

RESEARCH ARTICLE

Evolution of *Salmonella enterica* serotype Typhimurium driven by anthropogenic selection and niche adaptation

Matt Bawn^{1,2}, Nabil-Fareed Alikhan¹, Gaëtan Thilliez¹, Mark Kirkwood¹, Nicole E. Wheeler³, Liljana Petrovska⁴, Timothy J. Dallman⁵, Evelien M. Adriaenssens¹, Neil Hall², Robert A. Kingsley^{1,6*}

1 Quadram Institute Biosciences, Norwich Research Park, Norwich, United Kingdom, **2** Earlham Institute, Norwich Research Park, Norwich, United Kingdom, **3** Centre for Genomic Pathogen Surveillance, Wellcome Sanger Institute, Cambridge, United Kingdom, **4** Animal and Plant Health Agency, Addlestone, United Kingdom, **5** Gastrointestinal Bacteria Reference Unit, National Infection Service, Public Health England, London, United Kingdom, **6** University of East Anglia, Norwich, United Kingdom

* Rob.Kingsley@quadram.ac.uk



OPEN ACCESS

Citation: Bawn M, Alikhan N-F, Thilliez G, Kirkwood M, Wheeler NE, Petrovska L, et al. (2020) Evolution of *Salmonella enterica* serotype Typhimurium driven by anthropogenic selection and niche adaptation. PLoS Genet 16(6): e1008850. <https://doi.org/10.1371/journal.pgen.1008850>

Editor: Xavier Didelot, University of Warwick, UNITED KINGDOM

Received: January 22, 2020

Accepted: May 12, 2020

Published: June 8, 2020

Copyright: © 2020 Bawn et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Freely available from the European Nucleotide archive under the run accession numbers ERS007564, ERS007582, ERS007584, ERS007566, ERS007567, ERS007572, ERS007574, ERS007576, ERS007578, ERS007580, ERS007588, ERS007606, ERS007608, ERS007590, ERS007592, ERS007594, ERS007596, ERS007598, ERS007600, ERS007602, ERS007604, ERS007611, ERS007613, ERS007615, ERS008962, ERS008964, ERS008980, ERS015602, ERS015603,

Abstract

Salmonella enterica serotype Typhimurium (*S. Typhimurium*) is a leading cause of gastroenteritis and bacteraemia worldwide, and a model organism for the study of host-pathogen interactions. Two *S. Typhimurium* strains (SL1344 and ATCC14028) are widely used to study host-pathogen interactions, yet genotypic variation results in strains with diverse host range, pathogenicity and risk to food safety. The population structure of diverse strains of *S. Typhimurium* revealed a major phylogroup of predominantly sequence type 19 (ST19) and a minor phylogroup of ST36. The major phylogroup had a population structure with two high order clades (α and β) and multiple subclades on extended internal branches, that exhibited distinct signatures of host adaptation and anthropogenic selection. Clade α contained a number of subclades composed of strains from well characterized epidemics in domesticated animals, while clade β contained multiple subclades associated with wild avian species. The contrasting epidemiology of strains in clade α and β was reflected by the distinct distribution of antimicrobial resistance (AMR) genes, accumulation of hypothetically disrupted coding sequences (HDCS), and signatures of functional diversification. These observations were consistent with elevated anthropogenic selection of clade α lineages from adaptation to circulation in populations of domesticated livestock, and the predisposition of clade β lineages to undergo adaptation to an invasive lifestyle by a process of convergent evolution with of host adapted *Salmonella* serotypes. Gene flux was predominantly driven by acquisition and recombination of prophage and associated cargo genes, with only occasional loss of these elements. The acquisition of large chromosomally-encoded genetic islands was limited, but notably, a feature of two recent pandemic clones (DT104 and monophasic *S. Typhimurium* ST34) of clade α (SGI-1 and SGI-4).

ERS015598, ERS015599, ERS015600, ERS015601, ERS015604, ERS015613, ERS015614, ERS015605, ERS015606, ERS015607, ERS015608, ERS015609, ERS015610, ERS015611, ERS015612, ERS015615, ERS015624, ERS015625, ERS015616, ERS015617, ERS015618, ERS015619, ERS015620, ERS015621, ERS015622, ERS015623, ERS015626, ERS015635, ERS015636, ERS015627, ERS015628, ERS015629, ERS015630, ERS015631, ERS015632, ERS015633, ERS015634, ERS015637, ERS015646, ERS015647, ERS015638, ERS015639, ERS015640, ERS015641, ERS015642, ERS015643, ERS015644, ERS015645, ERS014081, ERS014078, ERS014080, ERS015648, ERS015657, ERS015658, ERS015649, ERS015650, ERS015651, ERS015652, ERS015653, ERS015654, ERS015655, ERS015656, ERS015659, ERS015668, ERS015660, ERS015661, ERS015662, ERS015663, ERS015664, ERS015665, ERS015666, ERS015667, ERS023488, ERS023489, ERS023490, ERS023491, ERS023492, ERS023493, ERS023494, ERS023502, ERS023504, ERS023505, ERS023506, ERS023520, ERS023523, ERS023525, ERS023527, ERS023528, ERS023529, ERS023530, ERS023532, PRJEB34598, PRJEB34597, PRJEB34599, PRJEB34596, PRJEB34595, PRJEB34594.

Funding: RAK was funded by the BBSRC Institute Strategic Programme Microbes in the Food Chain BB/R012504/1 and its constituent project(s) BBS/E/F/000PR10348 and BBS/E/F/000PR10352, and by projects BB/J004529/1, BB/M025489/1 and BB/N007964/1. NH was supported by a BBSRC funding for the Earlham Institute BB/CCG1720/1. EMA was supported by the BBSRC Institute Strategic Programme Gut Microbes and Health BB/R012490/1 and its constituent project BBS/E/F/000PR10356. The genome sequencing for this work was carried out by the Genomics Pipelines group at the Earlham Institute which is funded as a BBSRC National Capability (BB/CCG1720/1). EMA was funded by the BBSRC Institute Strategic Programme Gut Microbes & Health BB/R012490/1 and its constituent projects BBS/E/F/000PR10353 and BBS/E/F/000PR10356. NFA was supported by the Quadram Institute Bioscience BBSRC funded Core Capability Grant (project number BB/CCG1860/1). This research was supported in part by the NBI Computing infrastructure for Science (CiS) group through use of HPC resources. The funders had no role in study design, data collection

Author summary

Salmonella Typhimurium is a leading cause of foodborne illness worldwide. Our current understanding of the biology of *Salmonella* is largely based on studies using just two laboratory strains of *S. Typhimurium*, with similar characteristics. Yet this pathogen exhibits a remarkable diversity in host range, outcome of infection, and risk to human health. To investigate the genetic basis of this diversity, we have explored the genetic relationship of a collection of isolates that represent a substantial portion of the diversity of *S. Typhimurium*, using whole genome sequencing. *S. Typhimurium* evolved forming two major groups that differ in their distribution in livestock and wild avian species. The livestock-associated group contained isolates commonly affecting human health and were often drug resistant, while the wild avian-associated group was rarely associated with drug resistance and less frequently associated with human infection. We report distinct evolutionary processes acting on subgroups of *S. Typhimurium* including loss of information content of genomes, and gain or loss of genes predicted to affect functions such as antimicrobial resistance, disease potential, and environmental survival, and functional diversification. This study provides a framework in which to understand the epidemiology of *Salmonella* and improve assessment of its risk to animal and human health.

Introduction

Bacteria of the genus *Salmonella* are a common cause of foodborne disease. Most of the approximately 2500 serovars cause gastroenteritis in humans and other animals, while some have evolved host adaptation associated with extra intestinal disseminated infections in specific host species [1]. For example, *Salmonella enterica* serovar Typhimurium (*S. Typhimurium*) and *S. Enteritidis* circulate in multiple vertebrate host species and cause food borne infections in the human population. These and other non-typhoidal *Salmonella* serotypes result in an estimated 75 million cases and 27 thousand deaths from gastroenteritis worldwide [2]. *S. Typhi* and *S. Paratyphi A* circulate exclusively in the human population and cause an estimated 2.5 million infections resulting in 65 thousand deaths each year as a result of the disseminated disease typhoid and paratyphoid disease [2]. Similarly, other serotypes evolved host adaptation to specific non-human host species, such as *S. Gallinarum* with poultry, *S. Dublin* with cattle, and *S. Choleraesuis* with pigs, where they are associated with disseminated infections [1].

Although *S. Typhimurium* is considered to be a broad host range serotype, the epidemiological record of *S. Typhimurium* phage types identified several *S. Typhimurium* pathovariants with distinct host range, pathogenicity and risk to food safety [3, 4]. The pathovariant commonly associated with this serotype, has a broad host range and is associated with gastroenteritis in the human population. Such broad host range strains of *S. Typhimurium* account for the majority of those isolated by public health surveillance in England, presumably because they are common in many species of livestock and poultry, the primary zoonotic reservoir for human infections [5]. The epidemiological record of this pathovariant is characterised by successive waves of dominant clones identified historically by their phage type, that account for up to 60% of all human infections for several years, before being replaced by a subsequent strains [6]. Dominant clonal groups have been characterized by strains of phage types DT9, DT204/49 complex, DT104, and the current monophasic *S. Typhimurium* (*S.* 4,[5],12:i:-) sequence type 34 (ST34), since around the middle of the last century [7–10]. In contrast, some phage types are common in clonal groups typically associated with a restricted host range, and

and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

in some cases altered pathogenicity. For example, clonal groups of *S. Typhimurium* DT8, DT2 and DT56 circulate in populations of ducks, pigeon, and passerine birds, respectively, and only rarely cause gastroenteritis in the human population [11–13]. Also, specific clonal groups of *S. Typhimurium* ST313 are associated with disseminated disease (invasive non-typhoidal *Salmonella*, iNTS) in sub-Saharan Africa [14, 15].

In this study we report the population structure, gene flux, recombination and signatures of functional diversification in the whole genome sequence of 131 strains of *S. Typhimurium* with well characterised epidemiology. To assist in our analysis, we also report high quality complete and closed whole genome sequence of six additional reference genomes, representing diversity within the population structure not represented by previously reported sequence data.

Results

Population structure of *S. Typhimurium* consists of two high-order clades containing strains with distinct epidemiology

Variant sites (38739 SNPs) in the core genome sequence of 134 *S. Typhimurium* strains representing commonly isolated phage types revealed two diverse phylogroups composed of three strains of ST36 that clustered separately from the remaining 131 *Typhimurium* isolates (S1 Fig). The two phylogroups were more similar to one another than any other serotype, including a closely related isolate of serotype *S. Heidelberg*. Since the majority of *S. Typhimurium* formed a large number of relatively tightly clustered isolates, predominantly of ST19, we focussed on the analysis of the population structure and evolution of this phylogroup. A phylogenetic tree constructed using variant sites (8382 SNPs) in the core genome sequence of the 131 *S. Typhimurium* strains and rooted with *S. Heidelberg*, revealed a ‘star’ topology with relatively long internal branches extending from a hypothetical common ancestor, and diversification at the terminal branches (Fig 1).

The population structure determined using a three-level hierarchical Bayesian approach [16] resolved *S. Typhimurium* into two major clades, designated α and β , seven second level clades and 18 third-level clades (α 8–18 and β 1–7) (S1 Table). In many cases, third level clades corresponded to known epidemic clades, with the exception of β 1 that was a poorly defined basal clade, and we therefore focused on 1st and 3rd level clades. Clade β was defined by an internal branch that originated from a common ancestor of the basal clade α , defined by approximately 100 core genome SNPs.

Despite relatively few SNPs distinguishing clade α and β , these clades exhibited distinct epidemiology characterised by association predominantly with livestock, including cattle, pigs and poultry (clade α) or avian species, including wild species (clade β). Strikingly all pig isolates from our sampling were located in clade α . Cattle isolates were in both first order clades (11 in clade α 7 in clade β), but in clade β they had a relatively limited distribution with five of the isolates from a subclade containing the DT204/49 complex of strains associated with a cattle associated epidemic in the 1970’s [8]. Clade α contained strains from several previously described epidemics in livestock animal species, including in pigs (α 12, U288)[17], two clades associated with recent pandemic clonal groups associated with pigs, cattle and poultry (α 17, monophasic *Typhimurium* ST34 and α 15, DT104) [18–20], and potentially epidemic clades not previously described in the literature, consisting of isolates from pigs, cattle and poultry (α 8 and α 11). Clade- β was characterised by many long internal branches, indicative of a relatively high level of root to tip sequence divergence, relative to those in Clade- α . In contrast, clade- β contained several third-level clades previously described as host-adapted, particularly for avian species such as passerine birds (β 5, DT56), duck (β 2, DT8) and pigeon (β 3 DT2)

