Presence of phage-plasmids in multiple serovars of *Salmonella enterica*

Satheesh Nair[1], Clare R Barker[1,2], Matthew Bird[1,3], David R Greig[1,4,5], Caitlin Collins[1,2,6], Anaïs Painset[1], Marie Chattaway[1,2], Derek Pickard[7], Lesley Larkin[1], Saheer Gharbia[1,2,3], Xavier Didelot[2,4,8], Paolo Ribeca[1,2,4]

1) UK Health Security Agency, London, UK

2) NIHR Health Protection Research Unit in Genomics and Enabling Data, University of Warwick, UK

3) NIHR Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance, University of Oxford, UK

4) NIHR Health Protection Research Unit in Gastrointestinal Infections, University of Liverpool, UK

5) Division of Infection and Immunity, The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, UK

6) Department of Veterinary Medicine, University of Cambridge, Cambridge, UK

7) The Cambridge Institute for Therapeutic Immunology and Infectious Disease (CITIID), University of Cambridge, UK

8) School of Public Health and Department of Statistics, University of Warwick, UK

Corresponding authors: paolo.ribeca@ukhsa.gov.uk & satheesh.nair@ukhsa.gov.uk

Keywords: *Salmonella*, phage-plasmid, antimicrobial resistance, genomic surveillance

Abstract

Evidence is accumulating in the literature that the horizontal spread of antimicrobial resistance (AMR) genes mediated by bacteriophages and bacteriophage-like plasmid (phage-plasmid) elements is much more common than previously envisioned. For instance, we recently identified and characterised a circular P1-like phage-plasmid harbouring a $bla_{CTX\text{-}M\text{-}15}$ gene conferring extended-spectrum beta-lactamase (ESBL) resistance in *S.* Typhi. As the prevalence and epidemiological relevance of such mechanisms has never been systematically assessed in Enterobacterales, in this study we carried out a follow-up retrospective analysis of UK *Salmonella* isolates previously sequenced as part of routine surveillance protocols between 2016 and 2021. Using a high-throughput bioinformatics pipeline we screened 47,784 isolates for the presence of the P1 lytic replication gene *repL*, identifying 226 positive isolates from 25 serovars and demonstrating that phage-plasmid elements are more frequent than previously thought. The affinity for phage-plasmids appears highly serovar-dependent, with several serovars being more likely hosts than others; most of the positive isolates (170/226) belonged to *S.* Typhimurium ST34 and ST19. The phage-plasmids ranged between 85.8–99.5kb in size, with an average length of 92.6kb; detailed analysis indicated a high amount of diversity in gene content and genomic architecture. 132 phage-plasmids had the p0111 plasmid replication type, and 94 the IncY type; phylogenetic analysis indicated that both horizontal and vertical gene transmission mechanisms are likely to be involved in phage-plasmid propagation. Finally, phage-plasmids were present in isolates that were resistant and non-resistant to antimicrobials. In addition to providing a first comprehensive view of the presence of phage-plasmids in *Salmonella*, our work highlights the need for a better surveillance and understanding of phage-plasmids as AMR carriers, especially through their characterisation with long-read sequencing.

Data Summary

All of the FASTQ files examined in this study have been uploaded to the Sequence Read Archive under BioProject PRJNA248792. Accessions of individual isolates which were found to contain phage plasmids are listed in Supplementary Table S1.

Impact Statement

Bacteriophage-like plasmids are increasingly being recognised as important mobile elements in many species of bacteria, particularly due to their involvement in the transmission of antimicrobial resistance (AMR); however, few studies of their overall prevalence in clinical datasets have been undertaken to date. In this study we have performed the first large-scale surveillance of human disease-associated *Salmonella* genomes for the presence of P1-like phage-plasmids, showing that they are more common than previously thought. Furthermore, we highlight how valuable information about the evolution and transmission of phage-plasmids in *Salmonella* and other Enterobacterales can be revealed by linking phage-plasmid prevalence and genetic diversity to epidemiologically relevant metadata such as *S. enterica* serovar, outbreak clusters, time, and geography. Our work shows the ability to

use sequencing data and scalable bioinformatics workflows for the detection and characterisation of these extrachromosomal elements, highlights the importance of screening for novel mechanisms of AMR transmission, and provides a foundation for further surveillance studies of phage-plasmid prevalence.

Introduction

The global threat of rising antimicrobial resistance (AMR) incidence is well known, particularly among genera of Enterobacterales such as *Escherichia* and *Salmonella* (1,2). AMR genes can be transferred between bacteria via horizontal gene transfer (HGT), which involves many diverse molecular mechanisms mediated by mobile and integrative elements including transposons and integrons, plasmids, bacteriophages and genomic islands, all of which are well known for their role in the transfer of AMR genes (3) as well as other genes such as virulence and fitness factors. Temperate bacteriophages, which typically integrate as prophages into the host chromosome, are another important mechanism of HGT and further contribute to the spread of AMR genes, toxins and virulence factors (4,5). Moreover, certain temperate bacteriophages have been found to exist in their prophage forms as low-copy number, extrachromosomal-plasmids and thus can replicate autonomously within bacterial cells while remaining latent (6,7).

Large-scale computational analysis of such bacteriophage-like plasmids, or "phage-plasmids", has shown that they are associated with multiple specific types of bacteriophage including P1, P7 and SSU5 (8). The plasmid replication types linked to the P1-like phage-plasmid group are consistently either IncY or p0111 (9), suggesting both types can theoretically be compatible and maintained in a single host.

There is increasing evidence from the literature that the horizontal spread of AMR genes by these phage-plasmids is more common than previously envisioned, especially among the P1, SSU5 and AB types (10). It is speculated that bacteriophages carrying AMR genes would confer an evolutionary advantage to their hosts, especially during latent infection as phage-plasmids. This mechanism is underpinned by several complex molecular pathways currently being elucidated (11); it would eventually result in phage-plasmids with AMR genes being positively selected. For instance, phage-plasmids carrying extended-spectrum beta-lactamase (ESBL) and colistin resistance genes have been identified in *Escherichia coli* (9,12–19), *Klebsiella pneumoniae* (20) and *Salmonella enterica* (10,21–23) to date. Among *S. enterica*, reports suggest that phage-plasmids are associated more specifically with the AMR genes $bla_{\text{CTX-M-15}}$ (22), $bla_{\text{CTX-M-27}}$ (21,23) and *mcr-1* (23).

We recently identified, characterised and described for the first time a circular P1-like phage-plasmid harbouring a $bla_{\text{CTX-M-15}}$ gene conferring high-level ESBL resistance within *S. enterica* serovar Typhi (22). Following this discovery, we wanted to determine the extent to which such elements may be distributed among all the subspecies and serovars of *Salmonella* in our collection, and whether these phage-plasmids play a significant role in the increasing spread of antibiotic resistance.

The Gastrointestinal Bacteria Reference Unit (GBRU) at the UK Health Security Agency (UKHSA) is the UK national reference laboratory for enteric bacterial pathogens. Routine whole genome sequencing (WGS) for all *Salmonella* spp. was implemented in 2014, resulting in a collection of roughly 9000-10000 genomes per year from hundreds of different

108 serovars. In this study, we offer a retrospective examination of the genomic data associated
109 with this national surveillance. We developed a high-throughput process for detecting the
110 occurrence of P1-like phage-plasmids and applied it to examine the genomes of *Salmonella*
111 spp. isolates received by the reference laboratory. Our goal was to gain insight into the
112 distribution, prevalence, and connection of these phage-plasmids with antimicrobial
113 resistance genes.

Materials and Methods

Data collection & bacterial strains

A total of 47,784 *Salmonella* spp. isolates were sequenced by GBRU between January 2016 and December 2021. All of the FASTQ files have been uploaded to the Sequence Read Archive under BioProject PRJNA248792.

Whole genome sequencing and bioinformatics

Genomic DNA was extracted using the Qiasymphony system (Qiagen, UK). The Nextera XP kit (Illumina) was used to prepare the sequencing library for sequencing on the HiSeq 2500 and NextSeq 1000 instruments (Illumina), run with the fast protocol. Trimmomatic v0.27 (24) was utilised to remove bases with a PHRED score of <30 from the leading and trailing ends on the FASTQ reads, with reads <50bp after trimming discarded. The eBurst Group (eBG) was determined as previously described (25). Multilocus sequence type (MLST) assignment was performed using MOST v2.18 which is available at (https://github.com/phe-bioinformatics/MOST) (26), and the reads (FASTQ) underwent SNP typing and clustering using SnapperDB v0.2.9 (27). Routine detection of AMR genes and plasmid replication gene types was conducted using in-house software.

Detection of *repL* gene

The *repL* gene forms part of the L-replicon of the P1 family of temperate bacteriophages, which is the active replicon for DNA replication during the lytic cycle (6); therefore, we used the presence of *repL* as an indicator for carriage of a P1-like phage plasmid. For each of the 47,784 samples, reads were mapped to the *repL* gene sequence with the GEM mapper v3.6.0 (28) in default mode. Alignments were filtered to keep those of sufficient quality (having no more than 2 indels or clips; such that sum of the number of substitutions, indels and clips would not exceed 5% of the read length; and such that the sum of the number of mismatches and the lengths of indels and clips would not exceed 10% of the read length); all multimaps were kept. Pileups were generated with Samtools pileup v1.15.1 (using htslib v1.16) and parsed to determine coverage. We considered a sample to be a potential positive for the detection of *repL* whenever the number of bases having non-zero coverage was >60% of the total, and the average coverage of bases having non-zero coverage was >1. Further manual inspection of coverage plotted along *repL* and exclusion of a few genomes with only some gene sections covered and/or very low average coverage yielded 226 candidate samples (Table S1).

