**Supplementary Materials:**



# Supplementary Materials for

# **Distinguishable Epidemics Within Different Hosts of the Multidrug Resistant Zoonotic Pathogen** *Salmonella* **Typhimurium DT104**

A.E. Mather, S.W.J. Reid, D.J. Maskell, J. Parkhill, M.C. Fookes, S.R. Harris, D.J. Brown, J.E. Coia, M.R. Mulvey, M.W. Gilmour, L. Petrovska, E. de Pinna, M. Kuroda, M. Akiba, H. Izumiya, T.R. Connor, M.A. Suchard, P. Lemey, D.J. Mellor, D.T. Haydon, N.R. Thomson correspondence to: swjreid@rvc.ac.uk, nrt@sanger.ac.uk

#### **This PDF file includes:**

Materials and Methods Figs. S1 to S7 Tables S1 to S12

#### **Materials**

#### Data

A description of the number, country of origin, and purpose for selection of all isolates used in this study is presented in Table S1. The majority of animal isolates (70%) are of bovine origin (Tables S2 – S5), which reflects the primary animal reservoir (*20, 31*). Sequence data are deposited in the European Nucleotide Archive, under study accession numbers ERP000244, ERP000270, and ERP000994. Sequence data of the Japanese strains can be obtained from the DNA Data Bank of Japan, accession number DRA000942.

#### Whole genome sequencing

All isolates were sequenced using multiplex libraries on the Illumina HiSeq platform using 100 bp paired end reads (Table S1), unless otherwise stated. To create a high quality DT104 reference sequence, *S*. Typhimurium DT104 genomic DNA was fragmented by sonication, and several libraries were generated in pUC18 using size fractions ranging from 1.0 to 2.5 kb. The high quality finished DT104 genome was sequenced to a depth of 9x coverage from M13mp18 (insert size 1.4–2 kb) and pUC18 (insert size 2.2–4.2 kb) small insert libraries, using dye terminator chemistry on ABI3700 automated sequencers. End sequences from larger insert plasmid (pBACe3.6, 12–30 kb insert size) libraries were used as a scaffold. The sequence was assembled, finished, and annotated as described previously (*32*). The finished chromosome and plasmid sequences have been submitted to the European Molecular Biology Laboratory (accession numbers HF937208 and HF937209, respectively).

#### Scottish *S.* Typhimurium DT104

The surveillance programme that generated the Scottish animal and human *Salmonella* Typhimurium DT104 (hereafter, DT104) data used in this study is described in Mather *et al* (*11*). *Salmonella* is a reportable human and livestock pathogen in the UK, and all suspected *Salmonella* isolates identified at medical and veterinary diagnostic laboratories in Scotland are forwarded to the Scottish *Salmonella Shigella* and *Clostridium difficile* Reference Laboratory (SSSCDRL) for confirmation and typing. Both human and animal DT104 isolates were subject to the same microbiological and typing procedures. Serotyping of the isolates and phage typing was accomplished according to internationally standardized methods (*33-35*). Antimicrobial susceptibility was assessed using a modified breakpoint method, involving solid agar plates containing a pre-determined concentration of antimicrobial (Table S7), and isolates were classified as non-resistant or resistant (*36*). There were four sets of isolates selected from the collection held at the SSSCDRL. The total number of isolates submitted to the SSSCDRL over the period 1990 – 2004 (figure from the supplementary material of  $(11)$ ), and the number of isolates sequenced per year in this study are represented in Figure S7, demonstrating the coverage of the epidemic represented by the sequenced isolates.

1. Scottish domestically-acquired DT104 isolates – diversity of antimicrobial resistance As described in Mather *et al* (*11*), 2,439 animal isolates and 2,761 human isolates were collected over the epidemic period 1990 – 2004. Phenotypically, there were 52 profiles identified in human DT104 isolates during this time period, and 35 profiles in animal isolates. Twenty-two of these profiles were held in common by both animals and humans; overall, there were 65 unique

profiles (*11*). Human isolates were derived from domestically-acquired infections of DT104, from cases with no history of recent foreign travel.

We selected a subset of 156 of these DT104 isolates for sequencing. The selection process was the same for both human and animal isolates, as follows: All isolates with phenotypic AMR profiles observed only once or twice during the study period were selected for sequencing. For profiles observed three to nine times, two random isolates for each of these profiles were selected for sequencing. For each of the profiles comprised of ten or more isolates, with the exception of the most prevalent profile, three random isolates were selected for each. For the most prevalent profile, demonstrating phenotypic resistance to ampicillin, chloramphenicol, spectinomycin, streptomycin, sulphonamides, and tetracycline (ApClSpStSuTe), nine isolates were randomly selected. For any of the selected isolates which on retrieval from storage proved to be non-viable, additional DT104 isolates were randomly selected which were of the same profile; if no further isolates of the same profile were available, additional DT104 isolates from the most numerous profiles (ApClSpStSuTe and ApClSpStSuTe+trimethoprim) were randomly selected. A list of these isolates is presented in Table S2. All isolates of *Salmonella* at the SSSCDRL, once characterized, were inoculated on Dorset egg slopes for long-term storage. The isolates selected for sequencing were plated onto cysteine lactose electrolyte-deficient (CLED) agar and incubated overnight at 37ºC. A single colony from each culture was subcultured separately into 5 mL Brain Heart Infusion (BHI) broth, and incubated overnight at 37ºC. DNA was extracted using the Puregene Core Kit B (Qiagen). Nine (9) isolates were removed from further analysis following sequencing due to sample contamination or poor sequence quality.