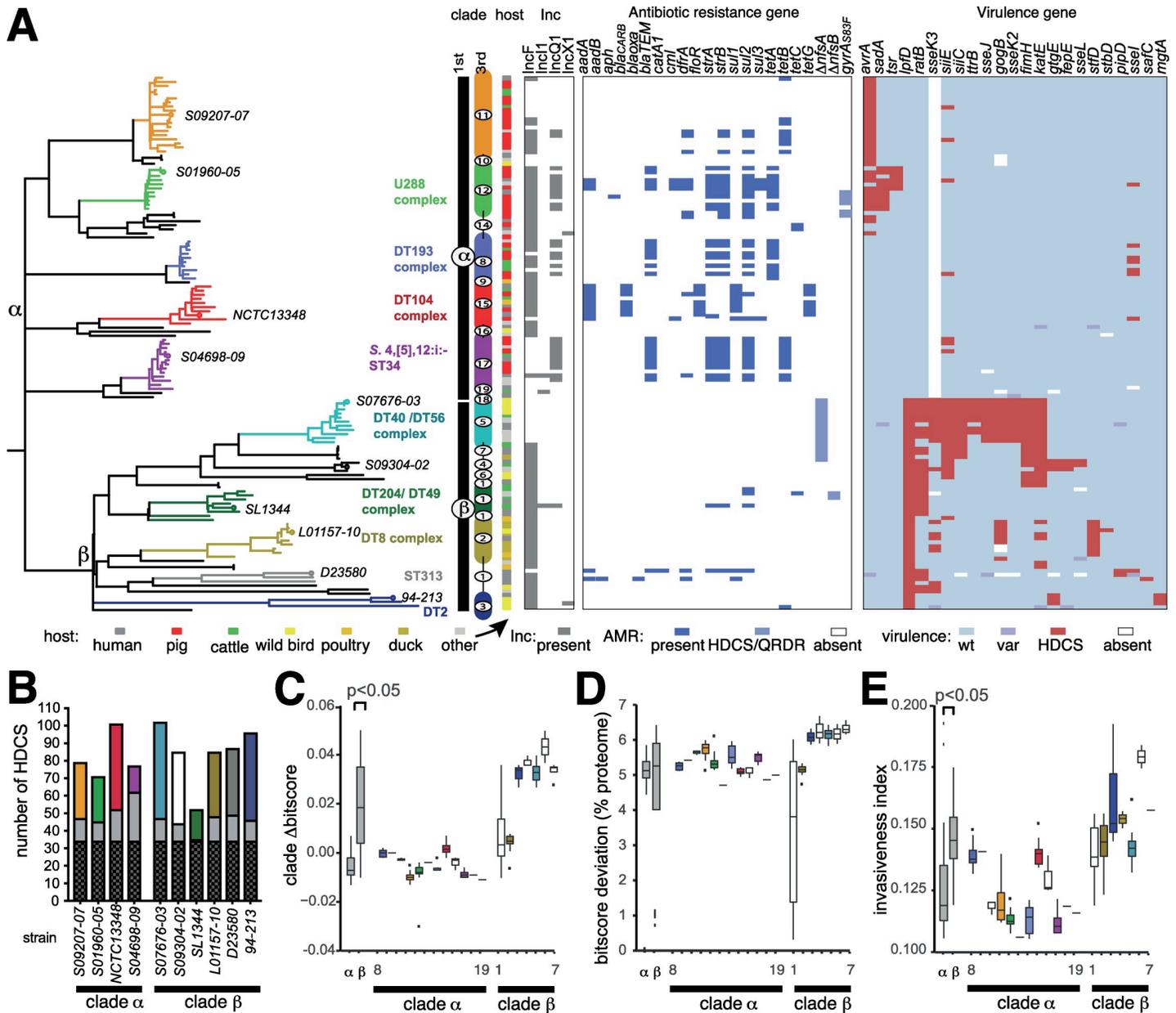


Fig 1. Phylogenetic relationship of the ST19 *Salmonella* Typhimurium phylogroup. (A) Maximum likelihood phylogenetic tree and based on sequence variation (SNPs) in the core genome with reference to *S. Typhimurium* strain SL1344. The root was identified using *S. Heidelberg* (accession number NC_011083.1) as the outgroup. 1st (α and β) and 3rd (α11–19 and β1–7) are indicated (vertical bars). Phage type complexes associated with the third-level clusters are indicated (bold type) colour coded with the lineages and representative strains from third level clusters (italicized type). The source of each isolate in the tree is indicated by filled boxes colour coded as indicated in the inset key (arrow). The presence of replicon sequence (grey box), antimicrobial resistance genes (blue box) and hypothetically disrupted coding sequence (HDCS) of virulence related genes (red box) in short read sequence data are indicated. HDCS in *nfsA* and *nfsB* resulting in resistance to nitrofurantoin antibiotics and ns SNPs resulting in substitutions in the QRDR of *GyrA* are indicated (light blue boxes). (B) Bars indicate the number of ancestral (black), phage or insertion sequence elements (grey), chromosomal gene (colour coded with lineages in Fig 1A) HDCS in the genome of representative strains from each third level clade. (C) Box plots indicate the mean Δbitscore (DBS: bitscore SL1344 –test strain bitscore) of proteomes in third level clades. (D) Box plot indicates the percentage of the proteome of the proteome of isolates from each third level clade with a non-zero bitscore (bitscore SL1344 –test strain bitscore >0 or <0) as an estimate of function divergence. (E) Box plots indicate the mean invasiveness index per genome, the fraction of random forest decision trees voting for an invasiveness phenotype based on training on the DBS of a subset of the proteome of ten gastrointestinal and extraintestinal pathovar serotypes.

<https://doi.org/10.1371/journal.pgen.1008850.g001>

[21], and ST313 that includes two sub-clades specifically associated with disseminated disease in sub-Saharan Africa [22].

To estimate the coverage of clinical isolates from the UK and global *S. Typhimurium* by our dataset, we compared the 131 *S. Typhimurium* genomes with 1697 *S. Typhimurium* genomes isolated from human clinical infection in the UK in 2014 and 2015 described previously [23], and 14,478 genomes from non-UK isolates, excluding ST36, in the Enterobase database [24] (S2 Fig). With both datasets, we grouped these genomes according to their cgMLST Hierarchical Clusters at the level of 100 alleles difference (HierCC HC100) as defined in Enterobase [24]. This provided an estimation of the genomic diversity within *S. Typhimurium*. We then determined the proportion of HierCC groups that contained a genome from our dataset, and the number of genomes in shared clusters. In the first comparison, 38.5% of hierarchical clusters were represented in our dataset and UK clinical isolates. Shared HC's contained 95.1% of genomes from clinical infections because most of the unrepresented HC's contained few genomes. In regard to the rest of the world, 17% of HC's were represented in our dataset, but shared HierCCs accounted for 71.5% of the Enterobase genomes. Therefore, although our dataset represents less than half of the known hierarchical clusters, it does cover the majority of genomes in available databases.

Antimicrobial resistance genes and plasmid replicons are common in isolates associated with livestock

The presence of multiple third-level clades associated with recent livestock associated epidemic strains in the first level clade α and the relative paucity in clade β suggested that they may be under differential anthropogenic selection pressure. A key anthropogenic selection pressure on microbial populations circulating in livestock is the widespread use of antimicrobial drugs in animal husbandry. Consistent with their distinct epidemiology, antimicrobial resistance (AMR) genes were common in clades α (mean of 2.7 per strain) and relatively rare in β (mean 0.38 per strain) (Fig 1A and S2 Table). Indeed, AMR genes in clade- β were restricted to strains from DT204/49 complex known to be associated with cattle [8] and the ST313 associated with disseminated disease in sub-Saharan Africa commonly treated with antibiotics [22].

Resistance to fluoroquinolone and nitrofurantoin antimicrobials is associated with sequence polymorphisms in the *gyrA* and *parC*, and *nsfA* and *nsfB*, respectively [25, 26]. Of four substitutions in the quinolone resistance determining region (QRDR) of GyrA and ParC, just GyrA S83F was present in six isolates, four from α 12 (pig-associated U288) and two from α 8 (DT104 complex), had a sequence polymorphism resulting in GyrA S83F. No other mutations affecting the QRDR were detected. Nitrofurantoin resistance has been linked to HDCS due to insertion sequence (IS) elements or sequence polymorphisms, primarily of the *nsfA* gene, but occasional secondary mutations in *nsfB* [26]. The *nsfA* gene was present as an HDCS in just two isolates from subclade β 1 and *nsfB* in 22 isolates from subclade β 5 (DT56 complex).

AMR genes are commonly present on plasmids and we therefore determined the presence of plasmid replicon sequence in short read sequence data from the 131 strains in clades α and β . The IncQ1 plasmid replicon, previously associated with antibiotic resistance [27] was widespread, particularly in clade α . The IncF replicon corresponding to the presence of the virulence plasmid pSLT [28] was also widespread in *S. Typhimurium* as expected, and associated with antibiotic resistance in a number of subclades including ST313 and U288 complex [14, 29]. The pSLT plasmid varied in size ranging from 96 to 167kb (Table 1), and assembly of pSLT from short read sequence from all 131 isolates indicated no significant difference in the mean size between clades α and β . The pSLT plasmid was notably absent from a number of third-level clades including β 5 (DT56), α 11 and α 17 (monophasic *S. Typhimurium* ST34).

Table 1. Characteristics of complete and closed whole genome sequence of *S. Typhimurium* reference strains.

strain	ST	PT ^a	accession no.	clade	chrom size ^b	pSLT ^c	P1 ^c	P2 ^c	ECC ^d	CDS ^e	SNPs ^f	Prophage ^g	total HAC ^h	spec. HAC	year isolation	source
SO9207-07	19	DT170B	PRJEB34598	α 11	4.92	-	-	-	0	4592	973	7 (5:1:1)	48	7	2007	Pig
SO1960-05	19	U288	PRJEB34597	α 12	4.90	154.4	19.4	18.2	3	4832	1272	8 (5:2:1)	57	16	2005	Pig
NCTC13348	19	DT104	HF937208.1	α 15	4.93	94.0	-	-	1	4751	1083	9 (5:3:1)	63	25	1988	Human
SO4698-09	34	DT193	LN999997.1	α 17	5.04	-	-	-	0	4754	1010	9 (6:2:1)	43	7	2009	Cattle
SO7676-03	19	DT56	PRJEB34599	β 5	4.88	-	-	-	0	4570	797	6 (4:1:1)	75	26	2003	Bird
SO9304-02	19	DT41	PRJEB34596	β 4	5.05	117.4	32.2		3	4685	943	8 (5:2:1)	78	15	2002	Cattle
SL1344	19	DT44	FQ312003.1	β 1	4.88	93.8	86.9	8.7	3	4771	0	9 (4:3:2)	42	2	~1966	Cattle
LO1157-10	19	DT8	PRJEB34595	β 2	4.86	93.8	23.9	22.2	3	4706	1101	7 (4:2:1)	62	19	2010	Duck
D23580	313	-	FN424405.1	β 1	4.88	117.0	84.6	-	2	4790	901	7 (5:2:0)	57	9	2004	Human
A130	313	DT56v	PRJEB34594	β 1	4.93	166.9	-	-	1	4812	901	8 (5:2:1)	62	13	2001	Human
94-213	98	DT2	HG326213.1	β 3	4.82	93.8	-	-	1	4598	903	5 (3:1:1)	66	20	1994	Pigeon

^a Phage type / Sequence type^b Chromosome size (Mbp)^c contig size (Kbp)^d extra chromosomal contigs (ECC), putative plasmid sequences^e Number of predicted coding sequences^f number of SNPs with reference to SL1344^g Number of prophage (intact, incomplete, candidate)^h total number of pseudogenes with reference to SL1344 allele, ⁱ number of clade-specific pseudogenes (S1 Table), ⁱ mean delta bitscore (DBS) with reference to SL1344 protein orthologue.<https://doi.org/10.1371/journal.pgen.1008850.t001>

Many key virulence genes of *Salmonella enterica* are present on *Salmonella* pathogenicity islands (SPIs) [30]. In general, SPI-1 to 6 were highly conserved, consistent with their key role in pathogenesis. However, SPI-4 containing the *sii* locus that encodes a giant adhesin secreted by a type I secretions system [31], exhibited elevated sequence divergence in several β subclades (S3 Fig). SPI-6 that encodes a type VI secretion system that mediates a cell contact dependent mechanisms of interbacterial antagonism involved in colonisation of the intestine [32], exhibited moderate sequence variation in isolates in subclade β 7.

Distinct patterns of genome degradation and signatures of functional divergence and invasiveness in clades α and β

Hypothetically disrupted coding sequences (HDCS) due to frameshift mutations or premature stop codons were determined in high quality finished and closed genome sequence of 11 representative strains from major subclades (Fig 1B and S3 Table). Representative strains of clade α generally contained fewer HDCS than those from clade β , with the exception of NCTC13348 (DT104) and SL1344 (DT204/49) that had atypically high and low numbers of HDCS, respectively (Fig 1B). However, none of the HDCS in NCTC13348 were in genes previously been implicated in pathogenesis, while SL1344 contained two virulence gene HDCS (*lpfD* and *ratB*) (Fig 1A). In general, clade α strains had 0–3 HDCS in virulence genes (mean 0.8 SD 0.9), while

clade β was characterised by multiple lineage containing three or more virulence gene HDCS (mean 5.0 SD 3.2) (Fig 1A and S4 Table). Isolates in clade β 5 (DT56, passerine bird associated [13]) contained up to eleven virulence gene HDCS (*lpfD*, *ratB*, *sseK3*, *siiE*, *siiC*, *ttrB*, *sseJ*, *gogB*, *sseK2*, *fimH* and *katE*). The greatest number of HDCS in clade α were observed in α 12 (U288, possibly pig adapted), in which three virulence genes were affected (*avrA*, *sadA* and *tsr*). *lpfD* was found to be the only HDCS (S3 Table) that segregated between clade- α and clade- β . A 10-nucleotide deletion causing a frameshift mutation in *lpfD* resulting in a truncation approximately half way into the protein in all isolates of clade- β .

To quantify the relative level of functional divergence in the proteome of isolates in each clade we used a profile hidden Markov Model approach, delta-bitscore (DBS profiling) [33] (Fig 1C and S5 Table). The method assigned a value (bitscore) to peptides of the proteome that indicated how well each sequence fitted the HMM. We determined the difference in bitscore of the proteome of each isolate relative to that of *S. Typhimurium* strain SL1344 (DBS = bitscore SL1344 proteome—bitscore test proteome). A greater DBS is therefore indicative of excess of polymorphisms that potentially alter protein function, and most likely a loss of function as it indicates divergence from the profile HMM. Mean DBS was significantly greater ($p < 0.05$, Wilcoxon test) for proteomes of strains in clade- β compared with clade- α (Fig 1C). In general, third-level clades in clade- α exhibited DBS of approximately zero, consistent with limited functional divergence. Notably, despite considerable numbers of HDCS in strain NCTC13348 (α 15, DT104), DBS was only moderately elevated in this clade. Proteomes of strains in clade β exhibited mean DBS of 0.03 and above with the exception of clades β 1 and β 2. The proportion of the proteome with any deviation in DBS was also greater in clade- β than clade- α (Fig 1D).

We also used a machine learning approach to predict the ability of strains to cause extraintestinal disease based on convergent patterns of mutation accumulation detected by delta bitscore (DBS), in 196 proteins that were recently determined and reported as the most predictive of the invasiveness phenotype in 13 serotypes (six extra-intestinal pathovars and seven gastrointestinal pathovars) of *S. enterica* subspecies I [34, 35] (Fig 1E). Protein sequences for the 196 genes were retrieved for each isolate in the study, scored by DBS, and run through the model. The invasiveness index metric is the fraction of decision trees in a random forest algorithm that vote for an invasive phenotype based on DBS values. The invasiveness index was significantly greater ($p < 0.05$, Wilcoxon test) in clade- β than clade- α , consistent with the epidemiology and pathogenicity of the isolates located in these clades [3, 4].

The clade-specific accessory genome is largely driven by acquisition of prophage genes and integrative elements

A pangenome analysis of *S. Typhimurium* (excluding the ST36 phylogroup) identified 9167 total gene families. The core genome (present in 99–100% of strains) was 3672 genes, soft core genome (95–99%) 388 genes. Shell genome (15–95%) 792 genes, and cloud genes (0–15%) 4315 genes (Fig 2A and 2B). We defined gene families of the accessory genome as non-prophage chromosomal, prophage, plasmid and undefined, based on their location and annotation in the complete and closed genomes of eleven reference strains phylogenetically distributed across *S. Typhimurium* (Table 1). Gene families not present in reference genomes were classified as ‘undefined’. The accessory genome was defined as genes present in 95% or fewer of isolates and thus represents the major source of genetic variation between strains.