Genome assembly and phylogenies

The 226 genomes in which the *repL* gene was present with high confidence according to the alignment pipeline were assembled *de novo* from the paired-end reads using SPAdes v3.8.0 (29). The assembled genomes were visualised using Bandage v0.8.1 (30), with the built-in BLAST (31) tool used to identify the phage-plasmid contig(s) associated with the *repL* gene (NC_005856.1) and the IncY (K02380.1) and p0111 (AP010962.1) plasmid replication genes, as well as to determine the location of any AMR genes. The putative phage-plasmid contigs

153 were manually curated, extracted and used to perform all downstream comparative
154 genomic and phylogenetic analyses. A total of 33/226 genome assemblies were of too poor
155 quality to extract the phage-plasmid contigs for further analysis. Panaroo v1.3.4 (32) was
156 used to produce core (present in >98% phage-plasmid contigs) gene alignments, which were
157 passed to FastTree v2.1.11 (33) to generate maximum-likelihood trees. The individual
158 phage-plasmid contigs were annotated using Bakta v1.8.2 (34), and visualised and compared
159 using Easyfig (35) and Clinker (36).

160 Phylogenetic analysis of plasmid replication type

161 To examine how the plasmid replication type varied across the *repL*-positive phage-plasmid
162 tree, we performed ancestral state reconstruction (37) and a phylogenetic randomisation
163 test. First, to determine if the two replication types were distributed randomly across the
164 tips of the tree without regard to tree structure, we performed a simple label-switching test.
165 The observed states, IncY (n=79/193) and p0111 (n=114/193), were randomly reassigned to
166 new terminal nodes. This process was repeated 100 times. In each case, a Welch t-test was
167 performed using the R package *stats* to assess whether the distribution of Euclidean genetic
168 distances between IncY and p0111 individuals differed between the observed and the tip-
169 randomised dataset. The mean Bonferroni-corrected p-value is 1.07e-15, indicating that
170 plasmid replication type is not distributed randomly across the tips of the tree if one
171 disregards tree structure.

172 However, as the presence of large clades with near-zero branch lengths might inflate the
173 appearance of clustering and vertical inheritance in the replication type, we needed to
174 account for phylogenetic structure. We performed ancestral state reconstruction using the
175 discrete maximum-likelihood method in R package *phangorn* (38) to infer the likely
176 replication type state on the internal branches of the phage-plasmid tree, identifying 17
177 changes in replication type along the tree (Fig. S1). To determine whether these 17 changes
178 occurred randomly across the phylogenetic history of the sample, we performed a
179 phylogenetically-informed randomisation test. The 17 changes were randomly reassigned to
180 new branches, with probability of reassignment proportional to branch length, and
181 IncY/p0111 states were simulated from root to tips according to these change-points with R
182 package *treeWAS* v1.1 (39). We performed 100 simulations and repeated the Welch t-test as
183 described above to determine whether the observed replication type states showed any
184 greater propensity for vertical inheritance than one might expect despite the number of
185 changes inferred.

186 # Results

187 ## Phage-plasmid prevalence

188 Using our mapping pipeline, the *repL* gene was detected in the reads of 226 isolates (0.47%),
189 comprising 0.2% to 1.1% of the dataset depending on the year (Table 1, Table S1). The *repL*
190 gene was present in 25 different *Salmonella* serovars, and was most prevalent among
191 isolates of *S.* Typhimurium (n=170), followed by *S.* Oslo (n=9), *S.* Livingstone (n=6), and *S.*
192 Indiana (n=5). Table 1 also shows serovar prevalence for the whole set of 47,784 isolates.
193 Comparison for serovars for which phage-plasmids were detected suggests that they tend to
194 occur much less than expected in *S.* Enteritidis (0.44% with phage-plasmids vs. 27%
195 prevalence of Enteritidis within our whole *Salmonella* collection); less than expected in *S.*
196 Agona (0.9% vs. 2.3%), *S.* Infantis (1.3% vs. 4.3%), *S.* Typhi (0.9% vs. 3.2%); much more than
197 expected (84% vs. 19%) in *S.* Typhimurium, *S.* Oslo (4.0% vs. 0.31%), and *S.* Livingstone (2.7%
198 vs. 0.21%); and more than expected in *S.* Indiana (2.2% vs 0.9%), *S.* Mbandaka (1.8% vs.
199 0.81%), and *S.* Mikawasima (1.8% vs. 0.76%). Phage-plasmids were also not detected in a
200 large number of serovars often occurring in the full surveillance datasets (the most frequent
201 being *S.* Newport, corresponding to 3.3% of the 47,784 isolates; *S.* Paratyphi A, 1.6%; *S.*
202 Kentucky, 1.5%; and *S.* Braenderup, 1.5%).

203 ## Phage-plasmid features

204 To explore the structure and content of the putative phage-plasmids, we performed *de novo*
205 assembly of the 226 *repL*-positive *Salmonella* genomes and extracted the contigs upon
206 which the *repL* gene was located, where possible. Analysis of these contigs (n=193) showed
207 that the phage-plasmids ranged between 85.8–99.5kb in size with an average length of
208 92.6kb, except for four isolates of *S.* Typhimurium which had notably shorter or longer
209 phage-plasmid sequences of 76-79kb and 104-109kb (Table 1, Table S1). Additionally, two
210 other *S*. Typhimurium genomes had phage-plasmids located on single contigs >115kb in
211 length; these were putatively determined to instead be integrated into chromosomal
212 sequences of these isolates (Fig. S2).

213 A search against a database of plasmid replication genes revealed that 94 of the phage-
214 plasmids had the IncY plasmid replication type and 132 had the p0111 replication type. To
215 assess the level of vertical or horizontal transmission from the correlation between phage-
216 plasmid phylogeny and plasmid replication type, a maximum-likelihood tree based on a
217 51,260bp alignment of the 54 core genes from the 193 phage-plasmid contigs was
218 constructed (Fig. 1). Statistical analysis showed that the two types are not distributed
219 completely randomly across the tips of the tree, if one disregards the tree structure (p-value
220 = 1.07e-15). That excludes a scenario where transmission occurs in a fully horizontal
221 manner. We then conducted a different analysis that considers the structure of the
222 phylogenetic tree. Ancestral state reconstruction (37) revealed that 17 changes in
223 replication type had occurred along the evolutionary history of the phage-plasmid sample
224 (Fig. S1). Comparison to simulated data suggested that these changes occur randomly,
225 falling with greater probability on longer branches, and result in a distribution of replication

types that does not differ statistically from a random distribution once branch lengths are considered (p-value = 0.15). Vertical inheritance of replication type thus occurs among clones and some close relatives, but it is not maintained over greater genetic distances. This supports a mixed model of transmission, whereby, as expected, both vertical and horizontal gene transfer mechanisms play a role.

231

Serovar	No.	Prev. phm.	Prev. surv.	ST(s)	Years	Rep type(s)	Av. Size (kb)	AMR
Typhimurium	170	84%	19%	34, 19	2015-2021	IncY (63) p0111 (107)	92.5	None
Oslo	9	4.0%	0.31%	1370	2016	IncY (5) p0111 (4)	92.6	None
Livingstone	6	2.7%	0.21%	457, 638	2018-2019, 2021	IncY (4) p0111 (2)	95.3	None
Indiana	5	2.2%	0.90%	17	2018	p0111	96.1	None
Mbandaka	4	1.8%	0.81%	413, 1602	2016, 2018, 2020	IncY (3) p0111 (1)	92.4	None
Mikawasima	4	1.8%	0.76%	1815	2016-2018	IncY	93.7	None
Virchow	4	1.8%	1.4%	16	2016-2018, 2020	IncY (2) p0111 (2)	96.2	None
Infantis	3	1.3%	4.3%	32	2018	p0111	98.4	None
Stanley	3	1.3%	1.6%	29	2018-2019	IncY	94.6	None
Agona	2	0.90%	2.3%	13	2019	p0111	96.6	None
Typhi	2	0.90%	3.2%	1	2019	IncY	96.2	$bla_{CTX-M-15}$
Adjame	1	0.44%	0.11%	3929	2021	IncY	97.9	None
Bareilley	1	0.44%	1.2%	203	2018	p0111	92.8	None
Corvallis	1	0.44%	0.42%	1541	2016	p0111	94.0	None
S. enterica 45:b:-	1	0.44%	0.66%	2524	2016	IncY	94.7	None
Enteritidis	1	0.44%	27%	11	2016	IncY	-	None
Florian	1	0.44%	0.002%	7166	2020	p0111	-	None
Florida	1	0.44%	0.021%	351	2017	IncY	93.6	None
Minnesota	1	0.44%	0.31%	548	2019	IncY	93.7	None
Mississippi	1	0.44%	0.25%	4711	2020	IncY	-	None
Monschaui	1	0.44%	0.08%	443	2017	IncY	97.0	None
Ohio	1	0.44%	0.12%	329	2017	IncY	-	None
Pomona	1	0.44%	0.084%	451	2016	p0111	90.1	None
Rissen	1	0.44%	0.19%	469	2017	IncY	99.5	None
Telaviv	1	0.44%	0.0063%	1068	2016	p0111	89.4	None

232 **Table 1** – Summary of metadata for the 226 *Salmonella* isolates possessing the *repL* gene;
233 see Table S1 for individual isolate information. In the column "Prev. surv." ("prevalence in
234 surveillance") we report the percentage of occurrences in the 47,784 *Salmonella* spp.
235 isolates that were sequenced by GBRU between January 2016 and December 2021; in the
236 column "Prev. phm." (short for "prevalence in phage-plasmids") we report the percentage
237 of occurrences in the set of sequences that were found to contain phage-plasmids as a
238 result of the current study.

239

240 ## Genetic diversity

241 In almost all cases, phage-plasmid sequences are grouped according to the host isolate SNP
242 clustering but are not necessarily closely related to others from the same serovar (Fig. 1, Fig.
243 2). However, there are examples of cross-serovar similarity, with a phage-plasmid from an *S.*
244 Stanley that is highly similar to several from *S.* Typhimurium (Fig. S3). Additionally, there are
245 also phage-plasmids from closely related, epidemiologically-linked *S.* Typhimurium strains
246 (forming a 10-SNP cluster) that display considerable genetic diversity and even possess
247 different plasmid replicon types (Fig. 3). In relation to the P1 phage-plasmid itself, several
248 regions are variable (Fig.5, Fig. S4) including: 3.4kbp – 9.6kbp (encoding a type 3 restriction-
249 modification enzyme); 22.8kbp - 24.1kbp (encoding phage protein) and 33.6kbp - 35.8kbp
250 (encoding phage tail fibre proteins). However, there are large sections of conserved
251 sequence surrounding the previously mentioned variable regions which contain a variety of
252 phage and plasmid associated genes.