#### 2. Scottish domestically-acquired DT104 isolates – diversity within the main resistance profile: ApClSpStSuTe

To assess the diversity within isolates demonstrating the main resistance pattern, conferring resistance to ampicillin, chloramphenicol, spectinomycin, streptomycin, sulphonamides, and tetracycline, an additional 47 animal isolates and 47 human isolates were selected for sequencing. These were selected stratified by year so that, along with the ApClSpStSuTe isolates selected in the first round of selection, the number of these isolates was proportional to the number of isolates demonstrating the ApClSpStSuTe profile that were submitted in each year from each host population (Fig. S7). DNA was extracted as described above, and isolates sequenced (see Table S3). One human isolate was subsequently found to be contaminated and so was removed from all further analysis.

#### 3. Scottish domestically-acquired DT104 isolates – post-epidemic

Twenty-four DT104 isolates (12 from animals, 12 from humans) were randomly selected from the post-epidemic period 2005 – 2011. DNA was extracted as described above, and isolates sequenced (Table S3). Two human isolates were subsequently found to be contaminated or had poor sequence quality and so were removed from all further analysis.

#### 4. Travel-associated DT104 isolates reported to the SSSCDRL

Over the period 1990 – 2004, there were 135 reported human infections of DT104 from patients with a recent history of foreign travel. To assess how these isolates fit within the DT104 phylogeny generated with the Scottish isolates, 28 isolates were selected across the diversity of countries that the patients reported visiting (Table S3). DNA was extracted as described above,

and isolates were sequenced. One isolate was subsequently found to be contaminated and so was removed from all further analysis.

#### Japanese *S.* Typhimurium DT104

To provide context to the Scottish DT104 isolates, five human and five animal DT104 isolates were sequenced at the Laboratory of Bacterial Genomics, Pathogen Genomics Center, at the National Institute of Infectious Diseases in Tokyo, Japan. The Illumina GAIIx machine, with 81 base paired end reads, was used to obtain whole genome sequences. These isolates were from the period 1994 – 2012. Phenotypic susceptibility was assessed through disc diffusion, using Clinical and Laboratory Standards Institute (CLSI) criteria and breakpoints (*37, 38*). Resistance to ampicillin, chloramphenicol, streptomycin, sulphonamides, tetracycline, nalidixic acid, ciprofloxacin, kanamycin, cefotaxime, trimethoprim/sulphamethoxazole, gentamicin and fosfomycin was assessed. The phenotypic resistance patterns and year of isolation for these isolates are presented in Table S4.

#### Canadian *S.* Typhimurium DT104

To provide context to the Scottish DT104 isolates, 51 human isolates of DT104 from Canada were sequenced (Table S4). These isolates were from across Canada, over the period 1999 – 2002, and collected through the Canadian Integrated Program for Antimicrobial Resistance Surveillance (39). Phenotypic susceptibility was assessed using the Sensititer Automated Microbiology System (Trek Diagnostic Systems Ltd, Westlake, OH) method (*40*), and isolates were classified as resistant or non-resistant to each antimicrobial, according to CLSI breakpoints (*41*). Resistance to amoxicillin-clavulanic acid, ampicillin, amikacin, gentamicin, kanamycin, streptomycin, ceftiofur, ceftriaxone, cefoxitin, nalidixic acid, ciprofloxacin, sulphamethoxazole, trimethoprim-sulphamethoxazole, tetracycline and chloramplenicol was assessed.

#### English and Welsh *S.* Typhimurium DT104

To provide context to the Scottish DT104 isolates, 12 human isolates and 12 isolates from other animals were sequenced using the Illumina MiSeq platform and 150 bp paired end reads. The human isolates were provided by Public Health England (PHE; formerly the Health Protection Agency), and were isolated between 1991 – 2005. The isolates from other animals were provided by the Animal Health and Veterinary Laboratories Agency (AHVLA), and were isolated between 1996 – 2004 (then the Veterinary Laboratories Agency). Phenotypic susceptibility of the animal isolates (AHVLA) was assessed using disc diffusion (Table S8); isolates were classified as resistant or susceptible according to the zone sizes in Table S8 (*42*). Phenotypic susceptibility of the human isolates (PHE) was assessed using a modified breakpoint technique (*43*) (Table S9). The phenotypic resistance patterns and year of isolation for these isolates are presented in Table S5. One animal isolate was subsequently found to be contaminated and so was removed from all further analysis.