Some gene families exhibited a distinct distribution in clade α or β , or within individual third-level clades of clade α or β (S4 Fig). Four non-phage chromosomal genes were specifically associated with clade- β strains, STM0038 a putative arylsulfatase, *tdcE* encoding a

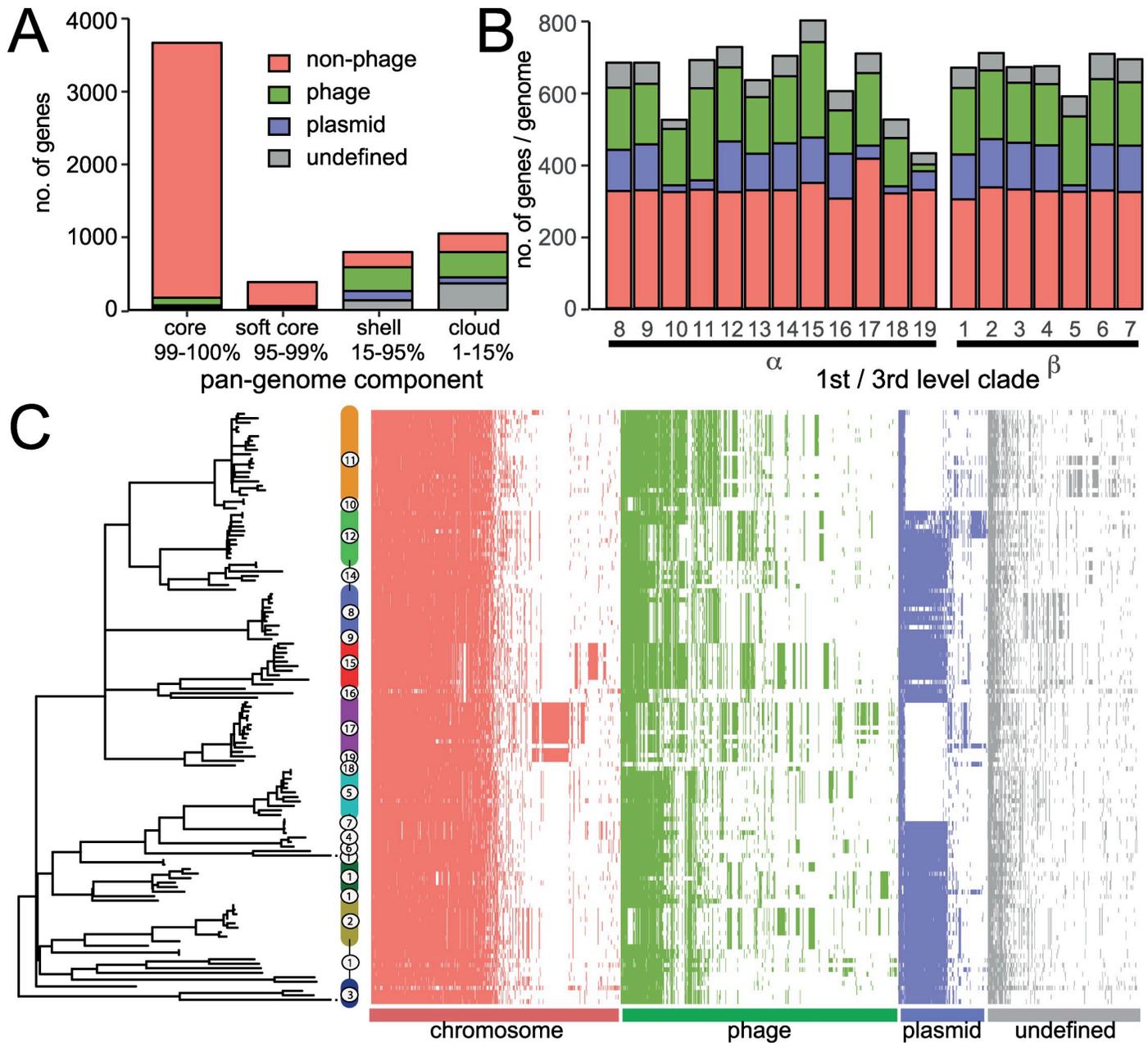


Fig 2. The pan genome of 131 *S. Typhimurium* isolates. Gene families were identified based on sequence alignment with a cut off of 90% sequence identity and assigned to non-phage chromosomal (red), prophage (green), plasmid (blue), or undefined (grey), based on their genome context in eleven annotated reference genomes from each third level clade. (A) Number of genome families in the core, softcore, shell and cloud components of the pangenome. (B) Number of genome families of each pan genome component in isolates from each *S. Typhimurium* third level clade. (C) Accessory genome (shell and cloud) in each isolate. Gene families present in more than 130 or less than 5 strains were excluded. Maximum Likelihood tree based on variation (SNPs) in the core genome with reference to *S. Typhimurium* SL1344. Third-level clades are indicated in colour coded in common with the phylogeny vertical bars.

<https://doi.org/10.1371/journal.pgen.1008850.g002>

pyruvate formate lyase 4, *aceF* acetyltransferase and *dinI* a DNA damage inducible protein. A series of plasmid genes in clade β 2 corresponded to a region of p2 seen in LO1157-10. This region is likely a transposon as it contains an IS200 transposase and an integrase. Also, a number of prophage-associated genes were present throughout clade- β due to apparent recombination in the ST64B prophage. The rate of gene flux in clade α and clade β was determined by

computing the number of accessory genes as a function of SNPs in each clade. By this measure, the rate of gene flux was nearly twice as high in clade α compared to clade β (S5 Fig).

Generally, gene flux in the non-prophage and non-plasmid gene families that were specific to individual third level clades was limited to individual genes or small blocks of genes (S4 Fig). The exception was two large genetic islands in clades α 15 (DT104 complex) and α 17 (monophasic *S. Typhimurium* ST34) corresponding to the presence of SGI1 [36] and SGI4 [37]. In addition, a chromosomal block of genes in clade β 2 corresponding to an insertion at the Thr-tRNA at 368274 containing a series of hypothetical proteins and a gene with similarity to the *trbL* gene involved in conjugal transfer (A0A3R0DZN6) and a site-specific integrase. Some of these genes were also present in isolates in clades β 1 and β 3. The greatest contribution to third level clade specific gene families was in the those with predicted functions in prophage (Fig 2C and S4 Fig).

Extant prophage repertoire is the result of recombination and infrequent loss of ancestral elements and acquisition of new phage

In order to investigate the flux of prophage genes resulting in clade-specific repertoires, we identified prophage in eleven complete and closed reference genomes of *S. Typhimurium* sequences. A total of 83 complete or partial prophage elements were identified in the eleven reference genomes (Fig 3). Prophage were present at twelve variably occupied chromosomal loci and the number per strain ranged from five in DT2 (strain 94–213) to nine in monophasic *S. Typhimurium* ST34 (strain SO4698-09) (Table 1). Clustering of gene families in the prophage pangenome identified 23 prophage, although in some cases blocks of genes were replaced resulted in mosaic prophage for example “*Salmonella* virus ST64BX” and “*Salmonella* virus Gifsy1X” (S6 Table), and the definition of families of prophage with a high confidence was consequently problematic (Fig 4). Thirteen prophage elements encoded at least one identifiable cargo gene, capable of modifying the characteristics of the host bacterial strain, including eleven genes previously implicated in virulence (S3 Table). Ten prophage families contained no recognisable cargo genes. The evolutionary history of prophage acquisition and loss was reconstructed based on principles of maximum parsimony. Six prophage (*Salmonella* viruses “BcepMuX”, “Gifsy1X”, “Gifsy2X”, “Fels1X”, “ST64BX” and “sal3X”, hereafter referred to as BcepMu, Gifsy1, Gifsy2, Fels1, ST64B and sal3) (S6 Table), that together accounted for 61 of the prophage in these genomes, were most likely present in the common ancestor of *S. Typhimurium*. Loss of two of these ancestral prophage by three isolates (NCTC13348, L01157-10 and D23580) represented the only evidence for decrease in prophage repertoire in the dataset.

A total of 22 additional prophage had a limited distribution within *S. Typhimurium* strains, present in three or fewer genomes *Salmonella* viruses (“TmEGF”, “TmSEN34”, “mTmII”, “mTmV”, BTP1, “TmST104”, “SPN9CC”, Fels2, “TmHP1/mTmHP1”, BTP5, “TmC3PO”, “RE2010”, “TmR2D2” and “TmSEN1”) (S6 Table) and are therefore likely to have been acquired during the evolution of *S. Typhimurium* (Fig 4). *Salmonella* virus BTP1 that was reported to be specific to the ST313 strains associated with epidemics of invasive NTS disease in sub-Saharan Africa [38], was also present in strain SO7676-03, a strain in clade β 5 (DT56 complex) adapted to circulation in wild bird (Passerine) species [12]. “*Salmonella* virus mTmV” of strain SO4698-09, that carries the *sopE* virulence gene in some monophasic *S. Typhimurium* ST34 isolates [18], was absent from all other *S. Typhimurium* reference strains. However, a second prophage “*Salmonella* virus mTmII” with similarity to SJ46 was also in strain SO4698-09, and shared several clusters of gene families in common with mTmV.

With the notable exception of BcepMu the prophage predicted to be present in the common ancestor of *S. Typhimurium* exhibited considerable variation, potentially due to

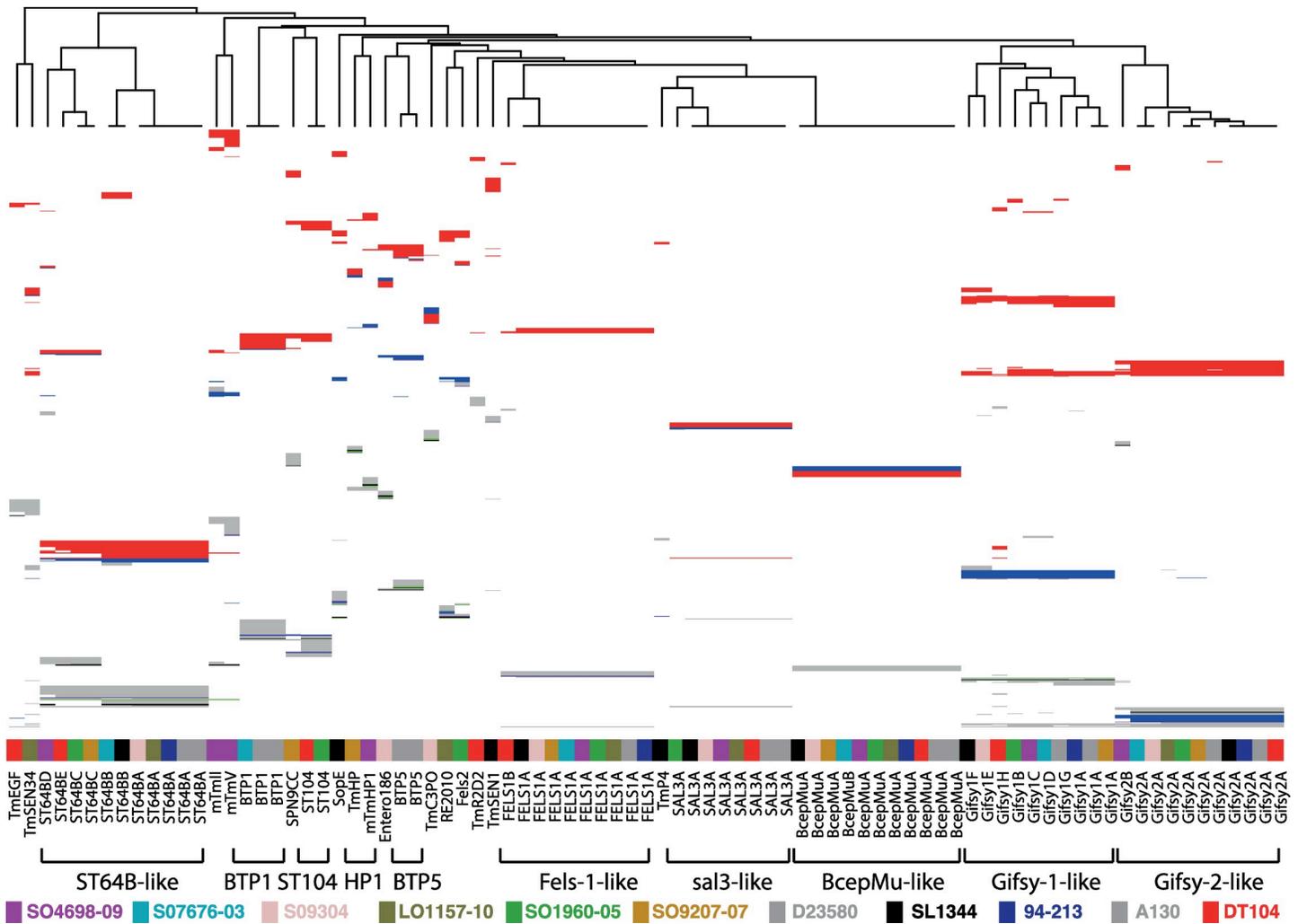


Fig 3. Clustering of prophage genes based on sequence identity indicates related families and potential recombination. Genes from all prophage identified in complete and closed whole genome sequence of eleven reference strains of *S. Typhimurium* were assigned to families based on sequence identity (>90% identity). Prophage genes (columns) were clustered to identify related prophage. The presence of a gene is indicated with a box predicted function based on *in silico* annotation are colour coded based on annotation, terminase (black), capsid (green), recombinase/integrase (purple), tail fibre (blue), other phage associated (red), and hypothetical protein (grey). A cladogram showing the relationship of prophage is based on the pattern of gene presence or absence is indicated (top).

<https://doi.org/10.1371/journal.pgen.1008850.g003>

recombination [39] (Fig 4). Recombination is a major source of genetic variation in bacteria, although the level of recombination seen in other bacteria and indeed other *Salmonella* serovars vary greatly. We identified potential recombination in the genome sequence of the 131 *S. Typhimurium* of the main phylogroup by the identification of atypical SNP density. Recombination was almost exclusively present in prophage regions resulting in clade specific sequence variation (Fig 5). Recombination resulted in replacement of large blocks of gene families in ancestral prophage elements. Fels1, sal3 and Gifsy2 were conserved in the most reference strains, with the exception of Fels1 in DT104 and Gifsy2 in monophasic *S. Typhimurium* ST34 that had large alternative blocks of gene families. Gifsy1 was highly variable in all strains, but retained a core set of genes suggesting common ancestry and frequent recombination. Variation in ST64B was also present in most strains, and variable blocks of genes distinguished strains in first order clades α from β , and resulted in the acquisition of *sseK3* virulence gene by the common ancestor of the latter.

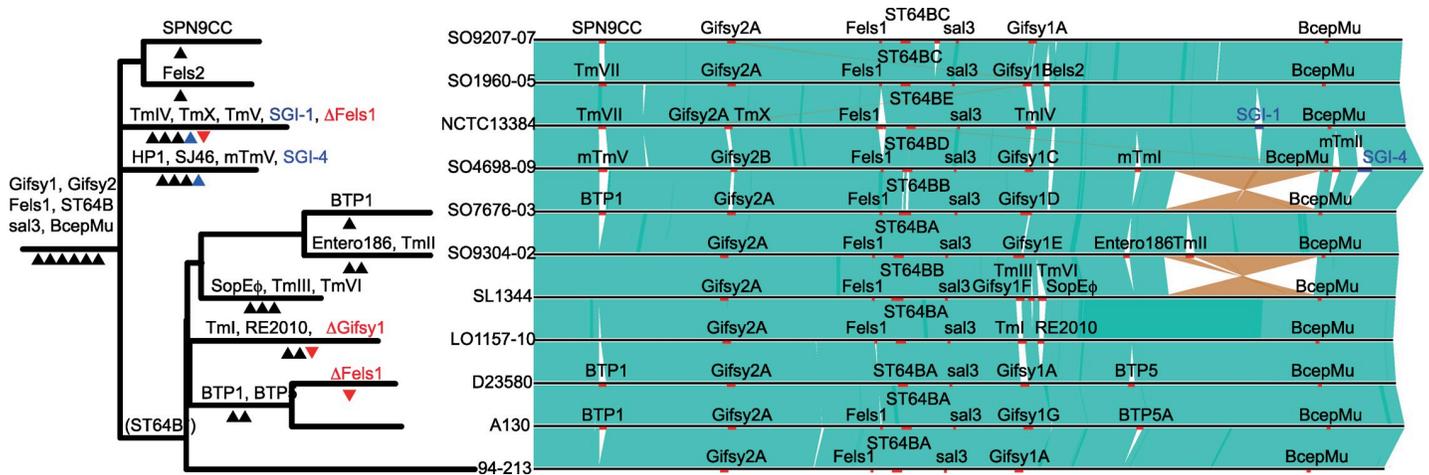


Fig 4. Genome alignment and phylogenetic relationship of complete and closed reference strains of *S. Typhimurium* or used in this study. Sequence with >90% nucleotide sequence identity are indicated where this is direct alignment (green) or reverse and complement (red). The location of prophage sequence (red bars) or integrative elements (blue bars) are indicated. A maximum likelihood tree based on sequence variation (SNPs) in the core genome with reference to *S. Typhimurium* strain SL1344 (left) is annotated with the most likely order of acquisition (black arrow) or loss (red arrow) of prophage and integrative elements, based on the principle of parsimony.