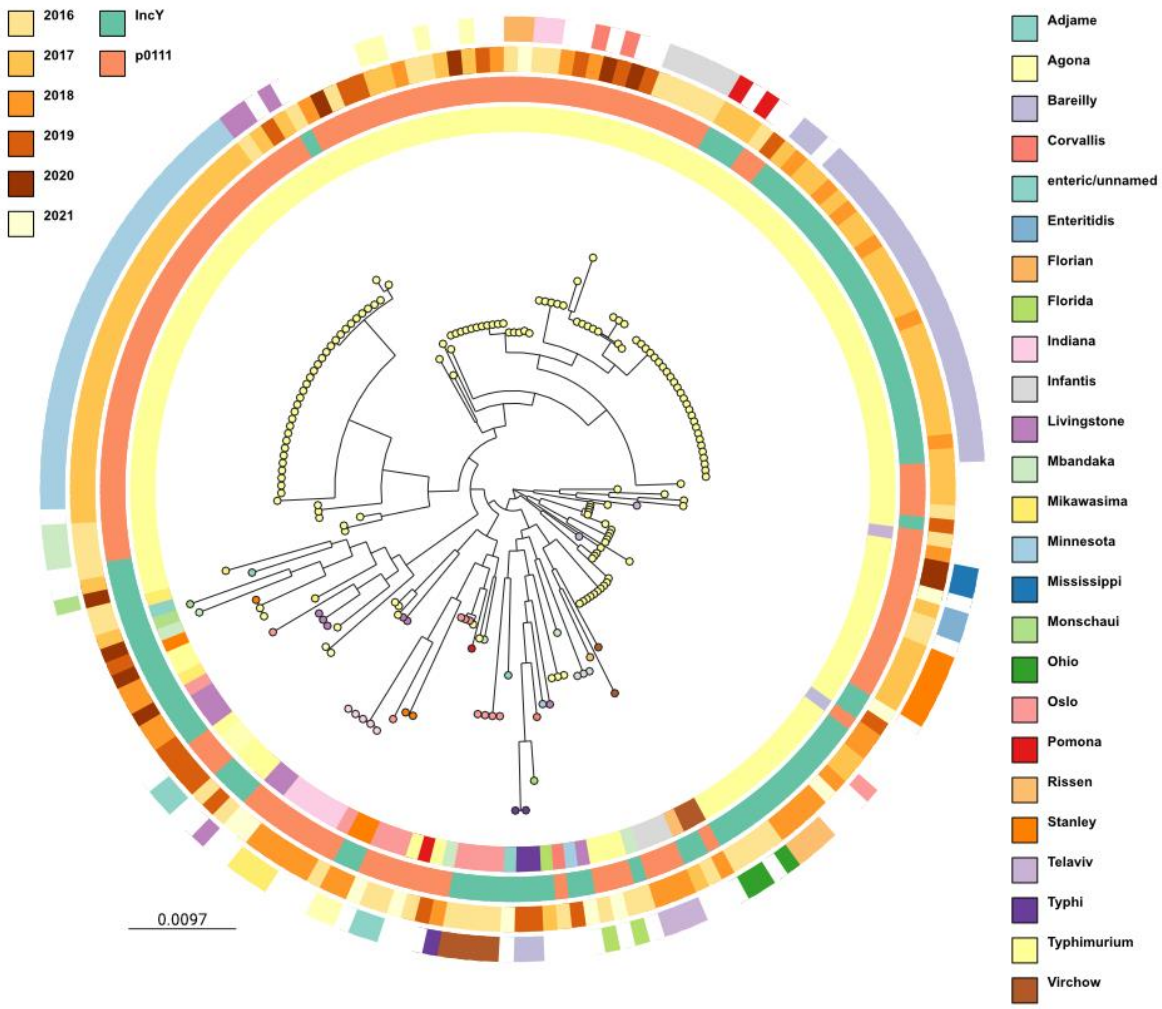

Fig. 1 – Midpoint-rooted maximum-likelihood tree of 51,260bp alignment of 54 core genes from the 193 *repL*-positive phage-plasmid contigs with tips coloured by serovar. Rings from inner to outer: serovar; replication gene type; year; host isolate SNP cluster.

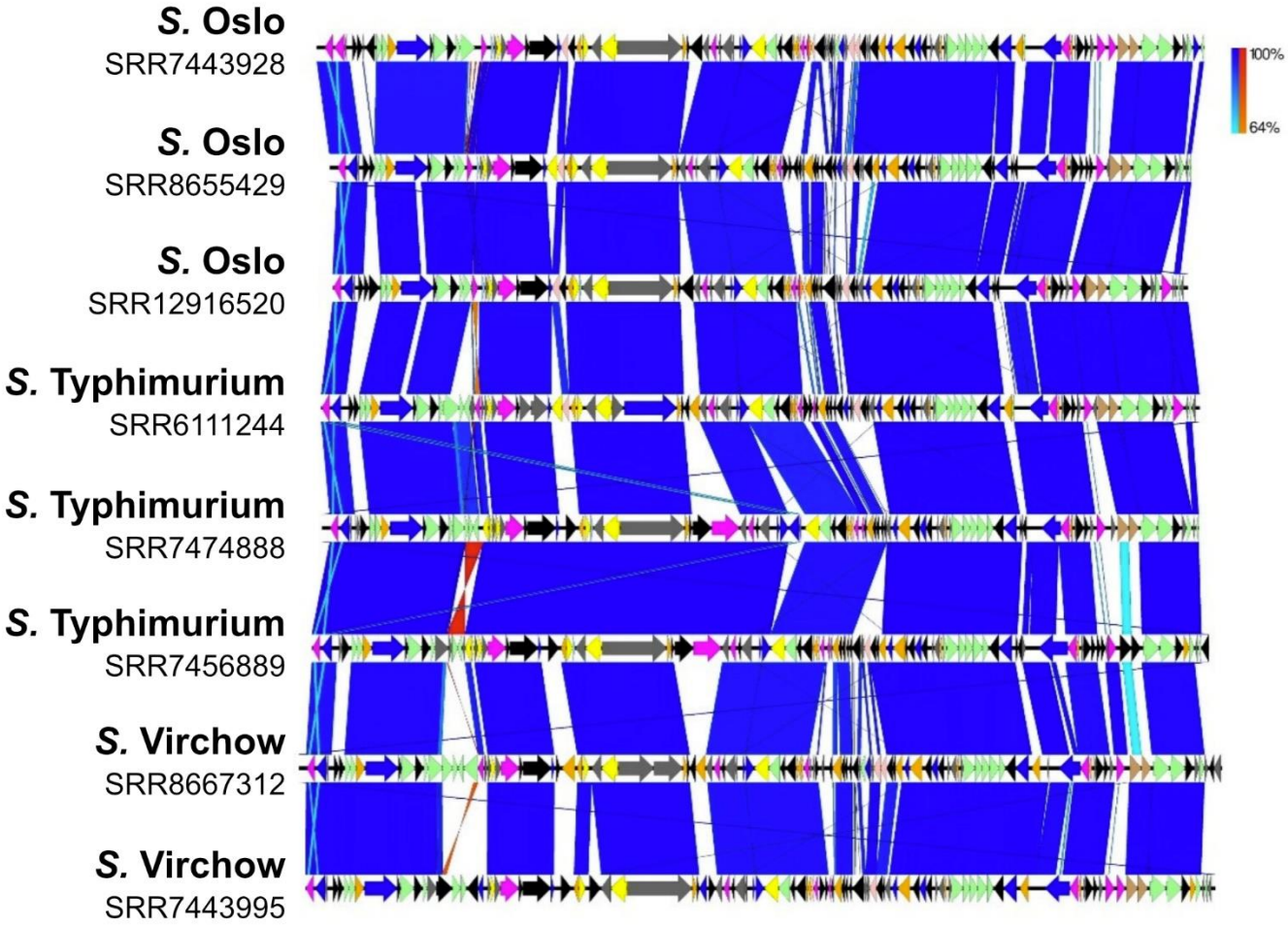

Fig. 2 – Plot showing how diversity occurs in sections in-between highly conserved regions, even among isolates belonging to the same serovar. Arrows indicate gene direction while colours indicate gene function. Hypothetical proteins are shown in grey; regulator genes are shown in blue; effector virulence genes are shown in pink; Phage-related genes are shown in light sky blue; Plasmid-related genes are shown in brown; Recombinant genes are shown in magenta; Metabolism/Signalling genes are shown in orange; Structure genes are shown in pale green; lysis genes are shown in yellow; uncharacterised or not well-established genes are shown in black. Scale bars indicate level of similarity for forward (blue) and reverse (red) sequences.

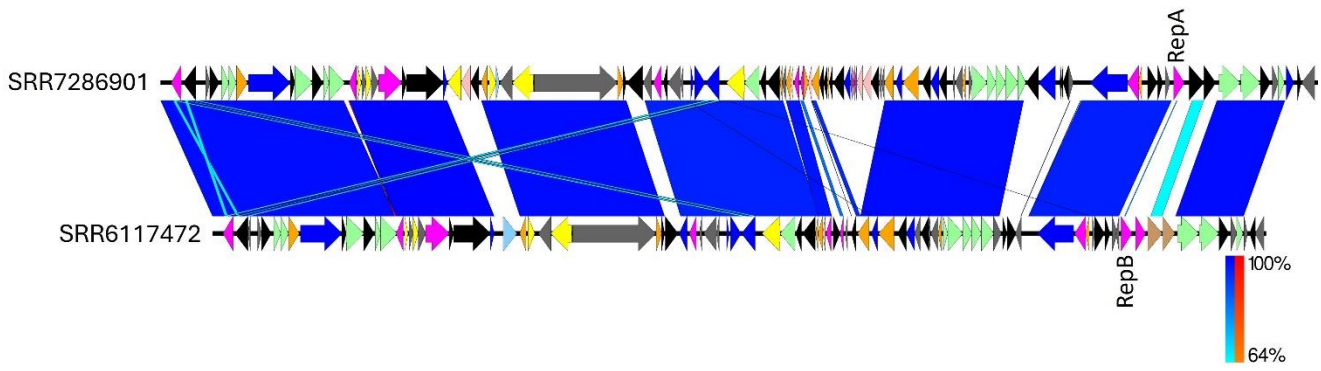

Fig. 3 – Plot showing variability in the sequences of two phage-plasmids from *S*. Typhimurium isolates that belong to the same 10-SNP cluster.

