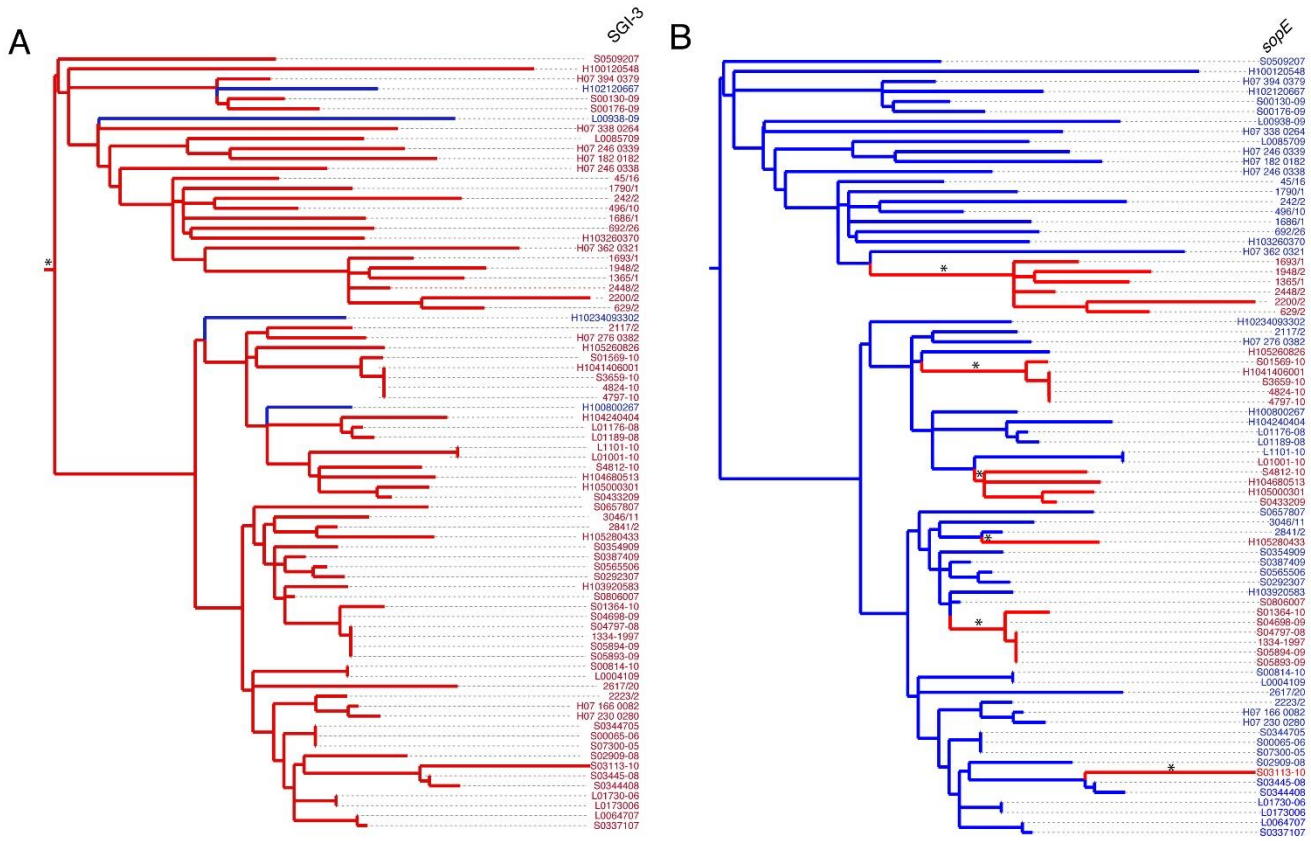
Year of isolation	Sample name	Species	Serotype	Phage type	Resistance pattern	*SopE*, by PCR
2005	S04612-05	ENVIRONMENTAL_FARM	4,12:I:-	NOPT	ASSuT	Neg
2005	S07300-05	CATTLE	4,12:I:-	NOPT	ND	Neg
2006	S06813-06	PIG	TYPHIMURIUM	DT 193	ASSuT	Neg
2006	S00065-06	CATTLE	4,12:I:-	NOPT	ND	Neg
2007	L00745-07	PIG	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



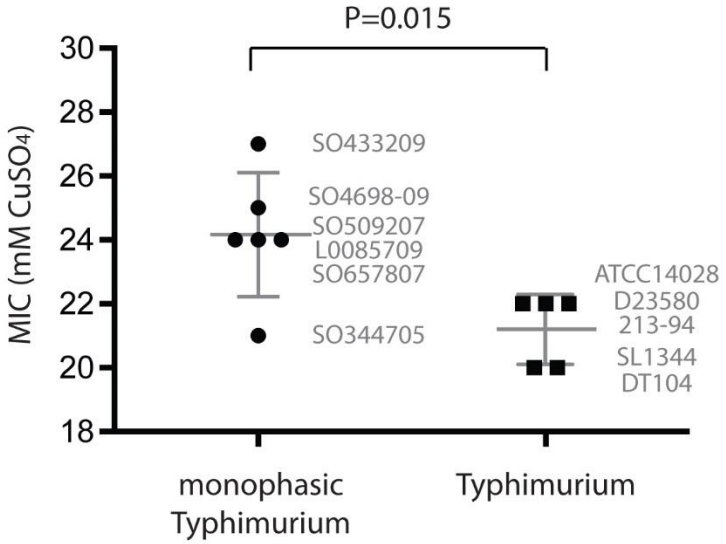
Technical Appendix 2 Figure 1. Presence of antibiotic resistance genes in the monophasic *Salmonella* Typhimurium epidemic strains from the UK. The presence (red) or absence (blue) of antibiotic resistance genes are shown in the context of the maximum likelihood tree described in Figure 2 in the main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



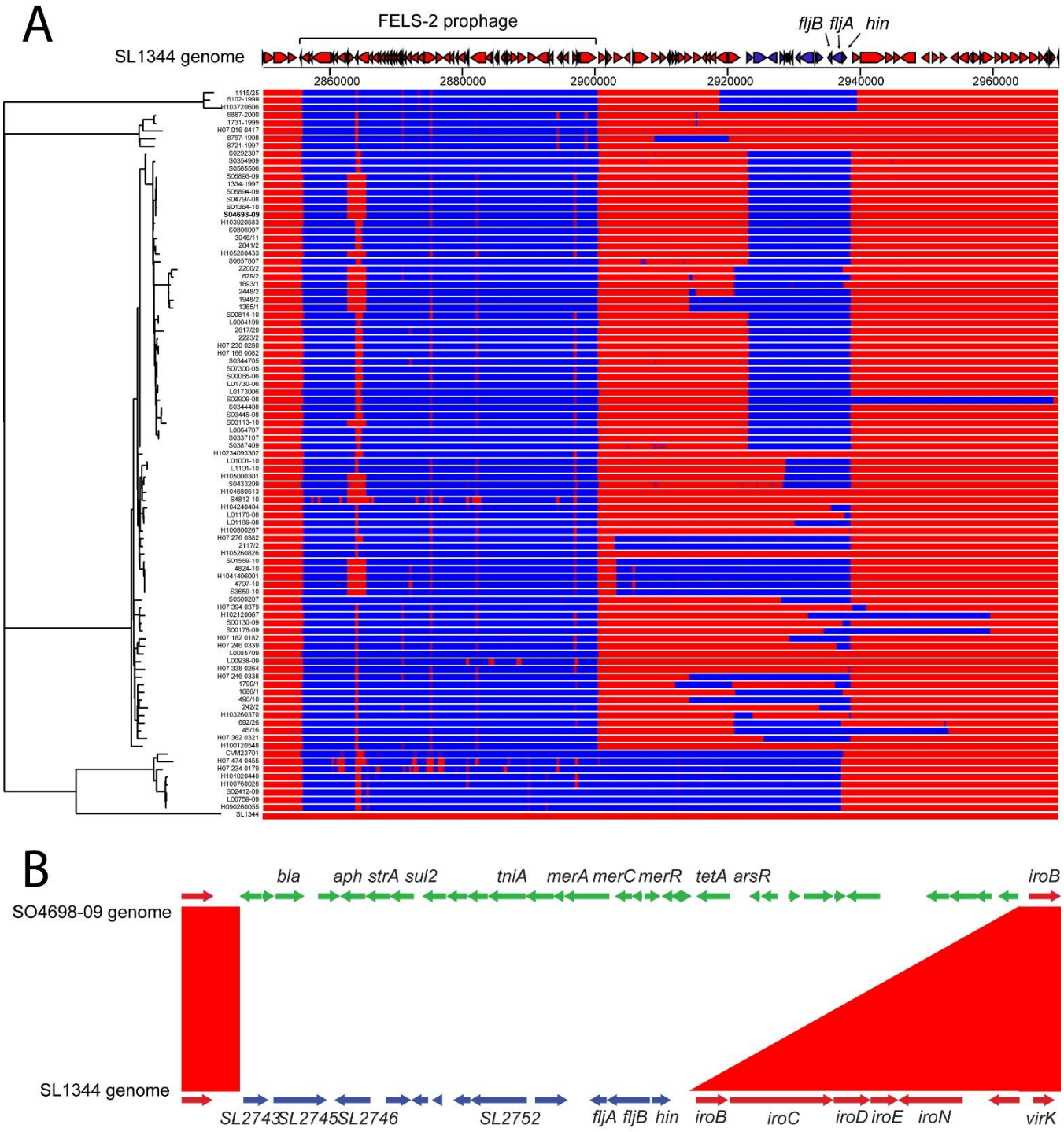
Technical Appendix 2 Figure 2. Gene arrangement of the novel genomic island of *Salmonella* 1,4,[5],12:i:- strain SO4698–09. Arrows indicate predicted genes within the island. The position of genes with predicted functions