<https://doi.org/10.1371/journal.pgen.1008850.g004>

Discussion

The population structure of *S. Typhimurium* consisted of two relatively distantly related clusters comprising strains of ST36 and a second (main) phylogroup predominantly ST19, containing the remainder of *S. Typhimurium*, consistent with previous reports of two distinct *S. Typhimurium* phylogroups [40, 41]. These two groups were more closely related to each other than to other serotypes of *S. enterica* subspecies I. The main Typhimurium phylogroup exhibited a star shaped phylogeny with multiple deeply rooted branches emerging from a common ancestor, with diversification at the terminal branches in some cases, associated with expansion of epidemic clonal groups. The topology of this *S. Typhimurium* phylogroup was similar to that of serotypes of *S. enterica* subspecies I [42, 43], with internal branches radiating from a common ancestor, defined by the accumulation of hundreds of SNPs in *S. Typhimurium* compared with tens of thousands of SNPs defining lineages of representative strains of distinct serotypes of *S. enterica* subspecies I [4].

The nested phylogenetic structure, rooted with the *S. Heidelberg* outgroup, was characterised by two high order clades (α and β), in which clade α was basal to clade β . Several deeply rooted lineages of clade α contained isolates almost entirely from livestock. A single lineage originating from the common ancestor of the main *S. Typhimurium* phylogroup gave rise to the common ancestor of clade β and diversification into multiple lineages was accompanied by apparent host adaptation to diverse host species, but notably many more avian species, compared with clade α . The β subclades included those associated with the DT56, DT2 and DT8 complexes that are well characterized host adapted clonal groups [11, 12, 21], contained exclusively isolates from avian species, and were present on relatively extended internal branches. This general phylogenetic topology is consistent with that described for distinct collections of *S. Typhimurium* strains from North America and Asia [44, 45], and Enterobase [24]. The Enterobase database contained over 40,000 *S. Typhimurium* genome sequences at the time of writing, with additional additional α and β subclades not present in our dataset, presumably because they are not present in the UK. A lack of isolates from wild avian hosts and incomplete metadata limited our ability to test many of our key findings with this larger dataset.

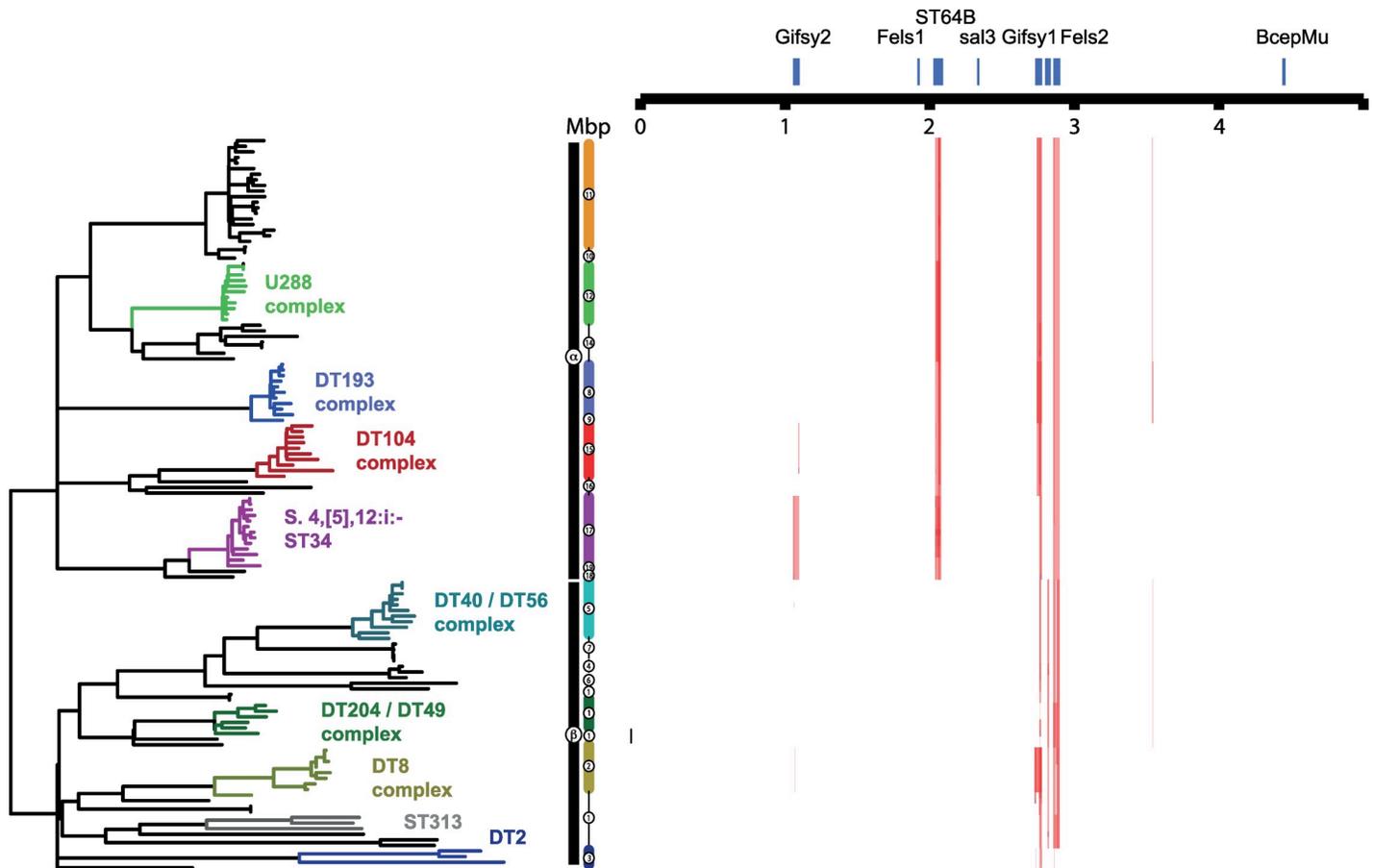


Fig 5. Recombination inferred by high SNP density in 131 *S. Typhimurium* strains. Regions of high SNP density (red) are indicated for each of the 131 isolates in the *S. Typhimurium* ST19 cluster with reference to the *S. Typhimurium* strain SL1344 genome. Recombination is shown with reference to the population structure and phylogeny of Typhimurium shown in Fig 1. The position of predicted prophage (blue) in the *S. Typhimurium* strain SL1344 genome are indicated (top).

<https://doi.org/10.1371/journal.pgen.1008850.g005>

Isolates in α subclades and some β subclades were under distinct anthropogenic selection pressure for the acquisition and maintenance of AMR, that correlated with their distribution in livestock or wild avian species [46]. Clade α isolates that were predominantly from livestock contained several lineages with multiple AMR genes, while most clade β isolates contained few or no AMR genes. Differential selection pressure for acquisition and maintenance of AMR genes is consistent with the idea that some *S. Typhimurium* genotypic variants are adapted to circulation in specific host populations that exert different selection pressure for the acquisition and maintenance of AMR genes. Antimicrobials have been used widely to control infection or as growth promoters in livestock, but wild animals are unlikely to encounter therapeutic levels of these drugs [47]. However, *S. Typhimurium* strains of DT56 and DT40 present in clade β and known to be associated with passerine birds, lacked AMR genes, yet are occasionally isolated from cattle and human clinical infections where they may be subject to selection for antimicrobial resistance [48]. We might therefore also expect to find AMR genes in DT56 and DT40 strains. One possibility is that DT56 and DT40 strains from clade β may transiently colonise the cattle host but are unable to circulate in this population and do not transmit back to the avian population with high frequency. Host adaptation to avian species therefore appears to create an effective barrier to circulation in livestock. Two clade β lineages did contain strains with multiple AMR genes, but in each case their epidemiology was atypical

for clade β in that they were associated with an epidemic in cattle (DT204/49 complex) or invasive NTS in people in sub-Saharan Africa (ST313) [7, 22], and therefore were likely to be under selection for AMR.

The molecular basis of the barrier to circulation of some clade β isolates in livestock is not known, but likely the result of genotypic changes affecting functional diversification of the proteome. The proteome delta bitscore (DBS) of clade β isolates exhibited elevated divergence from profile HMMs of protein families in gamma proteobacteria, compared to clade α isolates, potentially resulting in loss or altered protein function [21]. Similarly, divergence was reported in the *S. Gallinarum* proteome, a serotype highly host adapted to poultry where it is associated with fowl typhoid [33]. β subclades also exhibited an elevated invasiveness index, the fraction of decision trees using random forests that vote for the invasive (extraintestinal) disease outcome, a predictive score of host adaptation to an extraintestinal lifestyle. Consistent with this finding, increased invasiveness of strains from clade- β lineages has been reported previously, including *S. Typhimurium* DT2 isolates in day of hatch chicks [21] and ST313 isolates in day of hatch chicks [49]. Signatures of invasiveness of ST313 strains was also consistent with changes in interaction with mice and cattle in experimental models of infection [50, 51], and multicellular behaviour in the environment [52].

We used a machine learning approach to determine an invasiveness index that was not designed to identify mechanism of invasiveness, but instead discriminated genotypes associated with alternative pathotypes on the basis of shared proteomic signatures [34]. The relationship of invasiveness index and site of isolation was reported previously [53]. In our analysis DBS of 196 protein families most predictive of disseminated disease in *S. Typhi*, *S. Paratyphi A*, *S. Gallinarum*, *S. Dublin* and *S. Choleraesuis*, also predicted an extraintestinal lifestyle in *S. Typhimurium* β subclades, consistent with the epidemiological data [4]. The pattern of functional divergence in some *S. Typhimurium* β subclades may therefore at least in part be by a process of convergent evolution with that observed as a result of the evolution of several extraintestinal serotypes *S. enterica*, including *S. Typhi*, *S. Paratyphi A*, *S. Gallinarum*, *S. Dublin* and *S. Choleraesuis* [34].

Host adaptation of bacteria is commonly associated with the accumulation of HDCS, potential pseudogenes that contribute to genome degradation [54, 55]. A total of 24 genes previously implicated in virulence, adhesion or multicellular behaviour were HDCS in one or more clade β isolates. Notably, 15 of these genes (63%) were also HDCS in highly host adapted *S. Typhi* or *S. Paratyphi A*, serotypes that are restricted to humans and cause a disseminated disease [56]. HDCS were especially common in strains of DT56/DT40 complex (clade β 5), that are reported to be highly host adapted to passerine birds, in which eleven HDCS were observed. In *S. Paratyphi A*, many mutations and gene flux that occurred was reported to be neutral, indicated by their sporadic distribution within clades, and frequent loss from the population by purifying selection [57]. In *S. Typhimurium*, we also observed some examples of potential neutral mutations, but in many cases HDCS in virulence genes were present in multiple related strains from the same subclade, indicating that they were likely under selection, and stably maintained in the population (Fig 1). In contrast to clade β , genome degradation affecting virulence-associated genes was less frequent in isolates clade α . Just three virulence gene HDCS had a clade phylogenetic signature, *avrA*, *tsr* and *sadA*, in α 10, α 11 and α 12. Subclade α 12 (U288 complex) was the only clade to contain all of these HDCS, consistent with the U288 complex exhibiting apparent host-adaptation to pigs [17, 37]. Therefore, the presence of HDCS in virulence associated genes was almost exclusively associated with subclades containing strains with strong epidemiological evidence of host adaptation [4]. Despite the *S. Typhimurium* DT104 complex strain NCTC13348 (clade α 15) exhibiting a high level of genome degradation that was uncharacteristic for α subclades and the broad host range epidemiology of the

clonal group [58], no virulence or multicellular behaviour genes were HDCS [58]. Furthermore, the mean DBS for the proteome of strains from clade α 15 was similar to that of other α subclades, suggesting that functional divergence as a whole was not atypical from that of other clade α isolates.

While genes encoding components of the type III secretion systems (T3SS) 1 and 2 apparatus were never HDCS, several genes encoding effector proteins secreted by them were (*sseI*, *sseK2*, *sseK3*, *avrA*, *sseL*, *sseJ* and *gtgE*). The *sseI* gene is inactivated in ST313 due to insertion of a transposable element, and results in hyper dissemination of these strains to systemic sites of the host via CD11b⁺ migratory dendritic cells [51]. The *sseK2*, *sseK3*, *avrA* and *sseL* genes each encode effectors that inhibit the NF κ B signalling pathway thereby modulating the proinflammatory response during infection [59–61]. Furthermore, these effectors are commonly absent or degraded in serotypes of *Salmonella* serotypes associated with disseminated disease [62, 63], suggesting that altered interaction with the macrophage is essential for disseminated disease in diverse *Salmonella* variants and hosts. In addition, several genes encoding components of fimbrial or afimbrial adhesin systems (*sadA*, *ratB*, *lpfD*, *stfD*, *stbD*, *safC*, *fimH*, *siiC* and *siiE*), or anaerobic respiration (*ttrB*), and chemotaxis (*tsr*) were HDCS in one or more isolates. Many of these genes have been implicated in intestinal colonisation [31, 64–69], suggesting that their inactivation in host adapted variants may be associated with a loss of selection for functions no longer required in a reduced host range or where intestinal colonisation is no longer critical to transmission. The *sadA* and *katE* genes are involved in multicellular behaviour, biofilm formation and catalase activity that protects against oxidative stress during high density growth functions, respectively, and are commonly affected by genome degradation in host adapted pathovars of *Salmonella* [52, 70].

The only virulence-related HDCS that segregated clades α and β resulted from a 10 bp insertion in the *lpfD* gene of all clade β isolates. Within the host *S. Typhimurium* preferentially colonises Peyer's patches (PPs) [71], due to long-polar fimbriae *Lpf* binding to M-cells [72]. Similarly, *lpf* genes in *E. coli* pathotypes are required for interaction with Peyer's patches and intestinal colonisation [73]. Despite the disruption of *lpfD* in the clade β isolate SL1344, deletion of *lpf* reduced colonisation on the surface of chicken intestinal tissue, [74], suggesting that long polar fimbriae retain function. Differences in intestinal architecture of avian species that have the lymphoid organ the bursa of Fabricius containing numerous M cells compared with mammalian species that have Peyer's patches with relatively scarce M cells [75, 76] may explain the pattern of *lpfD* HDCS in *S. Typhimurium*. The function of long polar fimbriae expressing full length *LpfD* in clade α isolates has not been investigated, but its distribution in isolates from livestock and human infections mark it as of potential importance to human health. Other virulence genes including *gtgE*, *avrA* and *sseK3*, exhibited sporadic distribution or HDCS within third level clades, consistent with observations in models of infection suggesting that they may play a role in adaptation to an invasive lifestyle [77–80].

Bacterial genome diversity is largely driven by the flux of genes resulting from acquisition by horizontal gene transfer and deletion, rather than allelic variation [81]. The accessory genome of *S. Typhimurium* revealed few genes that segregated clade α and β , but distinct forms of ST64B prophage resulting from recombination that replaced a large block of genes were present in clade α and β , and resulted in the presence of *sseK3* specifically in clade β . The accessory genome contributed significantly to genetic variation that distinguished third order subclades in both clade α and β , especially phage and plasmid genes. Non-phage chromosomal genes exhibited relatively little clade specific accessory genome suggesting that the majority was the result of deletions or gene acquisition on small mobile genetic elements that were neutral and subsequently lost, as observed previously in *S. Paratyphi* [57]. However, three large genetic elements were acquired on the chromosome in α 15 (DT104) or α 17 (monophasic *S.*

Typhimurium ST34), the two most recent dominant MDR pandemic clonal groups that together account for over half of all *S. Typhimurium* infections in the human population Europe in the past 30 years. The acquired genes corresponded to SGI1 [36] in the DT104 complex, and SGI4 and a composite transposon in monophasic *S. Typhimurium* ST34 [18, 37], highlighting the likely importance of horizontal gene transfer in the emergence of epidemic clones.