We next performed a nucleotide BLASTN search of the GenBank database to identify other phage-plasmids that were related to those in our collection, including from other species of Enterobacterales (Table S2). We produced a tree containing selected representative phage-plasmid sequences from our collection, alongside 31 publicly available sequences labelled as either phage-plasmids, plasmids or bacteriophages. This confirmed the high amount of genetic diversity and additionally showed that there is no apparent correlation between phage-plasmid sequence and species (Fig. 4). We did however identify several annotated plasmid or bacteriophage sequences that are highly similar to phage-plasmids from our collection (Fig. S5).

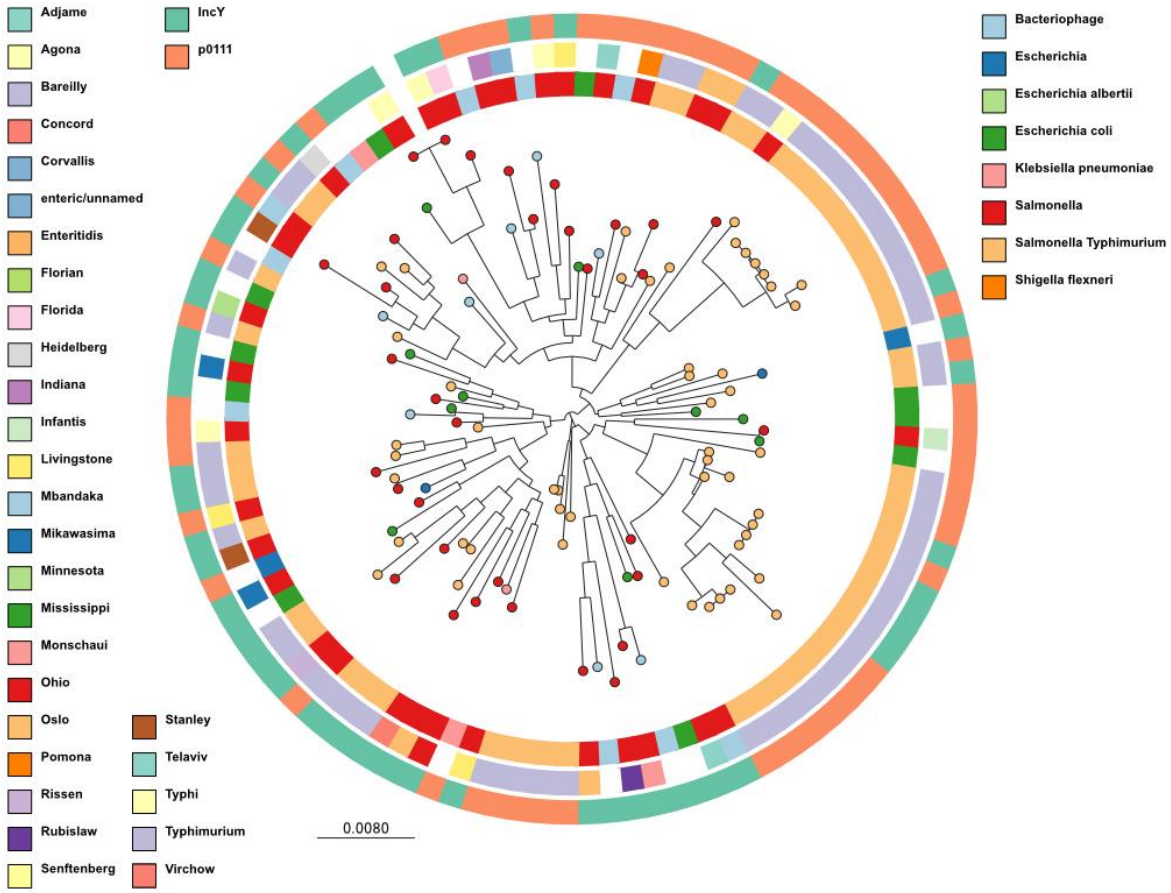

Fig. 4 – Midpoint-rooted maximum-likelihood tree of 42,606bp alignment of 46 core genes from 107 phage-plasmids, including 31 from public databases, with tips coloured by species. Rings from inner to outer: species; *S. enterica* serovar; replication gene type.

Antimicrobial resistance

AMR genes were detected in all 170 of the *S.* Typhimurium isolates and in 25/57 of the isolates belonging to other serovars (Table S1). Preliminary visual assessment of the whole genome sequences indicated that the majority of AMR genes were present on the chromosomes of these isolates rather than being associated with the phage-plasmid contigs. However, we did confirm the presence of $bla_{\text{CTX-M-15}}$ genes encoding ESBL resistance on phage-plasmids from two *S.* Typhi isolates (Fig. 5), both of which were associated with travel to Iraq. We have previously reported and described the phage-plasmid sequence in detail for one of these isolates (22). Further comparison with publicly-available plasmid and phage-plasmid sequences showed that the flanking region of $bla_{\text{CTX-M-15}}$ is variable, and that the gene is located next to an ISEc9/ISEcp1 element in the *S.* Typhi isolates that is similar to that alongside the $bla_{\text{CTX-M-55}}$ gene of *E. coli* phage-plasmid JL22 (ON018986), but different to that next to the $bla_{\text{CTX-M-27}}$ in *S.* Indiana phage SJ46 (KU760857).

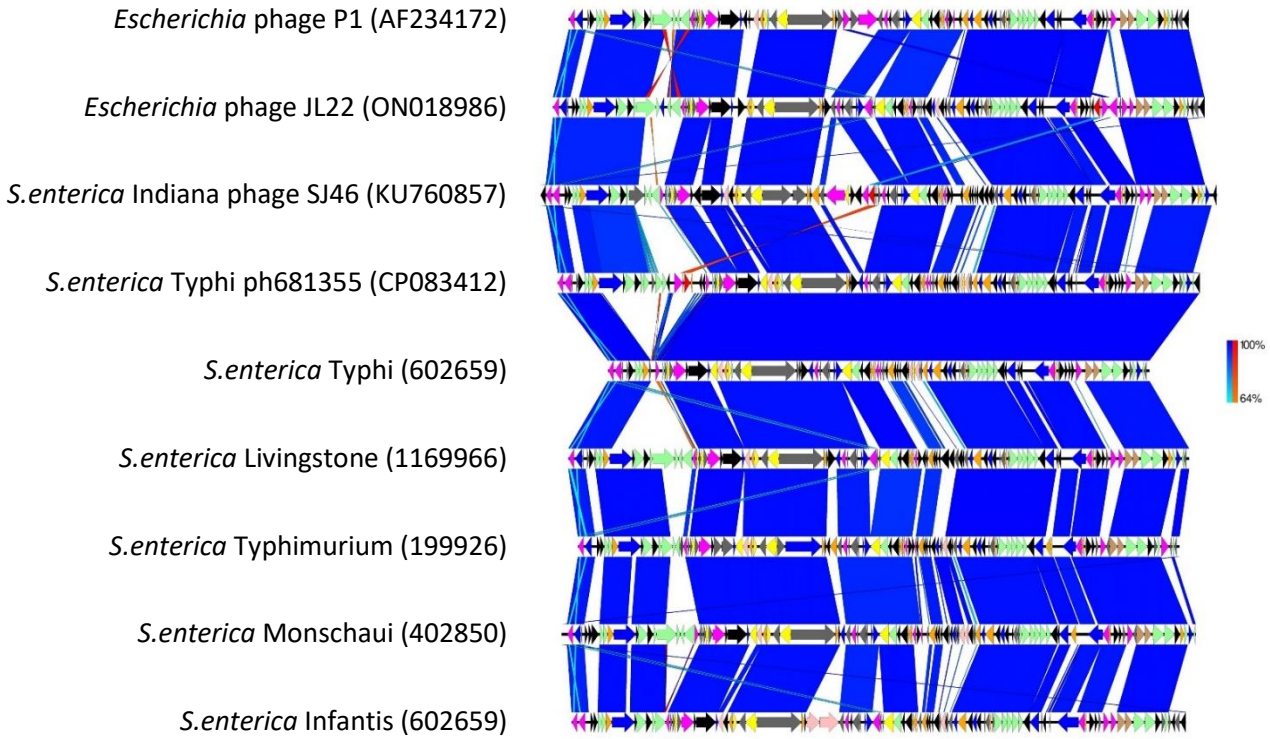

Fig. 5 – Comparison of *S*. Typhi ESBL-carrying phage-plasmid (CP083412: $bla_{\text{CTX-M-15}}$) to other ESBL-carrying phage-plasmids (ON018986: $bla_{\text{CTX-M-55}}$ & KU760857: $bla_{\text{CTX-M-27}}$). Arrows indicate gene direction while colours indicate gene function. Hypothetical proteins are shown in grey; regulator genes are shown in blue; effector virulence genes are shown in pink; Phage-related genes are shown in light sky blue; Plasmid-related genes are shown in brown; Recombinant genes are shown in magenta; Metabolism/Signalling genes are shown in orange; Structure genes are shown in pale green; lysis genes are shown in yellow; uncharacterised or not well-established genes are shown in black. Scale bars indicate level of similarity for forward (blue) and reverse (red) sequences. In the figure we can see that those encoding ESBL genes (CP083412, ON018986 & KU760857) shown by the red arrows, are encoded within variable regions but not within the same region.