#### **Methods**

#### Genomic analysis

#### Mapping

Following sequencing, the DT104 isolates, the *S.* Typhimurium LT2 and *S.* Typhimurium SL1344 reference sequences (accession numbers AE006468 and FQ312003, respectively) were mapped to the finished *S*. Typhimurium DT104 chromosome and plasmid (accession numbers

HF937208 and HF937209, respectively) using SMALT v0.5.8. (*44*). Prophage sequences, which are known to be highly variable (*45, 46*), the multidrug resistance region of *Salmonella* Genomic Island 1 and the virulence plasmid were then excluded from single nucleotide polymorphism (SNP) calling, leaving a core genome of 4,686,262 base pairs. Genome-wide identification of SNPs and small insertions or deletions in the core genome, compared to the reference genome, were called. The minimum base call quality to call a SNP was set at 50, and the minimum mapping quality to call a SNP was set at 30 (*47*). Recombination events were detected and removed as outlined in Croucher *et al*. (*48*). RAxML v7.0.4 was used to reconstruct a phylogenetic tree from the SNPs called from the core genome (*49*), using a general timereversible model with a gamma correction for among site variation. Support for nodes was assessed using 100 random bootstrap replicates. The tree was visualized using the Interactive Tree of Life (*50, 51*).

The mutation rate was calculated using the SNP alignment of the 359 typical Scottish and non-Scottish DT104 isolates, using the BEAST v1.7.4 software package (*18*). A proportion of invariant sites was included, as well as a discrete gamma distribution to model rate variation among sites. The isolation dates of samples in years were used to calibrate the time scale of the tree, and an uncorrelated lognormal relaxed molecular clock was used to accommodate rate variation among lineages (*52*). The exponential growth coalescent tree prior was used (*53, 54*), with a general time-reversible nucleotide substitution model. Four independent Markov Chain Monte Carlo (MCMC) analyses were run for 25 million states, sub-sampled once every 10,000 states. LogCombiner (*18*) was used to remove 10% as burn-in, resample every 50,000 states, and to combine those sub-samples from the four runs. The mutation rate was estimated to be 3.4 x  $10^{-7}$  substitutions/site/year (95% highest posterior density [HPD] interval: 3.1 x  $10^{-7} - 3.7$  x  $10^{-7}$ ).

#### Assembly

The raw Illumina data were used to create a *de novo* draft assembly of the genome of each sample using the VELVET v0.7.03 algorithm (55), creating multi-contig draft genomes.

#### Identification of antimicrobial resistance determinants

The 147 Scottish isolates of DT104 from humans and other animals selected to investigate the diversity of observed phenotypic resistance profiles were interrogated for antimicrobial resistance determinants. An antimicrobial resistance determinant is defined as either a gene that has been previously identified to be associated with AMR, or a SNP that has been previously identified to be associated with AMR in *Salmonella*, such as those in DNA gyrase subunit A and quinolone resistance. The presence or absence of acquired resistance genes and SNPs associated with resistance were identified in the following way: A non-redundant pan-resistance pseudomolecule was created as a 'sub-reference sequence', consisting of the DT104 reference chromosome and plasmid, and other resistance genes and related regions that have been previously found to be involved in resistance to the 13 antimicrobials for which there were phenotypic data. These included unique regions of different SGI1 variants found in the literature (*17, 56-58*), as well as resistance genes reported to be found in *Salmonella* (*59, 60*), and other genetic regions, mainly plasmids, from which most of these genes were found (Table S10). After re-mapping the Illumina reads per sample with SMALT (*44*) to the resulting new pseudomolecule composed of both the DT104 reference sequence and the pan-resistance subreference sequence, resistance genes and related determinants within the samples that are not

present in the DT104 reference genome were detected. To identify further resistance determinants that were not included in the pan-resistance pseudomolecule, the accessory genome regions of the draft genomes were searched using BLAST.

A list of every gene identified, known to be relevant for antimicrobial resistance, was compiled for each isolate. Antimicrobial resistance genes that were believed to be pseudogenes, either due to truncation or interruption by another gene, were included, as they provide an indication of the evolutionary history of the isolate with respect to AMR. The *gyrA*, *gyrB*, *parC* and *parE* genes were inspected for SNPs that have been previously described as conferring resistance to quinolone antimicrobials (*59*). Venn diagrams of the number of resistance determinants and number of genotypic resistance profiles, and of the number of phenotypic resistance profiles in the original 5,200 domestically-acquired DT104 isolates from 1990 – 2004 (*11*) were generated using the VennDiagram package (*61*) of R (*62*). The same methods were used to interrogate all isolates for the presence or absence of the same resistance determinants. Table S11 details the number of antimicrobial resistance determinants and number of unique resistance profiles for the 133 typical DT104 isolates.