by sequence comparison are indicated for arsenic resistance (red), cadmium, zinc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



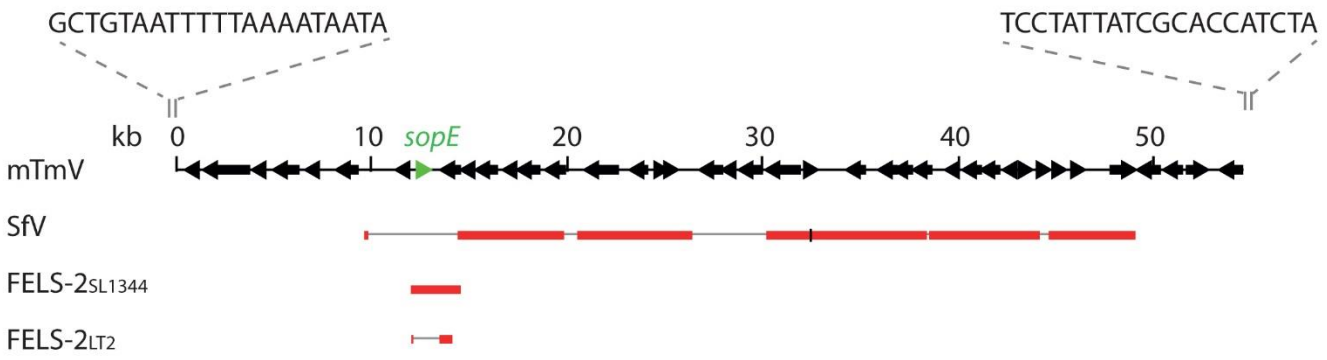
Technical Appendix 2 Figure 3. Ancestral state reconstruction of SGI-3 and *sopE* gene within the monophasic epidemic clade. Maximum likelihood trees for 77 UK and Italy monophasic isolates as previously described in Figure 2 in the main text. Ancestral state for presence (red edges) or absence (blue 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