317 ## Discussion

318 New and emerging research based on WGS highlights the prevalence of many different
319 phage-plasmids in the bacterial world. We performed retrospective genomic surveillance for
320 the presence of P1-like phage-plasmids within a collection of almost 48,000 *Salmonella*
321 isolates received over a five year period. Our work shows that phage-plasmids are
322 distributed among multiple serovars of *Salmonella* and have been present within the UK
323 population since at least 2016. While there is growing evidence of a significant phage-
324 plasmid presence among genera of Enterobacterales (8,10,16), targeted large-scale genomic
325 surveillance of *Salmonella* had not previously been performed.

326 While P1-like phage-plasmids have been reported in a small number of individual strains of
327 *Salmonella* (21–23), previous large-scale genomic surveillance studies of the
328 Enterobacterales have primarily focused on *E. coli*. Our work shows that the overall
329 prevalence of P1-like phage-plasmids among *Salmonella* appears to be considerably lower
330 than in *E. coli*, being just ~0.5% compared to 7-12.6% (12,16). However, this may vary
331 depending on country or isolate source, as a higher prevalence has been previously found in
332 *Salmonella* from pork suggesting an association between the phage-plasmid sequences and
333 their host strains, with closely related *S. enterica* isolates (<5 SNPs) possessing identical or
334 extremely similar phage-plasmids in China (21), though only among specific strains of *S.*
335 Derby and *S.* Indiana sampled from the same facility.

336 We confirmed the presence of P1 phage-plasmids in multiple *Salmonella* serovars, with the
337 majority (84%) occurring in *S.* Typhimurium. This represents a significantly higher fraction
338 than the fraction of *S.* Typhimurium recorded in our surveillance database during the same
339 time period. Interestingly, it has been reported that isolates belonging to *S.* Typhimurium
340 have a higher prophage burden than many other serovars of *Salmonella* (40), possibly
341 because phage-mediated recombination plays a role in the periodic emergence of epidemic
342 clones of *S.* Typhimurium (41). We also find in this study that other serovars seem to be
343 especially amenable to acquiring other such elements – examples are *S.* Oslo, *S.* Livingstone
344 and *S.* Indiana. Perhaps even more importantly, we notice that some serovars very rarely
345 acquire phage-plasmids, with the fraction of isolates containing phage-plasmids being much
346 smaller than what would be expected from the fraction of surveillance isolates. That is
347 especially true for *S.* Enteritidis (0.44% vs. 27%, or ~50 times less); other examples are *S.*
348 Agona, *S.* Infantis and *S.* Typhi. While the exact mechanism of affinity between serovars and
349 phage-plasmids is likely to depend on several factors, we speculate that a broad link with
350 the level of host specificity of an organism might exist, with generalists like *S.* Typhimurium
351 and *E. coli* requiring a higher level of genome adaptability via acquisition of malleable
352 extrachromosomal mobile elements; though this does not explain the low prevalence
353 observed in *S.* Enteriditis.

354 Similar to previous studies of P1-like phage-plasmids, we report that the majority of those
355 detected in our dataset fall into a narrow size range of 85.8-99.5kb, with the 92.6 kb

356 average matching well to those from other collections. Others have suggested that this
357 small size range may be due to a limit on the amount of genetic material that can be
358 packaged within the virion head, although there have been reports of larger phage-plasmids
359 ranging up to 155kb (42). The structure of the P1 phage-plasmids when comparing between
360 *Salmonella* strains are diverse, but nonetheless share a common genomic architecture. Our
361 results, alongside those of other studies (9,21), highlight large regions of conserved
362 sequence interspaced with more variable regions, which mainly encode hypothetical genes
363 and phage related genes rather than plasmid coding sequences. However, the conserved
364 regions also contain phage genes along with some plasmid associated genes.

365 Based on SNP typing of the corresponding *Salmonella* chromosomes there is an association
366 between the phage-plasmid sequences and chromosomal diversity, with closely related (<5
367 SNPs) isolates possessing identical or highly similar phage-plasmids. We also detected
368 several P1-like phage-plasmids in online databases that share considerable sequence
369 similarity with specific sequences from our collection. These public phage-plasmid
370 sequences derive from various sources and span multiple years, continents and even host
371 species, implying widespread horizontal transmission. Another study also detected an
372 identical phage-plasmid that was present in two serovars of *S. enterica* sampled from the
373 same facility (21). In combination with the link to chromosomal relatedness this suggests
374 that horizontal transfer may be followed by maintenance of the phage-plasmids during
375 clonal expansion of strains, with subsequent loss of the phage-plasmid occurring
376 sporadically in some isolates, rather than there having been long-term vertical transmission
377 and evolution at a serovar or species level. This is reinforced by our discovery that there is
378 only limited phylogenetic clustering of the two different plasmid replicon types: one type is
379 maintained across each shallow clade but frequent switching in replication type breaks this
380 association when considering deeper lineages. Additionally, our results suggest that overall
381 genetic diversity of the phage-plasmids is not consistent with the corresponding population
382 structure of their hosts at either a serovar or species level, which has also been observed in
383 other studies (43) and is indicative of their extensive recombination and horizontal method
384 of acquisition.

385 Phage-plasmids are known to be involved in the horizontal transmission of AMR genes,
386 antiseptic resistance genes (10) and virulence factors such as toxins (44), as well as being
387 involved in bacterial survival (42). The P1-like family in particular is the most common type
388 of phage-plasmid to be associated with AMR gene presence (10), with a prevalence that lies
389 between that of phages and plasmids. In our collection, P1-like phage-plasmids were
390 present both in isolates that were genotypically resistant and susceptible to antibiotics, with
391 patterns depending on the serovar. Most of these AMR genes were observed to be located
392 on the chromosome, and so while it is difficult to be certain without utilising long-read
393 sequencing, we did not find an association with AMR in almost all of the phage-plasmids in
394 this study. However, we previously described an IncY P1-like phage-plasmid carrying bla_{CTX-}
395 $_{M-15}$ (22) and in our current study we identified an identical phage-plasmid sequence in an

396 isolate that is genetically closely related as well as epidemiologically connected. This finding
397 provides further evidence of transduction, or maintenance during clonal expansion while
398 likely under the same selective pressures. Importantly, our previously described
399 antimicrobial susceptibility testing indicated phenotypic ESBL resistance, confirming the
400 functionality of the AMR genes present on phage-plasmids. The flanking regions and
401 associated IS elements responsible for mobilisation of *bla*CTX-M gene variants are widely
402 discussed in the literature (45,46). The two *S.* Typhi phage-plasmids described here possess
403 $bla_{\text{CTX-M-15}}$ flanked by ISEc9/ISEcp1, which is also implicated in many plasmids carrying
404 various ESBL resistance genes from multiple species (47,48); further evidence that phage-
405 plasmids can acquire AMR genes from diverse sources (43).

406 The presence of AMR-conferring genes in phage-plasmids lends itself to a straightforward
407 evolutionary explanation, whereby latent phage-plasmids would increase the fitness of their
408 host when antibiotics are present in the environment. Once the selective pressure of
409 antibiotics is removed, the phage-plasmids can enter a lytic stage to avoid the fitness cost
410 associated with resistance (49). However, many of the phage-plasmids we identified lack
411 genes with an explicit or known AMR function, and mostly contain unannotated genes. This
412 raises the obvious question of why these genes would be kept and propagated at all, given
413 that phages can only carry a limited payload and maintenance is likely to come at a cost for
414 the host. In fact, this line of thought strongly suggests that those genes with unknown
415 function might be particularly important, despite their function being unknown yet, and that
416 active surveillance of their appearance and distribution might be a valuable tool to
417 understand how bacteria harbouring phage-plasmids will evolve in the future. The unknown
418 genes carried by phage-plasmids might act as a reservoir of foreign material, allowing a
419 rapid response to selective environmental pressures such as antimicrobials, host defence
420 mechanisms or available substrates. The previously described spontaneous loss of the
421 $bla_{\text{CTX-M-15}}$ module from the *S.* Typhi phage-plasmid (22) may be evidence of such a dynamic
422 nature. It is also possible that P1-like phage-plasmids are able to be maintained at a
423 relatively low fitness cost to the bacterium due to their low copy number and restricted size.
424 Furthermore, as prophages have been implicated in driving the evolution of *Salmonella* via
425 re-assortment of virulence and fitness factors to form new pathogenic variants (50), it is
426 possible that phage-plasmids too may play a role in this process. Other recent work has also
427 highlighted phage-plasmids – in particular P1-like phage-plasmids – as mediating gene flow
428 between phages and plasmids, resulting in the creation of novel elements (43).

429 Our results would not be possible without the creation of a robust and scalable workflow
430 that allowed us to reprocess in one hour short-read sequencing data from hundreds of
431 isolates on a single high performance computing (HPC) node. Our method allows an
432 arbitrary number of genes to be detected, with high sensitivity and accuracy, and, given a
433 sufficient computational budget, it can be scaled up to an essentially arbitrary number of
434 samples. The *repL* gene has previously been used as a target for screening using both *in vitro*
435 (9,21,23) and *in silico* (16,23) based methods. We build on these to prove that *repL* presence

436 is a valid detection method to quickly screen very large genomic datasets for the presence
437 of P1-like phage-plasmids. A previous study did however detect an atypical P1-like phage-
438 plasmid that was missing its *repL* gene (16), suggesting that additionally targeting the
439 IncY/p0111 replication genes may enhance the robustness of the method.

440 The P1-like family of phage-plasmids was the target of this work due to its proven links with
441 AMR gene carriage, though we are determined to expand the scope of our surveillance to
442 include screening other phage-plasmid types including SSU5 (51,52) that are known to
443 encode AMR genes (10,53) as well as other enteric pathogens received by the UK reference
444 laboratory such as *E. coli* and *Shigella* spp.. Active surveillance of plasmids, plasmid replicon
445 types and AMR gene presence is already part of our routine bioinformatics pipeline, and so
446 the implementation of screening for phage-plasmids would be a valuable complement to
447 this service. Additionally, this method could be used to rapidly publicly available genome
448 collections, for example on the Sequence Read Archive (54) or Enterobase (55), to get a
449 more accurate idea of global P1-like phage-plasmid prevalence within *Salmonella* and other
450 species. Finally, future studies of phage-plasmids will benefit from the application of
451 techniques, such as long-read sequencing, that are able to capture much more easily the
452 structure and genomic architecture of complex plasmid-like elements, including those
453 carrying AMR modules.