It is worth noting that four resistance phenotypic profiles of the original 65 observed (*11*) were not represented in the sequencing analysis, due to non-viability of the archived isolate or contamination; of these, three were from humans. There were also two phenotypic resistance profiles of the original 35 observed in the animal isolates that that are not represented in the sequencing analysis, due to non-viability of the archived isolate or contamination, but which are represented in the sequenced human isolates. Similarly, two phenotypic resistance profiles of the original 52 observed in the human isolates are not represented in the sequencing analysis, due to non-viability of the archived isolate or contamination, but which are represented in the sequenced animal isolates.

To evaluate whether or not differential sampling bias could be, in part, responsible for our observation of a greater diversity of resistance determinants and profiles in the human isolates, we performed a rarefaction analysis using the vegan package (*63*) of R (*62*) on the dataset of 147 isolates. This examines the number of genotypic profiles (species richness) for a certain number of isolates, and evaluates whether or not there is additional, unsampled diversity. The diversities cannot be directly compared, as these particular isolates are a highly non-random subset of the overall sample collection from humans and animals (*n*=5,200). While the number of genotypic resistance profiles cannot be compared statistically, the greater diversity in the human isolates confirms that observed in the phenotypic resistance profiles. What be observed in Fig. 3D is that we have sampled the animal isolates as thoroughly, if not more so, than the human isolates, and thus suggests that our results cannot be accounted for by sampling bias.

#### Bayesian phylogenetic inference

The BEAST v1.7.4 software package (*18*) was used for Bayesian ancestral state reconstruction, using discrete phylogenetic diffusion models (*19*). Four models were set up, allowing for either bidirectional (both human-to-animal and animal-to-human transmission), or unidirectional transmission, and for the bidirectional models, either symmetric (equal two-way) or asymmetric (allowing unequal) transmission. The models therefore represented: 1) bidirectional asymmetric diffusion, 2) bidirectional symmetric diffusion, 3) unidirectional human-to-animal diffusion, 4)

unidirectional animal-to-human diffusion. The data for these models were the alignment of variable sites (SNPs) of the 248 Scottish DT104 isolates, 135 (54%) human and 113 (46%) animal, and the discrete trait representing the host population from which the isolates were obtained. A proportion of invariable sites was included, as well as a discrete gamma distribution to model rate variation among sites. These isolates were sampled from 1990 – 2011 (see Tables S2 and S3), excluding the 14 atypical DT104 isolates. The exponential growth coalescent tree prior was used (*53, 54*), with a general time-reversible nucleotide substitution model. Other tree priors were explored and compared using Bayes factors estimated through path sampling (*64*); the exponential prior was the preferred model. The isolation dates of samples in years were used to calibrate the time scale of the tree, and an uncorrelated lognormal relaxed molecular clock was used to accommodate rate variation among lineages (*52*). A conditional reference prior was specified on the overall rate scalar (clock rate) in the continuous-time Markov chain (CTMC) model for the phylogenetic diffusion of the discrete host population trait (*65*). We used stochastic mapping techniques to estimate both the transitions (Markov jumps) and the waiting times (Markov rewards) of the host trait diffusion process throughout the evolutionary history (*66, 67*). Markov jumps estimates provide expectations for the unobserved human-to-animal and animalto-human transitions along each branch of the tree; Markov rewards estimates provide corresponding expectations for the amount of time that is spent in each state, human or animal. Log marginal likelihoods obtained by path sampling and the resulting log Bayes factors revealed strong evidence against both the unidirectional scenarios. For these marginal likelihood estimations, we treated trees, independently estimated from the sequence data, as a discrete set of possibilities (*68*); the analysis for all four models integrated over the same empirical tree distribution. The best fitting model was the asymmetric bidirectional model (Table S12); four independent Markov Chain Monte Carlo (MCMC) analyses were run for 50 million states, subsampled once every 10,000 states, using BEAGLE (*69*) in conjunction with BEAST (*18*). LogCombiner (18) was used to remove 10% as burn-in, resample every 50,000 states, and to combine those sub-samples from the four runs. The maximum clade credibility tree from the resulting 3,600 trees was summarized with TreeAnnotator and visualized with FigTree (*18*). The posterior median number of unobserved animal-to-human transitions along branches was 39  $(95\%HPD: 27 - 55)$ , and the median number of unobserved human-to-animal transitions within branches was  $27$  (95%HPD:  $17 - 36$ ). Of the entire evolutionary time represented by Fig. 2A, the Markov rewards indicated the model spent a median of 400 (95%HPD: 318 – 521) years in the animal state, and a median of 666 (95%HPD:  $545 - 771$ ) years in the human state. To quantify and test the degree of host admixture we used a modified Association Index (AI) (*19*). Briefly, for each tree in our posterior distribution, we calculate the association value following Wang *et al* (*70*), which quantifies the association between phylogeny and host traits. We calculate the same value for a number of permutations  $(n = 10)$ , in which traits are randomly associated with the tree tips, and take the ratio of the association value for the real traits and the corresponding mean value for the permutations. Finally, we report the posterior distribution for this ratio by summarizing the AI for each tree in the posterior sample. A general deviation from the permuted distributions implies low AI values and suggests host structure in the phylogeny whereas AI values close to 1 suggest host admixture or no more clustering by host as expected from random association. The AI was 0.66 (95%HPD:  $0.57 - 0.76$ ), which rejects the null hypothesis (AI = 1), and indicates that clustering within the phylogenetic trees is not randomized.