Variable prophage repertoires are a major source of genetic diversity in *Salmonella* [82, 83], and may contribute to the emergence and spread by impacting the fitness during intra-niche competition due to lytic killing or lysogenic conversion of competing strains [84]. This view was supported by the considerable phage-associated gene flux observed in *S. Typhimurium*. Importantly, the phage component of the accessory genome in *S. Typhimurium* had a strong correlation with the third level clade, suggesting that although transfer was frequent, the acquisition or loss of prophage elements was not transient, consistent with selection within each clonal group [57]. Reconstruction of the evolutionary history of prophage elements in the main *S. Typhimurium* phylogroup indicated that the common ancestor likely contained six prophage, Gifsy1, Gifsy2, Fels1, ST64B, sal3 and BcepMu, that were well conserved during subsequent diversification. Just two of these ancestral prophage, Fels1 and Gifsy1, were lost from the genome, on three occasions in different lineages. The majority of the prophage flux was from the acquisition of between one to three prophage in each lineage, with the exception of a lineage containing clade β 3 (DT2), that only contained the ancestral prophage repertoire.

Together, our analyses are consistent with the view that the common ancestor of the main *S. Typhimurium* phylogroup was a broad host range pathogen with little genome degradation capable of circulating within multiple species of livestock. The age of the common ancestor of *S. Typhimurium* is not known and previous attempts to calculate this using Bayesian approaches have been frustrated by a weak molecular clock signal [85]. However, the common ancestor of *S. Paratyphi* A was estimated to have existed approximately 500 years ago and provides a frame of reference [57]. The main *S. Typhimurium* phylogroup exhibited greater genetic diversity than reported for *S. Paratyphi* A, with an estimated maximum root to tip SNP accumulation of approximately 750 and 250, respectively. Therefore, the common ancestor of the main *S. Typhimurium* phylogroup is likely to have existed earlier than that of *S. Paratyphi* A. However, the *S. Typhimurium* most recent common ancestor (MRCA) is unlikely to have predated the domestication of livestock, that began around ten thousand years ago [86], raising the possibility that the emergence of this phylogroup was linked to the anthropogenic selection provided by entry into a niche within the domesticated livestock. Subsequent to the emergence of the common ancestor of this phylogroup, a single lineage appears to have spawned multiple lineages, some of which have become highly host adapted to various wild avian species, by a process of convergent evolution with that observed in host adapted serotypes of *Salmonella* such as *S. Typhi*.

Materials and methods

Bacterial strains and culture

S. Typhimurium isolates and Illumina short read sequence used in this study have been described previously [87], selected based on phage type determined during routine surveillance by Public Health England (PHE) and the Animal and Plant Health Agency (APHA) in order to represent the diversity *S. Typhimurium* phage types as a proxy for genetic diversity. A strain collection of 134 *S. Typhimurium* or monophasic variant isolates was composed of 2 to 6 randomly selected strains from the top ten most frequent phage types from PHE and the top 20 most frequent phage types from APHA surveillance, from 1990–2010 (2–5 strains of each)

were used in this analysis. In addition, commonly used lab strain SL1344 [88], two reference strains of ST313 (D23580 and A130)[89] and three DT2 strains isolated from pigeon [90] were included. For routine culture, bacteria were stored in 25% glycerol at -80°C and recovered by culture on Luria Bertani agar plates, and single colonies were selected to inoculate LB broth that was incubated at 37 °C for 18 hours with shaking.

Short-read *de novo* assembly

Illumina generated fastq files were assembled using an in-house pipeline adapted from that previously described [91]. For each paired end reads, Velvet [92] (1.2.08) was used to generate multiple assemblies varying the k-mer size between 31 and 61 using Velvet Optimiser [93] and selecting the assembly with the longest N50. Assemblies were then improved using Improve_Assembly software [94] that uses SSPACE (version 3.0) [95] and GapFiller (version 1.0) [96] to scaffold and gap-fill. Ragout was used to order contigs [97] based on comparison to the long-read sequences. The finished genomes were then annotated using Prokka (version 1.11) [98].

Long-Read sequencing using Pacbio and sequence assembly

DNA for long-read sequencing on the Pacbio platform was extracted from 10 ml of cultured bacteria as previously described [87]. Data were assembled using version 2.3 of the Pacbio SMRT analysis pipeline (<https://smrt-analysis.readthedocs.io/en/latest/SMRT-Pipe-Reference-Guide-v2.2.0/>). The structure of the initial assembly was checked against a parallel assembly using Miniasm [99] which showed general agreement. The Pacbio best practice for circularizing contigs was followed using Minimus [100] and the chromosomal contiguous sequence in each assembly was re-orientated to begin at the *thrA* gene. Illumina short read sequence data were used to correct for SNPs and indels using iCORN2 (<http://icorn.sourceforge.net/>). The finished sequences were then annotated using Prokka [98].

Phylogenetic reconstruction and population structure analysis

The paired-end sequence files for each strain were mapped to the SL1344 reference genome (FQ312003) [88] using the Rapid haploid variant calling and core SNP phylogeny pipeline SNIPPY (version 3.0) (<https://github.com/tseemann/snippy>). The size of the core genome was determined using snp-sites (version 2.3.3) [101], outputting monomorphic as well as variant sites and only sites containing A,C,T or G. Variant sites were identified and a core genome variation multifasta alignment generated. The core genome of 134 *S. Typhimurium* (3686476 nucleotides) (S2 Table) contained 17823 variant sites. The core genome (3739972 nucleotides) of 131 *S. Typhimurium* non-ST36 contained 8382 variant sites. The sequence alignment of variant sites was used to generate a maximum likelihood phylogenetic tree with RAxML using the GTRCAT model implemented with an extended majority-rule consensus tree criterion [102]. The genome sequence of *S. Heidelberg* (NC_011083.1) was used as an outgroup in the analysis to identify the root and common ancestor of all *S. Typhimurium* strains as a previous study has indicated both Heidelberg and Saintpaul serovars to be appropriate [103]. HierBaps (hierarchical Bayesian analysis of Population Structure) [16] was used to estimate population structure using three nested levels of molecular variation and 10 independent runs of the optimization algorithm as reported previously [104]. The input for this analysis was the same SNP variant matrix for the 131 strains with reference to SL1344 that was used to generate the GTRCAT phylogeny above. Estimation of coverage for our dataset was by determination of CCs with fewer than 100 allelic variants in 1679 whole genome sequences from human clinical infections in the UK in 2014 to 2015, and approximately 20,000 non-UK *S. Typhimurium*

whole genome sequences for which SRA accession numbers were available (accessed April 2020)[24]. The union of CCs and the proportion of sequences was reported.

***In-silico* genotyping**

The presence of antibiotic resistance, virulence and plasmid replicon genes in short-read data was determined by the mapping and local assembly of short reads to data-bases of candidate genes using ARIBA [105]. The presence of candidate genes from the resfinder [106], VFDB [107] and PlasmidFinder [108] databases was determined. Reads were mapped to candidate genes using nucmer with a 90% minimum alignment identity. This tool was also used to determine the presence of specific genes or gene allelic variants. The results of the ARIBA determination of the presence or absence of the *lpfD* gene were confirmed using SRST2 [109] setting each alternative form of the gene as a potential allele. SRST2 was also used to verify the ARIBA findings of the VFDB data set, as the presence of orthologous genes in the genome was found to confound the interpretation of results. Candidate SNPs in key genes associated with AMR were determined from the SNP matrix created by SNIPPY and used for the phylogeny reconstruction. Sites were then verified by interrogation of the assembled and annotated short-read sequences.

Determination of hypothetically disrupted coding sequences (HDCS)

HDCS were identified in high-quality finished Pacbio sequences and previously published reference sequences by identifying putative altered open reading frames using the RATT annotation transfer tool [110]. The *S. Typhimurium* strains SL1344 annotation (accession no. FQ312003) was transferred to each assembled sequence and coding sequences identified as having altered length were manually curated by comparison of aligned sequences visualised using Artemis comparison tool (ACT) [111]. Genes that contained either a premature stop codon or a frameshift mutation were classified as HDCS. The identified HDCS were used to construct a database that could be used as a reference for SRST2 (above) to detect presence or absence in short-read sequence data. Alleles were called based on matching to 99% sequence identity and allowing one miss-match per 1000 nucleotides.

Determination of Delta Bitscore

Illumina short-read sequences were mapped to the SL1344 reference genome and annotated using PROKKA and then analysed in a pairwise fashion against SL1344 using delta-bit-score (DBS), a profile hidden Markov model based approach [33] with Pfam hidden Markov Models (HMMs) [112]. The mean DBS per genome and percentage of genes with mutations in Pfam domains (non-zero DBS) are reported.

Determination of Invasiveness Index

The invasiveness index [34] for each strain was calculated to scan for patterns of mutation accumulation common to *Salmonella* lineages adapted to an invasive lifestyle. To calculate the invasiveness index, Illumina reads were mapped to a core-genome reference using the snippy pipeline above, and annotated using PROKKA. Protein sequences were then screened using phmmer from the HMMER3.0 package [113] to identify the closest homologs to the 196 predictive genes used by the invasiveness index model. Missing gene sequences were set to NA, to account for the possibilities of being missed during sequencing or misannotated. These genes were then scored against profile hidden Markov models (HMMs) for these protein families from the EggnoG database [34] using hmmsearch [113], to test for uncharacteristic patterns of

sequence variation. Bitscores produced in the comparison of each protein sequence to its respective protein family HMM were then used as input to the model.

Phage location and cargo in assembled long-read sequence

The location of prophage elements in assembled long-read sequences and published reference genome was determined using PHASTER [114], which identified regions as being intact, questionable or incomplete. This yielded a total of 83 potential complete and partial sequences across the 11 representative strains. Prophage sequences were annotated using PROKKA that identified terminase, tail fibre, recombinase/integrase proteins capsid proteins, phage related proteins, and hypothetical proteins.

Determination of recombination

Recombination was inferred by identifying regions of high SNP density from whole genome alignments of short-read data to SL1344, using Gubbins [115]. The results were visualised using Phandango [116] and related to the predicted prophage locations in the SL1344 genome. Similar results were obtained using maximum likelihood inference using clonal frame [117].

Determination of the *S. Typhimurium* pangenome

The annotated assemblies of 131 predominantly *S. Typhimurium* ST19 isolates were used as the input to the pangenome pipeline ROARY [118]. The presence or absence of genes was determined without splitting orthologues. In order to characterise the contribution of prophage and plasmids to the pangenome, genes were assigned to one of four categories, non-prophage genes located on the chromosome, prophage genes, plasmid genes and undefined genes, based on the similarity to annotated genes of complete and closed whole genome sequence of eleven reference strains. Orthologous genes were identified based on > 90% nucleotide sequence identity using nucmer [119]. A core-genome reference sequence (genes present in at least 99% of reference strains), was also constructed and used to determine the invasiveness index.

Estimation of gene flux rates

Genes were assigned a score based on their presence in strains within a specific clade. The clade gene score was compared to the score determined for strains outside of the clade to determine whether the gene was more prevalent within the clade than without. Genes were classed as associated with the clade if their score was greater than the mean plus two standard deviations of the non-cladal score (corresponding to the top 95% in a normal distribution).

The number of clade associate genes was compared with the number of SNPs associated with a clade (this gives a measure of evolutionary time) to determine the level of gene flux for the clade. The level of gene flux in the two first-level clades was then compared.

Prophage classification

PHASTER[114] curated prophage sequences were classified into species and genus-level groupings based on the current criteria used by the Bacterial and Archaeal Viruses Subcommittee of the International Committee on Taxonomy of Viruses (ICTV)[120]. At the species level, genomes were clustered at 95% nucleotide identity over the whole genome length, meaning that two genomes belong to two different species if they differ in more than 5% of their genome. Clustering was performed with CD-HIT-EST at 95% nucleotide identity over 95% of the alignment length (99% of alignment length of shorter sequence)[121] and with Gegenees, a

pairwise nucleotide comparison tool, using the accurate settings of 200 bp fragment size and 100 bp step size[122]. The Gegenees output was used in combination with vConTACT2[123] to classify the prophage sequences into new or existing genera. Briefly, coding sequences were predicted with PROKKA[98] and transformed into a table linking genomes and their encoding proteins. This table was used as input into vConTACT along with the Viral RefSeq database v85[124]. vConTACT then used Diamond[125], Markov clustering MCL[126] and Cluster-ONE[127] to predict viral clusters based on shared protein content. The output was visualised using Cytoscape[128]. Genera were defined as vConTACT viral clusters which shared a significant (>50%) nucleotide identity. The clusters were then compared to the current and pending ICTV taxonomic classification (ictvonline.org) using blast and vConTACT viral cluster output containing reference genomes and all prophage sequences were assigned to new or existing taxa.

Assembly and comparison of pSLT plasmid sequence

Plasmid sequences were determined by assembling Illumina short-read sequences with spades-3.8.0 using the plasmidSPAdes algorithm [129] with varying with k-mer sizes of 31, 41 and 51. Contigs larger than 70 kb were then compared against the NCBI blast database to identify forms of pSLT.

Determination of the distribution of representatives among the rest of the world

A collection of 14,478 genomes from all available *S. Typhimurium* in EnteroBase to represent the global diversity outside of the United Kingdom. These genomes met several criteria; they were predicted to be serovar Typhimurium using SISTR software [130], were not ST36, the country of isolation was listed as a country other than the "United Kingdom", and the genome was within the HierCC:1100 (cEBG) '2' cluster group (S7 Table). These genomes were grouped according to their EnteroBase HierCC:100 cluster definition, which groups genomes together if the nearest neighbour is no more 100 cgMLST alleles different [24]. We then included representative genomes from this study and determined the proportion of defined cluster groups contained at least one of the representative genomes described here.

Supporting information

S1 Fig. Phylogenetic relationship of *S. Typhimurium* and diverse *S. enterica* serotypes.

Mid-point rooted maximum likelihood phylogenetic tree based on the variation (SNPs) in the core genome of 18 strains of *Salmonella* Typhimurium and 14 representative strain of diverse *S. enterica* subspecies *enterica* serotypes, with reference to *S. Typhimurium* strain SL1344 genome sequence. *S. Typhimurium* strains (red lineages and text) are present in two clusters, composed of 15 strains with isolates that are ST19, ST34, ST313, ST98 and ST568 and three more divergent isolates of ST36. The phylogeny is rooted with respect to *S. Heidelberg* and was calculated using SL1344 as a reference to create a core-genome variant-site alignment and the GTRCAT model in RAxML.

(PDF)

S2 Fig. Distribution of 131 non ST36 isolates in this study in UK clinical isolates and non-UK isolates from enterobase. Grapetree visualization based on EnteroBase cgMLST allele profiles, including of (A) 1,693 *S. Typhimurium* isolates from clinical infections in the UK in 2014 and 2015, and (B) 14,760 genomes selected as the global diversity of *S. Typhimurium* outside of the United Kingdom in Enterobase database. Nodes are colour coded by EnteroBase

HierCC HC100 cluster groups. HierCC groups containing non ST36 isolates in this study yellow circles. Scale indicates number of cgMLST alleles.

(PDF)

S3 Fig. Sequence variation in selected pathogenicity islands. The percent sequence variation including SNPs and deletions in SPI-1, SPI-2, SPI-3, SPI-4, SPI-5 and SPI-6 are indicated from 0% (green) to 0.5% (red).

(PDF)

S4 Fig. Accessory genome with strong clade association. Gene families with a strong clade association in clade α or β (A), or in one of the third level clades (B). Maximum likelihood phylogenetic tree and based on sequence variation (SNPs) in the core genome with reference to *S. Typhimurium* strain SL1344 (left). Third-level clades are indicated and colour coordinated with that in Fig 1. Genes in each clade were assigned a score based on the number of strains containing the gene within the clade. This score was also calculated for the strains outside the clade. Clade associated genes were defined as genes that had scores greater than the mean plus two SD of the score for all other clades. Genes are colour coded based assignment to non-prophage chromosomal (red), prophage (green), plasmid (blue), or undefined (grey).

(PDF)

S5 Fig. Gene flux rate metrics determined for non-singleton gene families first-level clades.