454 ## Conclusions

455 We developed and applied a robust, high-throughput method to perform the first large-
456 scale genomic surveillance of P1-like phage-plasmids in a national collection of *Salmonella*
457 isolates. Although the prevalence is lower than what has been reported for other species of
458 Enterobacterales, the phage-plasmids we identified are present across a number of
459 serovars. They carry genes conferring not only AMR, but also several so far unknown and
460 potentially relevant traits. The apparent ease of transmission across serovars, species and
461 continents, and potential to integrate into the chromosome, highlight the importance to
462 routinely monitor the presence of such phage-plasmids. We have expanded the knowledge
463 of P1-like phage-plasmids by linking their sequences back to host genomic diversity and
464 supplementing them with epidemiological metadata, showing that while such phage-
465 plasmids are understood to be ancient features, their acquisition by bacterial strains
466 appears to be a dynamic process driven by recent evolutionary pressure and so it is
467 important to examine them in the context of their host strains. This study highlights the
468 necessity of regular sequence data curation for the detection, characterisation and tracking
469 of novel mobile elements involved in AMR transmission and other factors that can change
470 the biological properties of bacterial pathogens.

Conflicts of interest

The authors declare that there are no conflicts of interest.

Funding Information

This study is funded by the National Institute for Health Research (NIHR) Health Protection Research Unit in Genomics and Enabling Data at University of Warwick in partnership with the UK Health Security Agency (UKHSA), in collaboration with University of Cambridge and Oxford. Paolo Ribeca, Clare Barker and Marie Chattaway are based at UKHSA. Caitlin Collins is based at UKHSA and University of Cambridge. Xavier Didelot is based at University of Warwick. Matthew Bird is affiliated to the NIHR Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance at University of Oxford in partnership with UKHSA, and is based at UKHSA. David Greig, Xavier Didelot and Paolo Ribeca are affiliated to the NIHR Health Protection Research Unit in Gastrointestinal Infections at University of Liverpool in partnership with UKHSA, in collaboration with University of Warwick. David Greig is based at UKHSA. The views expressed are those of the author(s) and not necessarily those of the NIHR, the Department of Health and Social Care or the UK Health Security Agency.

488 # References

489 1. European Centre for Disease Prevention and Control., World Health Organization. Antimicrobial
490 resistance surveillance in Europe 2022: 2020 data. [Internet]. LU: Publications Office; 2022 [cited
491 2023 Aug 2]. Available from: https://data.europa.eu/doi/10.2900/112339

492 2. European Food Safety Authority, European Centre for Disease Prevention and Control. The
493 European Union Summary Report on Antimicrobial Resistance in zoonotic and indicator bacteria
494 from humans, animals and food in 2019–2020. EFSA J [Internet]. 2022 Mar [cited 2023 Aug
495 2];20(3). Available from: https://data.europa.eu/doi/10.2903/j.efsa.2022.7209

496 3. Partridge SR, Kwong SM, Firth N, Jensen SO. Mobile Genetic Elements Associated with
497 Antimicrobial Resistance. Clin Microbiol Rev. 2018 Aug;31(4):10.1128/cmr.00088-17.

498 4. Brown-Jaque M, Calero-Cáceres W, Muniesa M. Transfer of antibiotic-resistance genes via phage-
499 related mobile elements. Plasmid. 2015 May 1;79:1–7.

500 5. Balcazar JL. Bacteriophages as Vehicles for Antibiotic Resistance Genes in the Environment. PLOS
501 Pathog. 2014 Jul 31;10(7):e1004219.

502 6. Łobocka MB, Rose DJ, Plunkett G, Rusin M, Samojedny A, Lehnherr H, et al. Genome of
503 Bacteriophage P1. J Bacteriol. 2004 Nov;186(21):7032–68.

504 7. Gilcrease EB, Casjens SR. The genome sequence of Escherichia coli tailed phage D6 and the
505 diversity of Enterobacteriales circular plasmid prophages. Virology. 2018 Feb 1;515:203–14.

506 8. Pfeifer E, Moura de Sousa JA, Touchon M, Rocha EPC. Bacteria have numerous distinctive groups
507 of phage–plasmids with conserved phage and variable plasmid gene repertoires. Nucleic Acids
508 Res. 2021 Mar 18;49(5):2655–73.

509 9. Wang M, Jiang L, Wei J, Zhu H, Zhang J, Liu Z, et al. Similarities of P1-Like Phage Plasmids and
510 Their Role in the Dissemination of blaCTX-M-55. Microbiol Spectr. 2022 Sep 7;10(5):e01410-22.

511 10. Pfeifer E, Bonnin RA, Rocha EPC. Phage-Plasmids Spread Antibiotic Resistance Genes through
512 Infection and Lysogenic Conversion. mBio. 2022 Sep 26;13(5):e01851-22.

513 11. Chee MSJ, Serrano E, Chiang YN, Harling-Lee J, Man R, Bacigalupe R, et al. Dual pathogenicity
514 island transfer by piggybacking lateral transduction. Cell. 2023 Aug 3;186(16):3414-3426.e16.

515 12. Billard-Pomares T, Fouteau S, Jacquet ME, Roche D, Barbe V, Castellanos M, et al.
516 Characterization of a P1-Like Bacteriophage Carrying an SHV-2 Extended-Spectrum β-Lactamase
517 from an Escherichia coli Strain. Antimicrob Agents Chemother. 2014 Nov;58(11):6550–7.

518 13. Bai L, Wang J, Hurley D, Yu Z, Wang L, Chen Q, et al. A novel disrupted mcr-1 gene and a
519 lysogenized phage P1-like sequence detected from a large conjugative plasmid, cultured from a
520 human atypical enteropathogenic Escherichia coli (aEPEC) recovered in China. J Antimicrob
521 Chemother. 2017 May 1;72(5):1531–3.

522 14. Li R, Xie M, Lv J, Wai-Chi Chan E, Chen S. Complete genetic analysis of plasmids carrying mcr-
523 1 and other resistance genes in an Escherichia coli isolate of animal origin. J Antimicrob
524 Chemother. 2017 Mar 1;72(3):696–9.

525 15. Zhang C, Feng Y, Liu F, Jiang H, Qu Z, Lei M, et al. A Phage-Like IncY Plasmid Carrying the mcr-
526 1 Gene in Escherichia coli from a Pig Farm in China. Antimicrob Agents Chemother. 2017 Feb
527 23;61(3):e02035-16.

528 16. Venturini C, Zingali T, Wyrsch ER, Bowring B, Iredell J, Partridge SR, et al. Diversity of P1
529 phage-like elements in multidrug resistant Escherichia coli. Sci Rep. 2019 Dec 11;9(1):18861.

530 17. Han H, Liu W, Cui X, Cheng X, Jiang X. Co-Existence of mcr-1 and blaNDM-5 in an Escherichia
531 coli Strain Isolated from the Pharmaceutical Industry, WWTP. Infect Drug Resist. 2020 Apr
532 15;13:851–4.

533 18. Kalová A, Gelbíčová T, Overballe-Petersen S, Litrup E, Karpíšková R. Characterisation of
534 Colistin -Resistant Enterobacterales and Acinetobacter Strains Carrying mcr Genes from Asian
535 Aquaculture Products. Antibiotics. 2021 Jul;10(7):838.

536 19. Yao M, Zhu Q, Zou J, Shenkutie AM, Hu S, Qu J, et al. Genomic Characterization of a
537 Uropathogenic Escherichia coli ST405 Isolate Harboring blaCTX-M-15-Encoding IncFIA-FIB
538 Plasmid, blaCTX-M-24-Encoding IncI1 Plasmid, and Phage-Like Plasmid. Front Microbiol. 2022;13.

539 20. Zhou W, Liu L, Feng Y, Zong Z. A P7 Phage-Like Plasmid Carrying mcr-1 in an ST15 Klebsiella
540 pneumoniae Clinical Isolate. Front Microbiol. 2018;9.

541 21. Yang L, Li W, Jiang GZ, Zhang WH, Ding HZ, Liu YH, et al. Characterization of a P1-like
542 bacteriophage carrying CTX-M-27 in Salmonella spp. resistant to third generation cephalosporins
543 isolated from pork in China. Sci Rep. 2017 Jan 18;7:40710.

544 22. Greig DR, Bird MT, Chattaway MA, Langridge GC, Waters EV, Ribeca P, et al. Characterization
545 of a P1-bacteriophage-like plasmid (phage-plasmid) harbouring blaCTX-M-15 in Salmonella
546 enterica serovar Typhi. Microb Genomics. 2022;8(12):000913.

547 23. Jiang L, Zhu H, Wei J, Jiang L, Li Y, Li R, et al. Enterobacteriaceae genome-wide analysis
548 reveals roles for P1-like phage-plasmids in transmission of mcr-1, tetX4 and other antibiotic
549 resistance genes. Genomics. 2023 Mar 1;115(2):110572.

550 24. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data.
551 Bioinformatics. 2014 Aug 1;30(15):2114–20.

552 25. Achtman M, Wain J, Weill FX, Nair S, Zhou Z, Sangal V, et al. Multilocus Sequence Typing as a
553 Replacement for Serotyping in Salmonella enterica. PLOS Pathog. 2012 Jun 21;8(6):e1002776.

554 26. Tewolde R, Dallman T, Schaefer U, Sheppard CL, Ashton P, Pichon B, et al. MOST: a modified
555 MLST typing tool based on short read sequencing. PeerJ. 2016 Aug 17;4:e2308.

556 27. Dallman T, Ashton P, Schafer U, Jironkin A, Painset A, Shaaban S, et al. SnapperDB: a
557 database solution for routine sequencing analysis of bacterial isolates. Bioinformatics. 2018 Sep
558 1;34(17):3028–9.