We also conducted additional analyses with two different subsets of the data, with two independent MCMC analyses each. The majority of animal isolates are from cattle, reflecting the main animal reservoir of DT104 (*20, 31*). Thus, we performed the same analysis examining the 135 Scottish human isolates, and the 83 Scottish cattle isolates. We also conducted the same analysis dividing the animal isolates into their respective species, excluding species which were represented less than five times in the 113 animal isolates, giving 83 bovine, seven ovine, eight porcine, and six poultry isolates (*n*=104). In both cases, as the number of animal isolates decreased in the dataset, the dominance of the human ancestral state in the evolutionary history increased, as one would expect. In the cattle-only model, compared to the model with the full dataset, there were higher human Markov rewards, lower cattle Markov rewards, and fewer cattle-to-human Markov jumps. In the model sub-dividing the animal species, compared to the full model there were higher human Markov rewards, although similar animal Markov rewards, fewer animal-to-human Markov jumps and more human-to-animal Markov jumps. These results substantiate our conclusions based on the more conservative analysis including all animal and human isolates.

Comparing phenotypic resistance profiles and the genomic backbone of *S.* Typhimurium DT104 Each isolate in the dataset that was submitted to the SSSCDRL, as well as the reference DT104 sequence (accession HF937208), was included in this comparison. The 275 isolates from the SSSCDRL included those acquired domestically in Scotland, and those submitted from Scottish patients with a recent history of foreign travel; only these isolates were included, as the same microbiological and antimicrobial susceptibility testing methods were used to characterize these isolates. A molecular phylogenetic tree was generated as described in the Mapping section, by mapping the isolates to the reference sequence, calling SNPs from the core genome, and using RAxML (*49*) to draw the phylogenetic relationships. The distribution of the main phenotypic profile, ApClSpStSuTe, throughout the molecular phylogenetic tree of the same 275 isolates was visualized by plotting this specific trait on the tree using the Interactive Tree of Life (*50, 51*) (Fig S3A). All phenotypic resistance profiles, the combinations of phenotypic resistance to the 13 antimicrobials assessed, were also plotted against the tree (Fig. S3B).



# **Fig. S1.**

Maximum likelihood phylogeny of all 373 *Salmonella* Typhimurium DT104, from Scotland and elsewhere, demonstrating a main clade and a subset of 14 isolates. Scale bar represents number of substitutions per single nucleotide polymorphism site per year.



# **Fig. S2**

Single nucleotide polymorphisms (SNPs; blue) and homoplastic SNPs (red) across the genome of *S.* Typhimurium DT104 in the 359 typical DT104 isolates. Non-synonymous SNPs found in >5 isolates primarily were found in genes encoding membrane proteins; genes related to peripheral metabolism, amino acid transport, transcription regulation, catabolic pathways, or disulphide bond formation; genes of unknown function; flagellin; DNA gyrase A; degenerate phage genes; virulence-related genes; or pseudogenes of various classes.



#### **Fig. S3**

Maximum likelihood phylogenetic tree, mid-point rooted, using single nucleotide polymorphisms of 275 *Salmonella* Typhimurium DT104 isolates processed by the Scottish *Salmonella Shigella* and *Clostridium difficile* Reference Laboratory, with A) isolates exhibiting resistance to the ApClSpStSuTe phenotypic resistance profile (red), putatively conferred by *Salmonella* Genomic Island 1, and other phenotypic resistance profiles (black), and B) all phenotypic resistance profiles colored individually. The asterisk indicates the location of the reference isolate HF937208.



### **Fig. S4.**

Phylogeny of Scottish and global *Salmonella* Typhimurium DT104, rooted on *S.* Typhimurium SL1344. The colored ring indicates the putative *Salmonella* Genomic Island 1 variant within each isolate; (Ps) indicates a pseudogene.



**Fig S5.** Phylogeny from Fig. 1 of Scottish and non-Scottish *Salmonella* Typhimurium DT104, rooted on *S.* Typhimurium SL1344, with bootstrap values.



# **Fig. S6**

Bayesian maximum clade credibility phylogenetic tree and most probable ancestral state reconstruction of host population for *Salmonella* Typhimurium DT104 in Scotland of Fig. 2A. Branches with a reconstructed state (host population) posterior probability are colored red for human, blue for animal; branch width is scaled by the posterior probability of reconstructed state.