(PDF)

S1 Table. *S. Typhimurium* strain collection used in this study related to determine population structures. Table can be viewed at <https://www.dropbox.com/sh/dh84yyc4tguirw3/AABnbbrSPtqEcjGY6IR6GggLa?dl=0>.

(XLSX)

S2 Table. Presence of plasmid operons, AMR genes (Resfinder) and virulence genes (VFDB) determined by *in-silico* genotyping using ARIBA software. The identifier column corresponds to the study-identifier column in S1 Table. Presence of genes are indicated by '1'. Table can be viewed at <https://www.dropbox.com/sh/dh84yyc4tguirw3/AABnbbrSPtqEcjGY6IR6GggLa?dl=0>.

(XLSX)

S3 Table. Genome degradation in long-read reference strains. Potential hypothetically disrupted coding sequences (HDCS) were identified in reference genomes through anomalies in annotation transfer from the SL1344 reference sequence using RATT software and manual curation to exclude false positive HDCS. Table can be viewed at <https://www.dropbox.com/sh/dh84yyc4tguirw3/AABnbbrSPtqEcjGY6IR6GggLa?dl=0>.

(XLSX)

S4 Table. Summary of HDCS in *S. Typhimurium* main phylogroup. The presence of HDCS alleles was determined *in-silico* using SRST2 reported in Fig 1. The study identifier refers to isolates in S1 Table. Alleles are specified as wild-type (WT) or HDCS. In some cases, multiple HDCS forms were determined to be present and these are denoted as HDCS1 or HDCS2 etc. Table can be viewed at <https://www.dropbox.com/sh/dh84yyc4tguirw3/AABnbbrSPtqEcjGY6IR6GggLa?dl=0>.

(XLSX)

S5 Table. Summary of DBS and Invasiveness Index analysis. Isolate identifier corresponds to the identifier column in S1 Table. The mean delta bitscore (DBS) for the proteome of each strain, the number of proteins with a DBS greater than ten. Invasiveness index for each isolate

is indicated.
(XLSX)

S6 Table. Characteristics of prophage elements present in complete and closed whole genome sequence of *S. Typhimurium* reference strains.
(DOCX)

S7 Table. Data used for construction of HierCC clusters with 14,478 genomes from *S. Typhimurium* in Enterobase to represent the global diversity outside of the United Kingdom.
(XLSX)

Acknowledgments

The authors also acknowledge advice and informatics support from Andrew Page and Andrea Telatin from the Quadram Institute Bioscience informatics support group.

Author Contributions

Conceptualization: Matt Bawn, Robert A. Kingsley.

Formal analysis: Matt Bawn, Nabil-Fareed Alikhan, Gaëtan Thilliez, Mark Kirkwood, Nicole E. Wheeler, Evelien M. Adriaenssens, Robert A. Kingsley.

Funding acquisition: Neil Hall, Robert A. Kingsley.

Investigation: Nabil-Fareed Alikhan.

Methodology: Timothy J. Dallman.

Resources: Liljana Petrovska, Timothy J. Dallman, Neil Hall, Robert A. Kingsley.

Supervision: Neil Hall, Robert A. Kingsley.

Visualization: Nabil-Fareed Alikhan.

Writing – original draft: Matt Bawn, Evelien M. Adriaenssens, Robert A. Kingsley.

Writing – review & editing: Nabil-Fareed Alikhan, Gaëtan Thilliez, Mark Kirkwood, Nicole E. Wheeler, Liljana Petrovska, Timothy J. Dallman, Evelien M. Adriaenssens, Neil Hall, Robert A. Kingsley.

References

1. Kingsley R, Bäumlér J. Host adaptation and the emergence of infectious disease: the *Salmonella* paradigm. *Mol Micro.* 2000; 36. <https://doi.org/10.1046/j.1365-2958.2000.01907.x> PMID: 10844686
2. Kirk MD, Pires SM, Black RE, Caipo M, Crump JA, Devleeschauwer B, et al. World Health Organization Estimates of the Global and Regional Disease Burden of 22 Foodborne Bacterial, Protozoal, and Viral Diseases, 2010: A Data Synthesis. *PLoS Med.* 2015; 12(12):e1001921. Epub 2015/12/04. <https://doi.org/10.1371/journal.pmed.1001921> PMID: 26633831; PubMed Central PMCID: PMC4668831.
3. Rabsch W, Andrews HL, Kingsley RA, Prager R, Tschape H, Adams LG, et al. *Salmonella enterica* serotype Typhimurium and its host-adapted variants. *Infect Immun.* 2002; 70(5):2249–55. <https://doi.org/10.1128/iai.70.5.2249-2255.2002> PMID: 11953356
4. Branchu P, Bawn M, Kingsley RA. Genome variation and molecular epidemiology of *Salmonella* Typhimurium pathovariants. *Infect Immun.* 2018; 86(8):e00079–18. Epub 2018/05/23. <https://doi.org/10.1128/IAI.00079-18> PMID: 29784861.
5. Anonymous. *Salmonella* in livestock production in Great Britain, 2017: gov.uk; 2018 [cited 2019 June 2019]. Available from: <https://www.gov.uk/government/publications/salmonella-in-livestock-production-in-great-britain-2017>.

6. Rabsch W. Salmonella Typhimurium Phage Typing for Pathogens. In: Schatten H, Eisenstark A, editors. *Salmonella, Methods and Protocols. Methods in Molecular Biology*. 394 ed. Totowa, New Jersey: Humana Press; 2007. p. 177–212.
7. Threlfall EJ, Ward LR, Rowe B. Spread of multiresistant strains of *Salmonella typhimurium* phage types 204 and 193 in Britain. *Br Med J*. 1978; 2(6143):997.
8. Rabsch W, Tschape H, Baumberg AJ. Non-typhoidal salmonellosis: emerging problems. *Microbes Infect*. 2001; 3(3):237–47. [https://doi.org/10.1016/s1286-4579\(01\)01375-2](https://doi.org/10.1016/s1286-4579(01)01375-2) PMID: 11358718.
9. Rabsch W, Truepschuch S, Windhorst D, Gerlach RG. Typing phages and prophages of *Salmonella*. Norfolk, UK: Caister Academic Press; 2011.
10. Tassinari E, Duffy G, Bawn M, Burgess CM, McCabe EM, Lawlor PG, et al. Microevolution of antimicrobial resistance and biofilm formation of *Salmonella Typhimurium* during persistence on pig farms. *Sci Rep*. 2019; 9(1):8832. Epub 2019/06/22. <https://doi.org/10.1038/s41598-019-45216-w> PMID: 31222015; PubMed Central PMCID: PMC6586642.
11. Ashton PM, Peters T, Ameh L, McAleer R, Petrie S, Nair S, et al. Whole Genome Sequencing for the Retrospective Investigation of an Outbreak of *Salmonella Typhimurium* DT 8. *PLoS Curr*. 2015; 7. Epub 2015/02/26. <https://doi.org/10.1371/currents.outbreaks.2c05a47d292f376afc5a6fcd8a7a3b6> PMID: 25713745; PubMed Central PMCID: PMC4336196.
12. Mather AE, Lawson B, de Pinna E, Wigley P, Parkhill J, Thomson NR, et al. Genomic Analysis of *Salmonella enterica* Serovar Typhimurium from Wild Passerines in England and Wales. *Appl Environ Microbiol*. 2016; 82(22):6728–35. Epub 2016/10/30. <https://doi.org/10.1128/AEM.01660-16> PMID: 27613688; PubMed Central PMCID: PMC5086566.
13. Hughes LA, Shopland S, Wigley P, Bradon H, Leatherbarrow AH, Williams NJ, et al. Characterisation of *Salmonella enterica* serotype Typhimurium isolates from wild birds in northern England from 2005–2006. *BMC Vet Res*. 2008; 4:4. Epub 2008/01/31. <https://doi.org/10.1186/1746-6148-4-4> PMID: 18230128; PubMed Central PMCID: PMC2257933.
14. Kingsley RA, Msefula CL, Thomson NR, Kariuki S, Holt KE, Gordon MA, et al. Epidemic multiple drug resistant *Salmonella typhimurium* causing invasive disease in sub-Saharan Africa have a distinct genotype. *Genome Res*. 2009; 19. <https://doi.org/10.1101/gr.091017.109> PMID: 19901036
15. Feasey NA, Dougan G, Kingsley RA, Heyderman RS, Gordon MA. Invasive non-typhoidal salmonella disease: an emerging and neglected tropical disease in Africa. *Lancet*. 2012; 379(9835):2489–99. Epub 2012/05/17. [https://doi.org/10.1016/S0140-6736\(11\)61752-2](https://doi.org/10.1016/S0140-6736(11)61752-2) PMID: 22587967; PubMed Central PMCID: PMC3402672.
16. Cheng L, Connor TR, Siren J, Aanensen DM, Corander J. Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. *Mol Biol Evol*. 2013; 30(5):1224–8. Epub 2013/02/15. <https://doi.org/10.1093/molbev/mst028> PMID: 23408797; PubMed Central PMCID: PMC3670731.
17. Hooton SP, Atterbury RJ, Connerton IF. Application of a bacteriophage cocktail to reduce *Salmonella Typhimurium* U288 contamination on pig skin. *Int J Food Microbiol*. 2011; 151(2):157–63. Epub 2011/09/09. <https://doi.org/10.1016/j.ijfoodmicro.2011.08.015> PMID: 21899907.
18. Petrovska L, Mather AE, AbuOun M, Branchu P, Harris SR, Connor T, et al. Microevolution of monophasic *Salmonella Typhimurium* during epidemic, United Kingdom, 2005–2010. *Emerging infectious diseases*. 2016; 22(4):617. <https://doi.org/10.3201/eid2204.150531> PMID: 26982594
19. Mather AE, Reid SWJ, Maskell DJ, Parkhill J, Fookes MC, Harris SR, et al. Distinguishable Epidemics of Multidrug-Resistant *Salmonella Typhimurium* DT104 in Different Hosts. *Science*. 2013; 341(6153):1514–7. <https://doi.org/10.1126/science.1240578> WOS:000324894600051. PMID: 24030491
20. Leekitcharoenphon P, Hendriksen RS, Le Hello S, Weill FX, Baggesen DL, Jun SR, et al. Global Genomic Epidemiology of *Salmonella enterica* Serovar Typhimurium DT104. *Appl Environ Microbiol*. 2016; 82(8):2516–26. Epub 2016/03/06. <https://doi.org/10.1128/AEM.03821-15> PMID: 26944846; PubMed Central PMCID: PMC4959494.
21. Kingsley RA, Kay S, Connor T, Barquist L, Sait L, Holt KE, et al. Genome and Transcriptome Adaptation Accompanying Emergence of the Definitive Type 2 Host-Restricted *Salmonella enterica* Serovar Typhimurium Pathovar. *mBio*. 2013; 4(5). <https://doi.org/10.1128/mBio.00565-13> PMID: 23982073
22. 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(11):1215–21. <https://doi.org/10.1038/ng.2423> PMID: 23023330; PubMed Central PMCID: PMC3491877.
23. Branchu P, Charity O, Bawn M, Thilliez G, Dallman TJ, Petrovska L, et al. SGI-4 in monophasic *Salmonella Typhimurium* ST34 is a novel ICE that enhances resistance to copper. *Frontiers in microbiology*. 2019; 10:1118. <https://doi.org/10.3389/fmicb.2019.01118> PMID: 31178839
24. Zhou Z, Alikhan NF, Mohamed K, Fan Y, Agama Study G, Achtman M. The Enterobase user's guide, with case studies on *Salmonella* transmissions, *Yersinia pestis* phylogeny, and *Escherichia coli*

- genomic diversity. *Genome Res.* 2020; 30(1):138–52. Epub 2019/12/07. <https://doi.org/10.1101/gr.251678.119> PMID: 31809257; PubMed Central PMCID: PMC6961584.
25. Turner AK, Nair S, Wain J. The acquisition of full fluoroquinolone resistance in *Salmonella* Typhi by accumulation of point mutations in the topoisomerase targets. *Journal of Antimicrobial Chemotherapy.* 2006; 58(4):733–40. RefWorks:212. <https://doi.org/10.1093/jac/dkl333> PMID: 16895934
 26. Garcia V, Montero I, Bances M, Rodicio R, Rodicio MR. Incidence and Genetic Bases of Nitrofurantoin Resistance in Clinical Isolates of Two Successful Multidrug-Resistant Clones of *Salmonella enterica* Serovar Typhimurium: Pandemic "DT 104" and pUO-StVR2. *Microb Drug Resist.* 2017; 23(4):405–12. Epub 2016/11/05. <https://doi.org/10.1089/mdr.2016.0227> PMID: 27809653.
 27. Oliva M, Monno R, D'Addabbo P, Pesole G, Dionisi AM, Scrascia M, et al. A novel group of IncQ1 plasmids conferring multidrug resistance. *Plasmid.* 2017; 89:22–6. Epub 2016/12/06. <https://doi.org/10.1016/j.plasmid.2016.11.005> PMID: 27916622.
 28. Lobato-Marquez D, Molina-Garcia L, Moreno-Cordoba I, Garcia-Del Portillo F, Diaz-Orejas R. Stabilization of the Virulence Plasmid pSLT of *Salmonella* Typhimurium by Three Maintenance Systems and Its Evaluation by Using a New Stability Test. *Frontiers in molecular biosciences.* 2016; 3:66. Epub 2016/11/02. <https://doi.org/10.3389/fmolb.2016.00066> PMID: 27800482; PubMed Central PMCID: PMC5065971.
 29. Hooton SP, Timms AR, Cummings NJ, Moreton J, Wilson R, Connerton IF. The complete plasmid sequences of *Salmonella enterica* serovar Typhimurium U288. *Plasmid.* 2014; 76:32–9. Epub 2014/09/02. <https://doi.org/10.1016/j.plasmid.2014.08.002> PMID: 25175817.
 30. Ilyas B, Tsai CN, Coombes BK. Evolution of *Salmonella*-Host Cell Interactions through a Dynamic Bacterial Genome. *Frontiers in cellular and infection microbiology.* 2017; 7:428. Epub 2017/10/17. <https://doi.org/10.3389/fcimb.2017.00428> PMID: 29034217; PubMed Central PMCID: PMC5626846.
 31. Gerlach RG, Jackel D, Stecher B, Wagner C, Lupas A, Hardt WD, et al. *Salmonella* Pathogenicity Island 4 encodes a giant non-fimbrial adhesin and the cognate type 1 secretion system. *Cell Microbiol.* 2007; 9(7):1834–50. <https://doi.org/10.1111/j.1462-5822.2007.00919.x> PMID: 17388786.
 32. Sana TG, Flaugnatti N, Lugo KA, Lam LH, Jacobson A, Baylot V, et al. *Salmonella* Typhimurium utilizes a T6SS-mediated antibacterial weapon to establish in the host gut. *Proc Natl Acad Sci U S A.* 2016; 113(34):E5044–51. Epub 2016/08/10. <https://doi.org/10.1073/pnas.1608858113> PMID: 27503894; PubMed Central PMCID: PMC5003274.
 33. Wheeler NE, Barquist L, Kingsley RA, Gardner PP. A profile-based method for identifying functional divergence of orthologous genes in bacterial genomes. *Bioinformatics.* 2016; 32(23):3566–74. Epub 2016/08/10. <https://doi.org/10.1093/bioinformatics/btw518> PMID: 27503221; PubMed Central PMCID: PMC5181535.
 34. Wheeler NE, Gardner PP, Barquist L. Machine learning identifies signatures of host adaptation in the bacterial pathogen *Salmonella enterica*. *PLoS genetics.* 2018; 14(5):e1007333–e. <https://doi.org/10.1371/journal.pgen.1007333> PMID: 29738521.
 35. Van Puyvelde S, Pickard D, Vandelannoote K, Heinz E, Barbe B, de Block T, et al. An African *Salmonella* Typhimurium ST313 sublineage with extensive drug-resistance and signatures of host adaptation. *Nat Commun.* 2019; 10(1):4280. Epub 2019/09/21. <https://doi.org/10.1038/s41467-019-11844-z> PMID: 31537784; PubMed Central PMCID: PMC6753159.
 36. Boyd D, Peters GA, Cloeckeaert A, Boumedine KS, Chasus-Dancla E, Imberechts H, et al. Complete nucleotide sequence of a 43-kilobase genomic island associated with the multidrug resistance region of *Salmonella enterica* serovar Typhimurium DT104 and its identification in phage type DT120 and serovar Agona. *J Bacteriol.* 2001; 183(19):5725–32. Epub 2001/09/07. <https://doi.org/10.1128/JB.183.19.5725-5732.2001> PMID: 11544236; PubMed Central PMCID: PMC95465.
 37. Branchu P, Charity OJ, Bawn M, Thilliez G, Dallman TJ, Petrovska L, et al. SGI-4 in Monophasic *Salmonella* Typhimurium ST34 Is a Novel ICE That Enhances Resistance to Copper. *Front Microbiol.* 2019; 10:1118. Epub 2019/06/11. <https://doi.org/10.3389/fmicb.2019.01118> PMID: 31178839; PubMed Central PMCID: PMC6543542.
 38. Owen SV, Wenner N, Canals R, Makumi A, Hammarlof DL, Gordon MA, et al. Characterization of the Prophage Repertoire of African *Salmonella* Typhimurium ST313 Reveals High Levels of Spontaneous Induction of Novel Phage BTP1. *Front Microbiol.* 2017; 8:235. <https://doi.org/10.3389/fmicb.2017.00235> PMID: 28280485; PubMed Central PMCID: PMC5322425.
 39. Summer EJ, Gonzalez CF, Carlisle T, Mebane LM, Cass AM, Savva CG, et al. Burkholderia cenocepacia phage BcepMu and a family of Mu-like phages encoding potential pathogenesis factors. *J Mol Biol.* 2004; 340(1):49–65. Epub 2004/06/09. <https://doi.org/10.1016/j.jmb.2004.04.053> PMID: 15184022.
 40. Gymoese P, Sorensen G, Litrup E, Olsen JE, Nielsen EM, Torpdahl M. Investigation of Outbreaks of *Salmonella enterica* Serovar Typhimurium and Its Monophasic Variants Using Whole-Genome