559 28. Marco-Sola S, Sammeth M, Guigó R, Ribeca P. The GEM mapper: fast, accurate and versatile
560 alignment by filtration. Nat Methods. 2012 Dec;9(12):1185–8.

561 29. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New
562 Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol. 2012
563 May;19(5):455–77.

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

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

568 32. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, et al. Producing
569 polished prokaryotic pangenomes with the Panaroo pipeline. Genome Biol. 2020 Jul 22;21(1):180.

570 33. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for
571 Large Alignments. PLOS ONE. 2010 Mar 10;5(3):e9490.

572 34. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. Bakta: rapid and
573 standardized annotation of bacterial genomes via alignment-free sequence identification. Microb
574 Genomics. 2021 Nov 5;7(11):000685.

575 35. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. Bioinformatics.
576 2011 Apr 1;27(7):1009–10.

577 36. Gilchrist CLM, Chooi YH. clinker & clustermap.js: automatic generation of gene cluster
578 comparison figures. Bioinformatics. 2021 Aug 25;37(16):2473–5.

579 37. Joy JB, Liang RH, McCloskey RM, Nguyen T, Poon AFY. Ancestral Reconstruction. PLOS
580 Comput Biol. 2016 Jul 12;12(7):e1004763.

581 38. Schliep KP. phangorn: phylogenetic analysis in R. Bioinformatics. 2011 Feb 15;27(4):592–3.

582 39. Collins C, Didelot X. A phylogenetic method to perform genome-wide association studies in
583 microbes that accounts for population structure and recombination. PLOS Comput Biol. 2018 Feb
584 5;14(2):e1005958.

585 40. Mottawea W, Duceppe MO, Dupras AA, Usongo V, Jeukens J, Freschi L, et al. Salmonella
586 enterica Prophage Sequence Profiles Reflect Genome Diversity and Can Be Used for High
587 Discrimination Subtyping. Front Microbiol. 2018;9.

588 41. Brüssow H, Canchaya C, Hardt WD. Phages and the Evolution of Bacterial Pathogens: from
589 Genomic Rearrangements to Lysogenic Conversion. Microbiol Mol Biol Rev. 2004 Sep;68(3):560–
590 602.

591 42. Kamal SM, Cimdins-Ahne A, Lee C, Li F, Martín-Rodríguez AJ, Seferbekova Z, et al. A recently
592 isolated human commensal Escherichia coli ST10 clone member mediates enhanced
593 thermotolerance and tetrathionate respiration on a P1 phage-derived IncY plasmid. Mol
594 Microbiol. 2021;115(2):255–71.

595 43. Pfeifer E, Rocha EPC. Phage-plasmids promote genetic exchanges between phages and
596 plasmids and create novel ones [Internet]. bioRxiv; 2023 [cited 2023 Sep 12]. p.
597 2023.08.08.552325. Available from:
598 https://www.biorxiv.org/content/10.1101/2023.08.08.552325v1

599 44. Lindler LE, Plano GV, Burland V, Mayhew GF, Blattner FR. Complete DNA Sequence and
600 Detailed Analysis of the Yersinia pestis KIM5 Plasmid Encoding Murine Toxin and Capsular
601 Antigen. Infect Immun. 1998 Dec;66(12):5731–42.

602 45. Bird MT, Greig DR, Nair S, Jenkins C, Godbole G, Gharbia SE. Use of Nanopore Sequencing to
603 Characterise the Genomic Architecture of Mobile Genetic Elements Encoding blaCTX-M-15 in
604 Escherichia coli Causing Travellers' Diarrhoea. Front Microbiol [Internet]. 2022 [cited 2023 Sep
605 18];13. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2022.862234

606 46. Nair S, Chattaway M, Langridge GC, Gentle A, Day M, Ainsworth EV, et al. ESBL-producing
607 strains isolated from imported cases of enteric fever in England and Wales reveal multiple
608 chromosomal integrations of blaCTX-M-15 in XDR Salmonella Typhi. J Antimicrob Chemother.
609 2021 Jun 1;76(6):1459–66.

610 47. Zeng S, Luo J, Chen X, Huang L, Wu A, Zhuo C, et al. Molecular Epidemiology and
611 Characteristics of CTX-M-55 Extended-Spectrum β-Lactamase-Producing Escherichia coli From
612 Guangzhou, China. Front Microbiol [Internet]. 2021 [cited 2023 Sep 18];12. Available from:
613 https://www.frontiersin.org/articles/10.3389/fmicb.2021.730012

614 48. Fernandes MR, Sellera FP, Cunha MPV, Lopes R, Cerdeira L, Lincopan N. Emergence of CTX-
615 M-27-producing Escherichia coli of ST131 and clade C1-M27 in an impacted ecosystem with
616 international maritime traffic in South America. J Antimicrob Chemother. 2020 Jun 1;75(6):1647–
617 9.

618 49. Andersson DI, Hughes D. Antibiotic resistance and its cost: is it possible to reverse
619 resistance? Nat Rev Microbiol. 2010 Apr;8(4):260–71.

620 50. Figueroa-Bossi N, Uzzau S, Maloriol D, Bossi L. Variable assortment of prophages provides a
621 transferable repertoire of pathogenic determinants in Salmonella. Mol Microbiol.
622 2001;39(2):260–72.

623 51. Octavia S, Sara J, Lan R. Characterization of a large novel phage-like plasmid in Salmonella
624 enterica serovar Typhimurium. FEMS Microbiol Lett. 2015 Apr 1;362(8):fnv044.

625 52. Colavecchio A, Jeukens J, Freschi L, Edmond Rheault JG, Kukavica-Ibrulj I, Levesque R, et al.
626 AnCo3, a New Member of the Emerging Family of Phage-Like Plasmids. Genome Announc. 2017
627 May 11;5(19):10.1128/genomea.00110-17.

628 53. Colavecchio A, Jeukens J, Freschi L, Edmond Rheault JG, Kukavica-Ibrulj I, Levesque RC, et al.
629 Complete Genome Sequences of Two Phage-Like Plasmids Carrying the CTX-M-15 Extended-
630 Spectrum β-Lactamase Gene. Genome Announc. 2017 May 11;5(19):10.1128/genomea.00102-17.

631 54. Katz K, Shutov O, Lapoint R, Kimelman M, Brister JR, O'Sullivan C. The Sequence Read
632 Archive: a decade more of explosive growth. Nucleic Acids Res. 2022 Jan 7;50(D1):D387–90.

633 55. Zhou Z, Alikhan NF, Mohamed K, Fan Y, Group the AS, Achtman M, et al. The EnteroBase
634 user's guide, with case studies on Salmonella transmissions, Yersinia pestis phylogeny, and
635 Escherichia core genomic diversity. Genome Res. 2020 Jan 1;30(1):138–52.

636

637

638

639

640

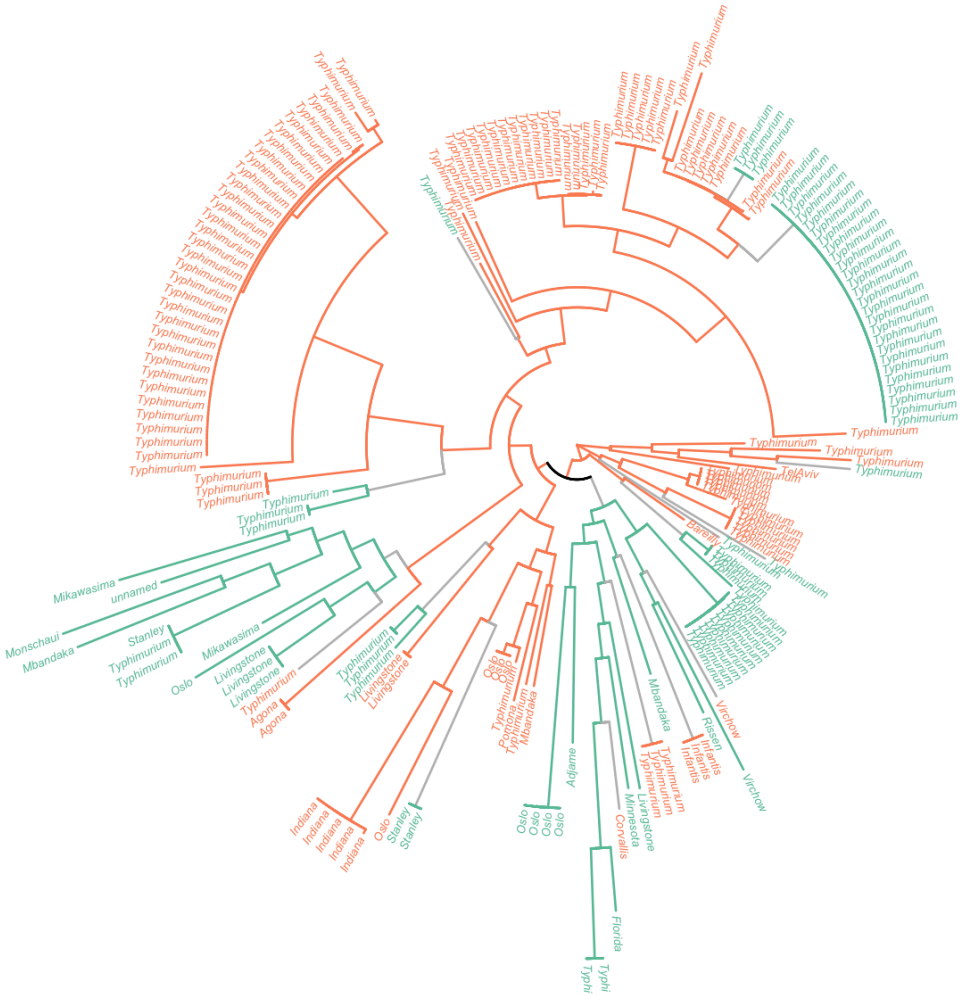

Fig. S1 Phylogenetic tree based on the 51,260bp alignment of 54 core genes of the 193 repL-positive phage-plasmid contigs (Fig. 1) coloured by plasmid replication type, as observed at the tips and inferred via ancestral state reconstruction along the branches (IncY = green, p0111 = orange), showing 17 changes in replication type state (IncY>p0111 (5/17) or p0111>IncY (12/17) = grey).