## **Fig. S7**

A) Number of *S.* Typhimurium DT104 domestically-acquired isolates submitted per year to the Scottish *Salmonella Shigella* and *Clostridium difficile* Reference Laboratory, 1990 – 2004 (reproduced from  $(11)$ ), and B) the number of sequenced domestically-acquired  $(n=262)$  *S*. Typhimurium DT104 isolates, 1990 – 2011, by year of isolation in the study dataset.





\* demonstrating resistance to ampicillin, chloramphenicol, streptomycin, spectinomycin, sulphonamides, tetracycline

§ see Table S3 for details

**Table S2.** Human and animal isolates of *S.* Typhimurium DT104 from Scotland, submitted to the Scottish *Salmonella Shigella* and *Clostridium difficile* Reference Laboratory, showing antimicrobial resistance (AMR) phenotypic profile and resistance determinants, in the subset of 147 isolates used to assess the diversity of resistance.  $Ap = \text{ampicillin}$ ,  $Cl = \text{chloramphenicol}$ ,  $Sp = \text{specinomycin}$ , St  $=$  streptomycin, Su = sulphonamides, Te = tetracycline, Ka = kanamycin, Cp = ciprofloxacin, Na = nalidixic acid, Gm = gentamicin,  $Ne$  = netilmicin,  $Tm$  = trimethoprim,  $Fz$  = furazolidone. Genes marked as  $(Ps)$  are believed to be pseudogenes, either through truncation (Ps) or interruption by another gene (Ps-2f).











\* Table S10 provides the Uniprot identifiers for each of the resistance determinants listed.

**Table S3.** Human and animal isolates of *S.* Typhimurium DT104 submitted to the Scottish *Salmonella Shigella* and *Clostridium difficile* Reference Laboratory, with antimicrobial resistance (AMR) phenotypic profile and resistance determinants, in the subset of 93 Scottish isolates used to assess the diversity of isolates demonstrating the main phenotypic resistance profile, the 22 Scottish isolates used to investigate the post-epidemic period  $2005 - 2011$ , and the 27 travel-associated isolates. Ap = ampicillin, Cl = chloramphenicol, Sp = spectinomycin, St = streptomycin, Su = sulphonamides, Te = tetracycline, Ka = kanamycin, Cp = ciprofloxacin, Na = nalidixic acid, Gm = gentamicin, Ne = netilmicin, Tm = trimethoprim, Fz = furazolidone. Genes marked as (Ps) are believed to be pseudogenes, either through truncation (Ps) or interruption by another gene (Ps-2f).











**Table S4.** Human isolates from Canada, as tested by the Canadian Integrated Program for Antimicrobial Resistance Surveillance, human and animal isolates from Japan, as tested by the National Institute of Infectious Diseases and the National Institute of Animal Health, of *S.* Typhimurium DT104 with antimicrobial resistance (AMR) phenotypic profile and resistance determinants.  $Ap =$ ampicillin,  $Cl =$  chloramphenicol,  $St =$  streptomycin,  $Su =$  sulphonamides,  $Te =$  tetracycline,  $Na =$  nalidixic acid,  $Tm =$  trimethoprim. Genes marked as (Ps) are believed to be pseudogenes, either through truncation (Ps) or interruption by another gene (Ps-2f).

		<b>AMR</b>		
	Isolated	phenotypic	Year of	
Isolate	from:	profile	isolation	Resistance determinants identified
CH <sub>1</sub>	Human	ApStSuTe	2001	bla(PSE-1), floR, aadA2, sull, tetG, tetR
CH <sub>2</sub>	Human	pansusceptible	2001	
CH <sub>3</sub>	Human	StSu	2001	aad $A2$ , sull, aph $(3')$ -I
CH <sub>4</sub>	Human	ApClStSuTe	2001	$bla(PSE-1)$ , floR, aadA2, sulI, tetG, tetR
CH <sub>5</sub>	Human	ApClStSuTe	2001	$bla(PSE-1)$ , floR, aadA2, sull, tetG, tetR
CH <sub>6</sub>	Human	ApClStSuTe	2001	$bla(PSE-1)$ , floR, aadA2, sulI, tetG, tetR
CH <sub>7</sub>	Human	ApStSuTe	2001	$bla(PSE-1)$ , floR, aadA2, sull, tetG, tetR
CH <sub>8</sub>	Human	ApClStSuTe	2001	$bla(PSE-1)$ , floR, aadA2, sulI, tetG, tetR, aph(3')-I
CH <sub>9</sub>	Human	ApClStSuTe	2001	$bla(PSE-1)$ , floR, aadA2, sulI, tetG, tetR, aph(3')-I, dfrA12, sulIII, bla(TEM-1b), mefE
CH <sub>10</sub>	Human	StSu	2001	aadA2, sulI
CH <sub>11</sub>	Human	pansusceptible	2001	
CH <sub>12</sub>	Human	ApClStSuTe	2001	$bla(PSE-1)$ , floR, aadA2, sull, tetG, tetR
CH <sub>13</sub>	Human	pansusceptible	2001	
CH14	Human	Te	2001	$aph(3')-I$
CH <sub>15</sub>	Human	pansusceptible	2001	
CH <sub>16</sub>	Human	ApClStSuTe	2001	$bla(PSE-1)$ , floR, aadA2, sull, tetG, tetR
CH17	Human	ApClStSuTe	2001	$bla(PSE-1)$ , floR, aadA2, sulI, tetG, tetR, aph(3')-I
CH18	Human	ApClStSuTe	2001	$bla(PSE-1)$ , floR, aadA2, sulI, tetG, tetR, aph(3')-I
CH <sub>19</sub>	Human	SuTe	2001	tetA, tetR2, sulII, hygBr, $aac(3)$ -IV, aph(3')-I
<b>CH20</b>	Human	ApClStSuTeTm	2001	$bla(PSE-1)$ , floR, aadA2, sull, tetG, tetR, tetA(A), tetR(A), aph(3')-I, dfrA12
CH <sub>21</sub>	Human	ApClStSuTe	2001	$bla(PSE-1)$ , floR, aadA2, sull, tetG, tetR, aph(3')-I
<b>CH22</b>	Human	pansusceptible	2001	