- Sequencing, Denmark. *Emerg Infect Dis.* 2017; 23(10):1631–9. Epub 2017/09/21. <https://doi.org/10.3201/eid2310.161248> PMID: 28930002; PubMed Central PMCID: PMC5621559.
41. Sun J, Ke B, Huang Y, He D, Li X, Liang Z, et al. The molecular epidemiological characteristics and genetic diversity of salmonella typhimurium in Guangdong, China, 2007–2011. *PLoS One.* 2014; 9(11):e113145. Epub 2014/11/08. <https://doi.org/10.1371/journal.pone.0113145> PMID: 25380053; PubMed Central PMCID: PMC4224511.
 42. Alikhan NF, Zhou Z, Sergeant MJ, Achtman M. A genomic overview of the population structure of *Salmonella*. *PLoS Genet.* 2018; 14(4):e1007261. Epub 2018/04/06. <https://doi.org/10.1371/journal.pgen.1007261> PMID: 29621240; PubMed Central PMCID: PMC5886390.
 43. Lan R, Reeves PR, Octavia S. Population structure, origins and evolution of major *Salmonella enterica* clones. *Infect Genet Evol.* 2009; 9(5):996–1005. Epub 2009/04/28. <https://doi.org/10.1016/j.meegid.2009.04.011> PMID: 19393770.
 44. Zhang S, Li S, Gu W, den Bakker H, Boxrud D, Taylor A, et al. Zoonotic Source Attribution of *Salmonella enterica* Serotype Typhimurium Using Genomic Surveillance Data, United States. *Emerg Infect Dis.* 2019; 25(1):82–91. Epub 2018/12/19. <https://doi.org/10.3201/eid2501.180835> PMID: 30561314; PubMed Central PMCID: PMC6302586.
 45. Mather AE, Phuong TLT, Gao Y, Clare S, Mukhopadhyay S, Goulding DA, et al. New Variant of Multi-drug-Resistant *Salmonella enterica* Serovar Typhimurium Associated with Invasive Disease in Immunocompromised Patients in Vietnam. *mBio.* 2018; 9(5). Epub 2018/09/06. <https://doi.org/10.1128/mBio.01056-18> PMID: 30181247; PubMed Central PMCID: PMC6123440.
 46. Van Boeckel TP, Brower C, Gilbert M, Grenfell BT, Levin SA, Robinson TP, et al. Global trends in antimicrobial use in food animals. *Proc Natl Acad Sci U S A.* 2015; 112(18):5649–54. Epub 2015/03/21. <https://doi.org/10.1073/pnas.1503141112> PMID: 25792457; PubMed Central PMCID: PMC4426470.
 47. McEwen SA, Fedorka-Cray PJ. Antimicrobial use and resistance in animals. *Clin Infect Dis.* 2002; 34 Suppl 3:S93–S106. Epub 2002/05/04. <https://doi.org/10.1086/340246> PMID: 11988879.
 48. Horton RA, Wu G, Speed K, Kidd S, Davies R, Coldham NG, et al. Wild birds carry similar *Salmonella enterica* serovar Typhimurium strains to those found in domestic animals and livestock. *Res Vet Sci.* 2013; 95(1):45–8. <https://doi.org/10.1016/j.rvsc.2013.02.008> PMID: 23481141.
 49. Parsons BN, Humphrey S, Salisbury AM, Mikoleit J, Hinton JC, Gordon MA, et al. Invasive non-typhoidal *Salmonella typhimurium* ST313 are not host-restricted and have an invasive phenotype in experimentally infected chickens. *PLoS Negl Trop Dis.* 2013; 7(10):e2487. <https://doi.org/10.1371/journal.pntd.0002487> PMID: 24130915; PubMed Central PMCID: PMC3794976.
 50. Okoro CK, Barquist L, Connor TR, Harris SR, Clare S, Stevens MP, et al. Signatures of adaptation in human invasive *Salmonella typhimurium* ST313 populations from sub-Saharan Africa. *PLoS Negl Trop Dis.* 2015; 9(3):e0003611. Epub 2015/03/25. <https://doi.org/10.1371/journal.pntd.0003611> PMID: 25803844; PubMed Central PMCID: PMC4372345.
 51. Carden SE, Walker GT, Honeycutt J, Lugo K, Pham T, Jacobson A, et al. Pseudogenization of the Secreted Effector Gene *ssel* Confers Rapid Systemic Dissemination of *S. typhimurium* ST313 within Migratory Dendritic Cells. *Cell Host Microbe.* 2017; 21(2):182–94. Epub 2017/02/10. <https://doi.org/10.1016/j.chom.2017.01.009> PMID: 28182950; PubMed Central PMCID: PMC5325708.
 52. Singletary LA, Karlinsey JE, Libby SJ, Mooney JP, Lokken KL, Tsoilis RM, et al. Loss of Multicellular Behavior in Epidemic African Nontyphoidal *Salmonella enterica* Serovar Typhimurium ST313 Strain D23580. *mBio.* 2016; 7(2):e02265. <https://doi.org/10.1128/mBio.02265-15> PMID: 26933058; PubMed Central PMCID: PMC4810497.
 53. Feasey NA, Hadfield J, Keddy KH, Dallman TJ, Jacobs J, Deng X, et al. Distinct *Salmonella enteritidis* lineages associated with enterocolitis in high-income settings and invasive disease in low-income settings. *Nat Genet.* 2016; 48(10):1211–7. Epub 2016/08/23. <https://doi.org/10.1038/ng.3644> PMID: 27548315; PubMed Central PMCID: PMC5047355.
 54. Abraham S, O'Dea M, Trott DJ, Abraham RJ, Hughes D, Pang S, et al. Isolation and plasmid characterization of carbapenemase (IMP-4) producing *Salmonella enterica* Typhimurium from cats. *Sci Rep.* 2016; 6:35527. Epub 2016/10/22. <https://doi.org/10.1038/srep35527> PMID: 27767038; PubMed Central PMCID: PMC5073282 Neoculi. All other authors have none to declare.
 55. Parkhill J, Dougan G, James KD, Thomson NR, Pickard D, Wain J, et al. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature.* 2001; 413(6858):848–52. <https://doi.org/10.1038/35101607> PMID: 11677608.
 56. Holt KE, Thomson NR, Wain J, Langridge GC, Hasan R, Bhutta ZA, et al. Pseudogene accumulation in the evolutionary histories of *Salmonella enterica* serovars Paratyphi A and Typhi. *BMC Genomics.* 2009; 10:36. <https://doi.org/10.1186/1471-2164-10-36> PMID: 19159446.
 57. Zhou Z, McCann A, Weill FX, Blin C, Nair S, Wain J, et al. Transient Darwinian selection in *Salmonella enterica* serovar Paratyphi A during 450 years of global spread of enteric fever. *Proc Natl Acad Sci U S*

- A. 2014; 111(33):12199–204. Epub 2014/08/06. <https://doi.org/10.1073/pnas.1411012111> PMID: 25092320; PubMed Central PMCID: PMC4143038.
58. Threlfall EJ. Epidemic *Salmonella* typhimurium DT 104—a truly international multiresistant clone. *Journal of Antimicrobial Chemotherapy*. 2000; 46(1):7–10. <https://doi.org/10.1093/jac/46.1.7> WOS:000089125100002. PMID: 10882682
59. Collier-Hyams LS, Zeng H, Sun J, Tomlinson AD, Bao ZQ, Chen H, et al. Cutting edge: *Salmonella* AvrA effector inhibits the key proinflammatory, anti-apoptotic NF-kappa B pathway. *J Immunol*. 2002; 169(6):2846–50. Epub 2002/09/10. <https://doi.org/10.4049/jimmunol.169.6.2846> PMID: 12218096.
60. Yang Z, Soderholm A, Lung TW, Giogha C, Hill MM, Brown NF, et al. SseK3 Is a *Salmonella* Effector That Binds TRIM32 and Modulates the Host's NF-kappaB Signalling Activity. *PLoS One*. 2015; 10(9):e0138529. Epub 2015/09/24. <https://doi.org/10.1371/journal.pone.0138529> PMID: 26394407; PubMed Central PMCID: PMC4579058.
61. Geng S, Wang Y, Xue Y, Wang H, Cai Y, Zhang J, et al. The SseL protein inhibits the intracellular NF-kappaB pathway to enhance the virulence of *Salmonella* Pullorum in a chicken model. *Microb Pathog*. 2019; 129:1–6. Epub 2019/02/01. <https://doi.org/10.1016/j.micpath.2019.01.035> PMID: 30703474.
62. Nuccio S-P, Bäumlér AJ. Comparative Analysis of *Salmonella* Genomes Identifies a Metabolic Network for Escalating Growth in the Inflamed Gut. *mBio*. 2014; 5(2). <https://doi.org/10.1128/mBio.00929-14> PMID: 24643865
63. Johnson R, Mylona E, Frankel G. Typhoidal *Salmonella*: Distinctive virulence factors and pathogenesis. *Cell Microbiol*. 2018; 20(9):e12939. Epub 2018/07/22. <https://doi.org/10.1111/cmi.12939> PMID: 30030897.
64. Kingsley RA, Humphries AD, Weening EH, De Zoete MR, Winter S, Papaconstantinopoulou A, et al. Molecular and phenotypic analysis of the CS54 island of *Salmonella enterica* serotype typhimurium: identification of intestinal colonization and persistence determinants. *Infect Immun*. 2003; 71(2):629–40. <https://doi.org/10.1128/iai.71.2.629-640.2003> PMID: 12540539.
65. Bäumlér AJ, Tsolis RM, Bowe F, Kusters JG, Hoffmann S, Heffron F. The *pef* fimbrial operon mediates adhesion to murine small intestine and is necessary for fluid accumulation in infant mice. *Infect Immun*. 1996; 64:61–8. PMID: 8557375
66. Weening EH, Barker JD, Laarakker MC, Humphries AD, Tsolis RM, Baumler AJ. The *Salmonella enterica* serotype Typhimurium *lpf*, *bcf*, *stb*, *stc*, *std*, and *sth* fimbrial operons are required for intestinal persistence in mice. *Infect Immun*. 2005; 73(6):3358–66. <https://doi.org/10.1128/IAI.73.6.3358-3366.2005> PMID: 15908362.
67. Winter SE, Thiennimitr P, Winter MG, Butler BP, Huseby DL, Crawford RW, et al. Gut inflammation provides a respiratory electron acceptor for *Salmonella*. *Nature*. 2010; 467(7314):426–9. Epub 2010/09/25. <https://doi.org/10.1038/nature09415> [pii] PMID: 20864996; PubMed Central PMCID: PMC2946174.
68. Rivera-Chavez F, Lopez CA, Zhang LF, Garcia-Pastor L, Chavez-Arroyo A, Lokken KL, et al. Energy Taxis toward Host-Derived Nitrate Supports a *Salmonella* Pathogenicity Island 1-Independent Mechanism of Invasion. *mBio*. 2016; 7(4). Epub 2016/07/21. <https://doi.org/10.1128/mBio.00960-16> PMID: 27435462; PubMed Central PMCID: PMC4958259.
69. Bourret TJ, Liu L, Shaw JA, Husain M, Vazquez-Torres A. Magnesium homeostasis protects *Salmonella* against nitrooxidative stress. *Sci Rep*. 2017; 7(1):15083. Epub 2017/11/10. <https://doi.org/10.1038/s41598-017-15445-y> PMID: 29118452; PubMed Central PMCID: PMC5678156.
70. MacKenzie KD, Wang Y, Musicha P, Hansen EG, Palmer MB, Herman DJ, et al. Parallel evolution leading to impaired biofilm formation in invasive *Salmonella* strains. *PLoS Genet*. 2019; 15(6):e1008233. Epub 2019/06/25. <https://doi.org/10.1371/journal.pgen.1008233> PMID: 31233504; PubMed Central PMCID: PMC6611641.
71. Gonzales AM, Wilde S, Roland KL. New Insights into the Roles of Long Polar Fimbriae and Stg Fimbriae in *Salmonella* Interactions with Enterocytes and M Cells. *Infect Immun*. 2017; 85(9). Epub 2017/06/21. <https://doi.org/10.1128/iai.00172-17> PMID: 28630073; PubMed Central PMCID: PMC5563581.
72. Bäumlér AJ, Tsolis RM, Heffron F. The *lpf* fimbrial operon mediates adhesion of *Salmonella typhimurium* to murine Peyer's patches. *Proc Natl Acad Sci USA*. 1996; 93:279–83. <https://doi.org/10.1073/pnas.93.1.279> PMID: 8552622
73. Lloyd SJ, Ritchie JM, Rojas-Lopez M, Blumentritt CA, Popov VL, Greenwich JL, et al. A double, long polar fimbria mutant of *Escherichia coli* O157:H7 expresses Curli and exhibits reduced in vivo colonization. *Infect Immun*. 2012; 80(3):914–20. Epub 2012/01/11. <https://doi.org/10.1128/IAI.05945-11> PMID: 22232190; PubMed Central PMCID: PMC3294650.
74. Ledebøer NA, Frye JG, McClelland M, Jones BD. *Salmonella enterica* Serovar Typhimurium Requires the *Lpf*, *Pef*, and *Tafi* Fimbriae for Biofilm Formation on HEp-2 Tissue Culture Cells and Chicken