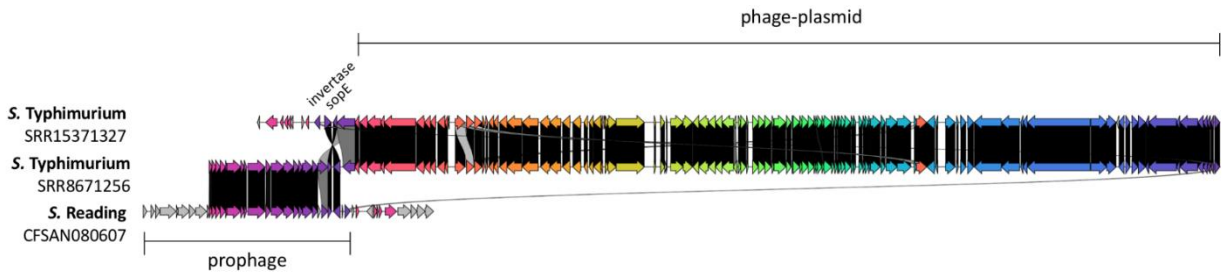

Fig. S2 Phage-plasmids in *S.* Typhimurium putatively integrated in chromosome at site of prophage.

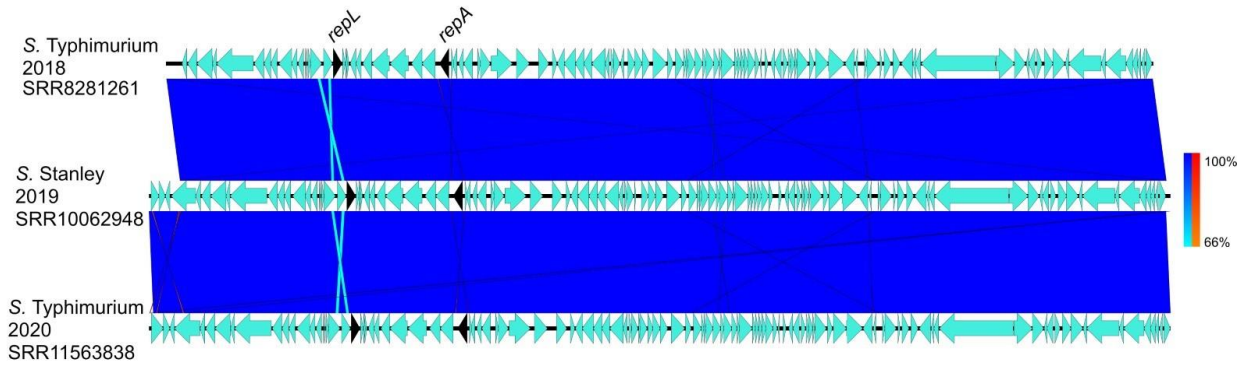

Fig. S3 Highly similar phage-plasmid sequences in different serovars: *S.* Stanley from 2019, and *S.* Typhimurium from 2018 & 2020. Arrows indicate gene direction and scale bars indicate level of similarity for forward (blue) and reverse (red) sequences.

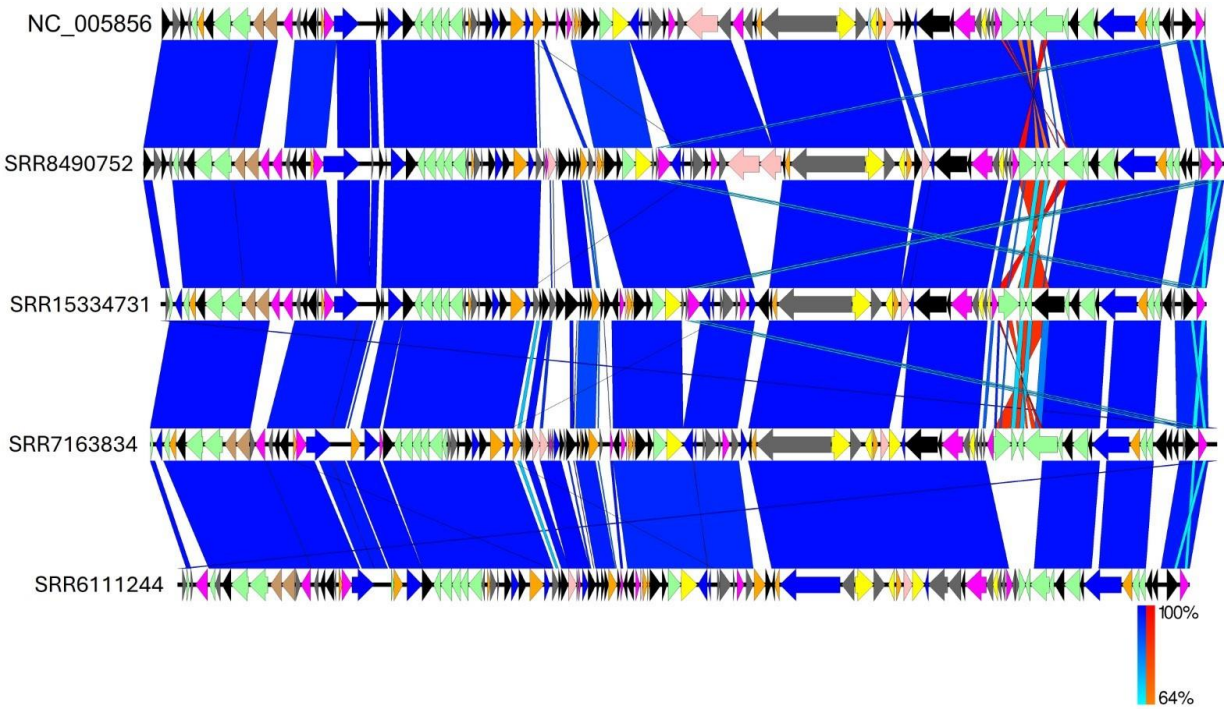

Fig. S4 Variation in selected phage-plasmids from multiple serovars (*S.* Infantis, *S.* Livingstone*, S.* Monschaui and *S.* Typhimurium) compared to P1 bacteriophage. Arrows indicate gene direction and scale bars indicate level of similarity for forward (blue) and reverse (red) sequences. Hypothetical proteins are shown in grey; regulator genes are shown in blue; effector virulence genes are shown in pink; Phage-related genes are shown in light sky blue; Plasmid-related genes are shown in brown; Recombinant genes are shown in magenta; Metabolism/Signalling genes are shown in orange; Structure genes are shown in pale green; lysis genes are shown in yellow; uncharacterised or not well-established genes are shown in black.

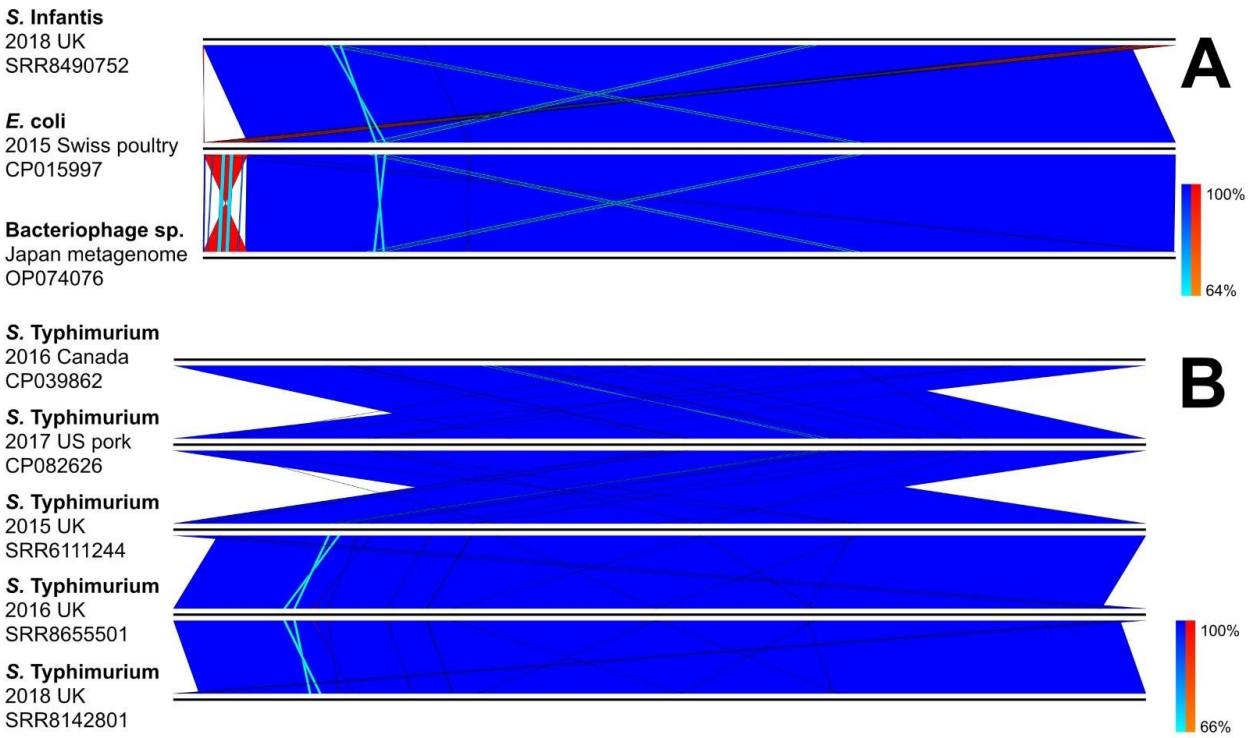

Fig. S5 Highly similar phage-plasmid sequences from public databases showing conserved sequence across: **A)** different species; **B)** isolates of *S.* Typhimurium from different years and countries.