**Table S5.** Human and animal isolates, as tested by Public Health England (formerly the Health Protection Agency) and Animal Health and Veterinary Laboratories Agency respectively (animal isolates processed by the Veterinary Laboratories Agency), from England and Wales of *S*. Typhimurium DT104 with antimicrobial resistance (AMR) phenotypic profile and resistance determinants. Ap = ampicillin,  $Cl =$  chloramphenicol,  $St =$  streptomycin,  $Sp =$  spectinomycin,  $Su =$  sulphonamide compounds,  $Te =$  tetracycline,  $Tm =$ trimethoprim, Fz = furazolidone, Na = nalidixic acid,  $Cp$  = ciprofloxacin, Tm = trimethoprim, Sxtm = sulphamethoxazole/trimethoprim. Genes marked as (Ps) are believed to be pseudogenes, either through truncation (Ps) or interruption by another gene (Ps-2f).



**Table S6.** Antimicrobial resistance profiles determined by presence and absence of genomic resistance determinants as represented in Figure 1.



**Table S7.** Concentrations of antimicrobials used in the susceptibility testing of the animal and human *S.* Typhimurium DT104 isolates by the Scottish *Salmonella Shigella* and *Clostridium difficile* Reference Laboratory, 1990 – 2004, and 2005 – 2011.

Antimicrobial	Breakpoint (mg/l) 1990 - 2004	Breakpoint (mg/l) 2005 - 2011
Ampicillin	50	8
Chloramphenicol	20	8
Ciprofloxacin	0.5	0.5
Furazolidone	20	8
Gentamicin	20	4
Kanamycin	20	16
Nalidixic acid	40	16
Netilmicin	20	20
Spectinomycin	100	64
Streptomycin	20	16
Sulphamethoxazole	100	64
Tetracycline	10	8
Trimethoprim	10	2

**Table S8.** Concentrations used in the disc diffusion susceptibility testing of the animal *S.* Typhimurium DT104 isolates from England and Wales by the Animal Health and Veterinary Laboratories Agency (isolates processed by the Veterinary Laboratories Agency).



**Table S9.** Concentrations of antimicrobials used in the susceptibility testing of the human *S.* Typhimurium DT104 isolates from England by Public Health England (formerly the Health Protection Agency).

Antimicrobial	Breakpoint (mg/l)
Amikacin	4
Ampicillin	8
Cefotaxime	1
Ceftriaxone	
Cefuroxime	16
Cephalexin	16
Cephradine	16
Chloramphenicol	8
Ciprofloxacin	0.125
Colomycin	8
Furazolidone	8
Gentamicin	$\overline{4}$
Kanamycin	16
Nalidixic acid	16
Neomycin	8
Spectinomycin	64
Streptomycin	16
Sulphonamides	64
Tetracycline	8
Trimethoprim	$\overline{2}$