- Intestinal Epithelium. *Infection and Immunity*. 2006; 74(6):3156–69. <https://doi.org/10.1128/IAI.01428-05> PMC1479237. PMID: 16714543
75. Nakato G, Fukuda S, Hase K, Goitsuka R, Cooper MD, Ohno H. New approach for m-cell-specific molecules screening by comprehensive transcriptome analysis. *DNA Res*. 2009; 16(4):227–35. Epub 2009/08/14. <https://doi.org/10.1093/dnares/dsp013> PMID: 19675110; PubMed Central PMCID: PMC2725790.
 76. Kozuka Y, Nasu T, Murakami T, Yasuda M. Comparative studies on the secondary lymphoid tissue areas in the chicken bursa of Fabricius and calf ileal Peyer's patch. *Vet Immunol Immunopathol*. 2010; 133(2–4):190–7. Epub 2009/09/09. <https://doi.org/10.1016/j.vetimm.2009.08.003> PMID: 19735947.
 77. Spano S, Galan JE. A Rab32-dependent pathway contributes to *Salmonella typhi* host restriction. *Science*. 2012; 338(6109):960–3. Epub 2012/11/20. <https://doi.org/10.1126/science.1229224> PMID: 23162001; PubMed Central PMCID: PMC3693731.
 78. Wu H, Jones RM, Neish AS. The *Salmonella* effector AvrA mediates bacterial intracellular survival during infection in vivo. *Cell Microbiol*. 2012; 14(1):28–39. Epub 2011/09/09. <https://doi.org/10.1111/j.1462-5822.2011.01694.x> PMID: 21899703; PubMed Central PMCID: PMC3240734.
 79. Du F, Galan JE. Selective inhibition of type III secretion activated signaling by the *Salmonella* effector AvrA. *PLoS Pathog*. 2009; 5(9):e1000595. Epub 2009/09/26. <https://doi.org/10.1371/journal.ppat.1000595> PMID: 19779561; PubMed Central PMCID: PMC2742890.
 80. Günster RA, Matthews SA, Holden DW, Thurston TLM. SseK1 and SseK3 Type III Secretion System Effectors Inhibit NF- κ B Signaling and Necroptotic Cell Death in *Salmonella*-Infected Macrophages. *Infection and Immunity*. 2017; 85(3).
 81. Rodriguez-Valera F, Martin-Cuadrado AB, Rodriguez-Brito B, Pasic L, Thingstad TF, Rohwer F, et al. Explaining microbial population genomics through phage predation. *Nat Rev Microbiol*. 2009; 7(11):828–36. Epub 2009/10/17. <https://doi.org/10.1038/nrmicro2235> PMID: 19834481.
 82. Mottawea W, Duceppe MO, Dupras AA, Usongo V, Jeukens J, Freschi L, et al. *Salmonella enterica* Prophage Sequence Profiles Reflect Genome Diversity and Can Be Used for High Discrimination Subtyping. *Front Microbiol*. 2018; 9:836. Epub 2018/05/22. <https://doi.org/10.3389/fmicb.2018.00836> PMID: 29780368; PubMed Central PMCID: PMC5945981.
 83. Figueroa-Bossi N, Uzzau S, Malorli D, Bossi L. Variable assortment of prophages provides a transferable repertoire of pathogenic determinants in *Salmonella*. *Mol Microbiol*. 2001; 39(2):260–72. <https://doi.org/10.1046/j.1365-2958.2001.02234.x> PMID: 11136448
 84. Bossi L, Fuentes JA, Mora G, Figueroa-Bossi N. Prophage contribution to bacterial population dynamics. *J Bacteriol*. 2003; 185(21):6467–71. Epub 2003/10/18. <https://doi.org/10.1128/jb.185.21.6467-6471.2003> PMID: 14563883; PubMed Central PMCID: PMC219396.
 85. Hawkey J, Edwards DJ, Dimovski K, Hiley L, Billman-Jacobe H, Hogg G, et al. Evidence of microevolution of *Salmonella* Typhimurium during a series of egg-associated outbreaks linked to a single chicken farm. *BMC Genomics*. 2013; 14:800. Epub 2013/11/20. <https://doi.org/10.1186/1471-2164-14-800> PMID: 24245509; PubMed Central PMCID: PMC3870983.
 86. Zeder MA. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proc Natl Acad Sci U S A*. 2008; 105(33):11597–604. Epub 2008/08/14. <https://doi.org/10.1073/pnas.0801317105> PMID: 18697943; PubMed Central PMCID: PMC2575338.
 87. Petrovska L, Mather AE, AbuOun M, Branchu P, Harris SR, Connor T, et al. Microevolution of Monophasic *Salmonella* Typhimurium during Epidemic, United Kingdom, 2005–2010. *Emerging Infectious Diseases*. 2016; 22(4):617–24. <https://doi.org/10.3201/eid2204.150531> PMC4806966. PMID: 26982594
 88. Kroger 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(20):E1277–86. Epub 2012/04/28. <https://doi.org/10.1073/pnas.1201061109> PMID: 22538806; PubMed Central PMCID: PMC3356629.
 89. Kingsley RA, Msefula CL, Thomson NR, Kariuki S, Holt KE, Gordon MA, et al. Epidemic multiple drug resistant *Salmonella* Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype. *Genome research*. 2009; 19(12):2279–87. Epub 2009/11/11. <https://doi.org/10.1101/gr.091017.109> PMID: 19901036; PubMed Central PMCID: PMC2792184.
 90. Kingsley RA, Kay S, Connor T, Barquist L, Sait L, Holt KE, et al. Genome and transcriptome adaptation accompanying emergence of the definitive type 2 host-restricted *Salmonella enterica* serovar Typhimurium pathovar. *MBio*. 2013; 4(5):e00565–13. Epub 2013/08/29. <https://doi.org/10.1128/mBio.00565-13> PMID: 23982073; PubMed Central PMCID: PMC3760250.
 91. Makendi C, Page AJ, Wren BW, Le Thi Phuong T, Clare S, Hale C, et al. A Phylogenetic and Phenotypic Analysis of *Salmonella enterica* Serovar Weltevreden, an Emerging Agent of Diarrheal Disease

- in Tropical Regions. *PLoS Negl Trop Dis*. 2016; 10(2):e0004446. Epub 2016/02/13. <https://doi.org/10.1371/journal.pntd.0004446> PMID: 26867150; PubMed Central PMCID: PMC4750946.
92. Zerbino D, Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*. 2008; 18(5):821–9. velvet-2008. <https://doi.org/10.1101/gr.074492.107> PMID: 18349386
 93. Zerbino DR. Using the Velvet de novo assembler for short-read sequencing technologies. *Current protocols in bioinformatics*. 2010;Chapter 11:Unit 11.5. Epub 2010/09/14. <https://doi.org/10.1002/0471250953.bi1105s31> PMID: 20836074; PubMed Central PMCID: PMC2952100.
 94. Page AJ, De Silva N, Hunt M, Quail MA, Parkhill J, Harris SR, et al. Robust high-throughput prokaryote de novo assembly and improvement pipeline for Illumina data. *Microb Genom*. 2016; 2(8):e000083. Epub 2017/03/30. <https://doi.org/10.1099/mgen.0.000083> PMID: 28348874; PubMed Central PMCID: PMC5320598.
 95. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics*. 2011; 27(4):578–9. Epub 2010/12/15. <https://doi.org/10.1093/bioinformatics/btq683> PMID: 21149342.
 96. Nadalin F, Vezzi F, Policriti A. GapFiller: a de novo assembly approach to fill the gap within paired reads. *BMC Bioinformatics*. 2012; 13(14):1–16. <https://doi.org/10.1186/1471-2105-13-s14-s8> PMID: 23095524
 97. Kolmogorov M, Raney B, Paten B, Pham S. Ragout—a reference-assisted assembly tool for bacterial genomes. *Bioinformatics*. 2014; 30(12):i302–9. Epub 2014/06/17. <https://doi.org/10.1093/bioinformatics/btu280> PMID: 24931998; PubMed Central PMCID: PMC4058940.
 98. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014; 30(14):2068–9. <https://doi.org/10.1093/bioinformatics/btu153> PMID: 24642063
 99. Li H. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics*. 2016; 32(14):2103–10. <https://doi.org/10.1093/bioinformatics/btw152> PMID: 27153593
 100. Sommer DD, Delcher AL, Salzberg SL, Pop M. Minimus: a fast, lightweight genome assembler. *BMC Bioinformatics*. 2007; 8(1):64. <https://doi.org/10.1186/1471-2105-8-64> PMID: 17324286
 101. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, et al. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb Genom*. 2016; 2(4):e000056. Epub 2017/03/30. <https://doi.org/10.1099/mgen.0.000056> PMID: 28348851; PubMed Central PMCID: PMC5320690.
 102. Stamatakis A. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 2006; 22(21):2688–90. raxml-2006. <https://doi.org/10.1093/bioinformatics/btl446> PMID: 16928733
 103. Almeida F, Seribelli AA, Medeiros MIC, Rodrigues DdP, MelloVarani Ad, Luo Y, et al. Phylogenetic and antimicrobial resistance gene analysis of *Salmonella* Typhimurium strains isolated in Brazil by whole genome sequencing. *PLOS ONE*. 2018; 13(8):e0201882. <https://doi.org/10.1371/journal.pone.0201882> PMID: 30102733
 104. Hayden HS, Matamouros S, Hager KR, Brittnacher MJ, Rohmer L, Radey MC, et al. Genomic Analysis of *Salmonella enterica* Serovar Typhimurium Characterizes Strain Diversity for Recent U.S. Salmonellosis Cases and Identifies Mutations Linked to Loss of Fitness under Nitrosative and Oxidative Stress. *mBio*. 2016; 7(2). <https://doi.org/10.1128/mBio.00154-16> PMID: 26956590
 105. Hunt M, Mather AE, Sánchez-Busó L, Page AJ, Parkhill J, Keane JA, et al. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. 2017. <https://doi.org/10.1101/118000>
 106. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother*. 2012; 67(11):2640–4. Epub 2012/07/12. <https://doi.org/10.1093/jac/dks261> PMID: 22782487; PubMed Central PMCID: PMC3468078.
 107. Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. *Nucleic Acids Res*. 2016; 44(D1):D694–7. Epub 2015/11/19. <https://doi.org/10.1093/nar/gkv1239> PMID: 26578559; PubMed Central PMCID: PMC4702877.
 108. Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O, Villa L, et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother*. 2014; 58(7):3895–903. Epub 2014/04/30. <https://doi.org/10.1128/AAC.02412-14> PMID: 24777092; PubMed Central PMCID: PMC4068535.
 109. Inouye M, Dashnow H, Raven LA, Schultz MB, Pope BJ, Tomita T, et al. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med*. 2014; 6(11):90. Epub 2014/11/26. <https://doi.org/10.1186/s13073-014-0090-6> PMID: 25422674; PubMed Central PMCID: PMC4237778.

110. Otto TD, Dillon GP, Degraeve WS, Berriman M. RATT: Rapid Annotation Transfer Tool. *Nucleic Acids Res.* 2011; 39(9):e57. Epub 2011/02/11. <https://doi.org/10.1093/nar/gkq1268> PMID: 21306991; PubMed Central PMCID: PMC3089447.
111. Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J. ACT: the Artemis Comparison Tool. *Bioinformatics.* 2005; 21(16):3422–3. <https://doi.org/10.1093/bioinformatics/bti553> PMID: 15976072.
112. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. *Nucleic Acids Research.* 2018:gky995–gky. <https://doi.org/10.1093/nar/gky995> PMID: 30357350
113. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 2011; 39(Web Server issue):W29–37. Epub 2011/05/20. <https://doi.org/10.1093/nar/gkr367> PMID: 21593126; PubMed Central PMCID: PMC3125773.
114. Arndt D, Marcu A, Liang Y, Wishart DS. PFAST, PFASTER and PFASTEST: Tools for finding prophage in bacterial genomes. *Brief Bioinform.* 2017. Epub 2017/10/14. <https://doi.org/10.1093/bib/bbx121> PMID: 29028989.
115. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Research.* 2015; 43(3):e15–e. <https://doi.org/10.1093/nar/gku1196> PMID: 25414349
116. Hadfield J, Croucher NJ, Goater RJ, Abudahab K, Aanensen DM, Harris SR. Phandango: an interactive viewer for bacterial population genomics. *Bioinformatics.* 2017. Epub 2017/10/14. <https://doi.org/10.1093/bioinformatics/btx610> PMID: 29028899; PubMed Central PMCID: PMC5860215.
117. Didelot X, Wilson DJ. ClonalFrameML: Efficient Inference of Recombination in Whole Bacterial Genomes. *PLOS Computational Biology.* 2015; 11(2):e1004041. <https://doi.org/10.1371/journal.pcbi.1004041> PMID: 25675341
118. 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(22):3691–3. Epub 2015/07/23. <https://doi.org/10.1093/bioinformatics/btv421> PMID: 26198102; PubMed Central PMCID: PMC4817141.
119. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. *Genome Biol.* 2004; 5(2):R12. <https://doi.org/10.1186/gb-2004-5-2-r12> PMID: 14759262.
120. Adriaenssens E, Brister JR. How to Name and Classify Your Phage: An Informal Guide. *Viruses.* 2017; 9(4). Epub 2017/04/04. <https://doi.org/10.3390/v9040070> PMID: 28368359; PubMed Central PMCID: PMC5408676.
121. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics.* 2012; 28(23):3150–2. <https://doi.org/10.1093/bioinformatics/bts565> PMID: 23060610; PubMed Central PMCID: PMC3516142.
122. Agren J, Sundstrom A, Hafstrom T, Segerman B. Gegenees: fragmented alignment of multiple genomes for determining phylogenomic distances and genetic signatures unique for specified target groups. *PLoS One.* 2012; 7(6):e39107. Epub 2012/06/23. <https://doi.org/10.1371/journal.pone.0039107> PMID: 22723939; PubMed Central PMCID: PMC3377601.
123. Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat Biotechnol.* 2019; 37(6):632–9. Epub 2019/05/08. <https://doi.org/10.1038/s41587-019-0100-8> PMID: 31061483.
124. O’Leary NA, Wright MW, Brister JR, Ciuffo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 2016; 44(D1):D733–45. Epub 2015/11/11. <https://doi.org/10.1093/nar/gkv1189> PMID: 26553804; PubMed Central PMCID: PMC4702849.
125. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nature methods.* 2015; 12(1):59–60. Epub 2014/11/18. <https://doi.org/10.1038/nmeth.3176> PMID: 25402007.
126. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* 2002; 30(7):1575–84. <https://doi.org/10.1093/nar/30.7.1575> PMID: 11917018; PubMed Central PMCID: PMC101833.
127. Nepusz T, Yu H, Paccanaro A. Detecting overlapping protein complexes in protein-protein interaction networks. *Nat Methods.* 2012; 9(5):471–2. Epub 2012/03/20. <https://doi.org/10.1038/nmeth.1938> PMID: 22426491; PubMed Central PMCID: PMC3543700.
128. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003; 13(11):2498–504. Epub 2003/11/05. <https://doi.org/10.1101/gr.1239303> PMID: 14597658; PubMed Central PMCID: PMC403769.

129. Antipov D, Hartwick N, Shen M, Raiko M, Lapidus A, Pevzner PA. plasmidSPAdes: assembling plasmids from whole genome sequencing data. *Bioinformatics*. 2016; 32(22):3380–7. Epub 2016/07/29. <https://doi.org/10.1093/bioinformatics/btw493> PMID: 27466620.
130. Yoshida CE, Kruczkiewicz P, Laing CR, Lingohr EJ, Gannon VP, Nash JH, et al. The Salmonella In Silico Typing Resource (SISTR): An Open Web-Accessible Tool for Rapidly Typing and Subtyping Draft Salmonella Genome Assemblies. *PLoS One*. 2016; 11(1):e0147101. Epub 2016/01/23. <https://doi.org/10.1371/journal.pone.0147101> PMID: 26800248; PubMed Central PMCID: PMC4723315.