	UniProt	the antimicropial resistance genes investigated in the D. Typininumum DTT0+ isolates. <b>ENA</b> accession		UniProt	
Gene name	ID	no.	Gene name	ID	ENA accession no.
$bla(PSE-1)$	Q7BL37	AAK02055.1	mph(A)	Q5QJG2	AAR05762.1
$f$ lo $R$	Q7BL41	AAK02049.1	mefE	Q5J436	AAS76329.1
sulI	E0D898	AEX00802.1	dhfrXVI	O85802	AAC32186.1
aadA2	Q7BL43	AAK02046.1	dfr1	Q6J3S3	AAT36680.1
tetR	Q7BL40	AAK02050.1	dfrA27	B <sub>2</sub> ZNP <sub>4</sub>	ACD56152.1
tetG	Q7BL39	AAK02051.1	dfrA23	Q5W314	CAG34233.2
$aac(6')$ -Ib-cr	<b>B9VR93</b>	ACM24779.1	dfrA21	Q6Q8S1	AAS66087.1
$aac(3)-IV$	P08988	X01385	dfrA19	Q8VVE6	CAC81324.1
dfrA14	A7WNT6	CAM98046.1	cm1A	Q5J429	AAS76336.1
sat2	Q75QQ2	BAD10975.1	cat	D0R779	CBA11382.1
aadA1	B0FGV6	ABY50547.1	catB8	Q79PD0	AAM92461.1
dfrA15	Q0ZB28	ABG36698.1	catB3	O86929	CAA08841.1
dhfrA7	E5G6I0	ADP08975.1	catB2	Q8KLQ3	CAD31710.1
<b>B</b> 1dhfrVII	Q79K64	AAO89216.1	cat2	Q5J470	AAS76295.1
dhfrX	Q79S90	AAL13155.1	ble	A8R700	BAF93087.1
dhfrIII	P12833	AAA25550.1	$bla(TEM-1b)$	B5SZN3	ACH85856.1
dfrA17	Q83ZN7	AAP23220.1	$bla(per-2)$	P74842	CAA63714.1
dfrA12	Q8GLV1	ACF21684.1	$bla(OXA-30)$	Q6QLX3	AAS46622.1
hygBr	<b>H9TI80</b>	AFG20898.1	$bla(OXA-53)$	Q7WTW0	AAP43641.1
$aph(3')$ -II	P00552	AAA73390.1	$bla(OXA-2)$	P0A1V8	AAA98357.1
tetR2	Q79VX4	BAB91577.1	$bla(KPC-2)$	Q7B856	AAM10643.1
tetR(A)	B7ZJI1	ACK44536.1	$bla(SHV-2)$	P0AA00	AAA75015.1
tetD	Q9S453	BAB91574.1	$bla(DHA-1)$	O54216	CAB40919.1
tetC	Q93F25	BAB91575.1	$bla(CTX-M-2)$	P74841	CAA63263.1
tetA	Q9K2Y4	BAB91576.1	$aph(3')-I$	Q5QJP8	AAR05693.1
tetA(A)	A7DY41	CAO00285.1	ampC	Q5J3Z2	AAS76373.1
sulII	D0R7A7	CBA11366.1	$aac(3)-II$	Q5QJN0	AAR05727.1
sulIII	Q7WZL0	AAP82508.1	$acc-1$	Q49JG6	AAX52125.1
sugE	E7DBI0	ADV39907.1	$aar-3$	Q83ZU8	ACD56151.1
strB	B7ZJI3	ACK44538.1	aadB	Q79LX7	AAO46870.1
strA	B7ZJI4	ACK44539.1	aadA7	Q6SIX0	AAR21615.1
sph	A8R701	BAF93088.1	aadA5	Q75T47	BAD07296.1
qnrS	<b>B7TZ43</b>	ACJ24509.1	aadA16	<b>B3V3X6</b>	ACF17980.1
qnr	Q3Y8H2	AAZ78355.1	aacC	A4IVL4	ABO41023.1
qnrB19	C6H187	CAZ67058.1	aacC1	O86934	CAA08847.1
pef	A5H8A5	ABN13922.1	aacA4	Q8KLQ4	CAD31708.1
$q \times B$	<b>F4MK98</b>	CBL62366.1	$aac(6') - 130$	Q7WTV9	AAP43642.1
qxA	<b>F4MK97</b>	CBL62365.1	aac6-II	Q79PC7	AAM92464.1
mrx	Q5QJG3	AAR05761.1	$aac(3)-Id$	Q6SIX1	AAR21614.1

**Table S10.** Gene, UniProt identifier, and European Nucleotide Archive (ENA) accession number for the antimicrobial resistance genes investigated in the *S.* Typhimurium DT104 isolates.

**Table S11.** The number of antimicrobial resistance (AMR) determinants, AMR profiles based on presence/absence of AMR determinants in the 133 typical Scottish human and animal isolates *S.* Typhimurium DT104 investigated for AMR diversity. The numbers of determinants or profiles unique to the particular host population (animal or human) are represented in brackets.

Host population	# Determinants $(\# \text{ unique})$	# Profiles (# unique)
Human	21(7)	21 (14)
Animal	20(6)	14
Shared	$\overline{4}$	
Total	77	28

**Table S12.** Log marginal likelihoods of the four assessed models for the Bayesian phylogenetic analysis: bidirectional asymmetric, bidirectional symmetric, unidirectional human-to-animal, unidirectional animal-to-human, and log Bayes factors (logBF) for each model compared to the bidirectional asymmetric model.

Model	Log marginal likelihood	Log(BF)
Bidirectional asymmetric	$-207.52$	
Bidirectional symmetric	$-208.35$	$-0.83$
Unidirectional: human-to-animal	$-241.36$	$-33.84$
Unidirectional: animal-to-human	$-267.26$	-59.74