Phage-plasmid borne methionine tRNA ligase mediates epidemiologically relevant antimicrobial persistence

Yi Ling Tam*[1], P. Malaka De Silva*[2], Clare R. Barker[3,11], Ruizhe Li[4], Leanne Santos[5], Gherard Batisti Biffignandi[2,6], Charlotte E. Chong[2], Lewis C. E. Mason[7,8], Satheesh Nair[9,10], Paolo Ribeca[3,8,10,11], Sion C. Bayliss[12], Claire Jenkins[8,9], Somenath Bakshi[4], James P.J. Hall[5], Lauren Cowley[1], Kate S. Baker[^2,8]

* authors contributed equally ^ author to whom correspondence should be addressed: kb827@cam.ac.uk

Affiliations

1. Milner Centre of Evolution, Department of Life Sciences, University of Bath
2. Department of Genetics, University of Cambridge, Cambridge, CB2 3EH
3. Gastrointestinal Infections & Food Safety (One Health), UK Health Security Agency, UK
4. Department of Engineering, University of Cambridge, Trumpington St, Cambridge, CB21PZ
5. Department of Evolution, Ecology and Behaviour, Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool
6. European Bioinformatics Institute, European Molecular Biology Laboratory, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, United Kingdom
7. Department of Clinical Infection, Microbiology, and Immunology; Institute for Infection, Veterinary and Ecological Sciences, Liverpool, UK, L697ZB
8. NIHR HPRU in Gastrointestinal Infections at University of Liverpool, Liverpool, UK
9. Gastrointestinal Bacteria Reference Unit, UK Health Security Agency, UK
10. NIHR HPRU in Genomics & Enabling Data at University of Warwick, Warwick, UK
11. Biomathematics and Statistics Scotland, The James Hutton Institute, Edinburgh, UK
12. Bristol Veterinary School, Faculty of Health and Life Sciences, University of Bristol, UK.

Abstract

Antimicrobial resistance (AMR) is a global public health crisis with few options for control. As such, moving to predictive frameworks for identification of emerging bacterial strains capable of rapidly evolving AMR for early intervention is key. Although antimicrobial tolerance and persistence are thought to be precursor phenotypes for AMR, little evidence exists to support their importance in real-world scenarios. Here we leveraged national genomic surveillance data of the diarrhoeal pathogen *Shigella sonnei* (n=3745) to agnostically identify common genetic signatures among lineages convergently evolving toward AMR (n=15) using bacterial genome-wide association. This revealed an association of an AMR trajectory with a multi- and highly variable second copy of *metG*, borne on a phage-plasmid we called pWPMR2. Further bioinformatic analyses revealed that pWPMR2 was present across clinical isolates of other enteric pathogens globally, including previous major outbreaks. And, that the mechanism of bearing additional *metG* copies on mobile genetic elements was present across multiple bacterial phyla. Subsequent functional microbiology and experimental evolution studies revealed that the expression of additional *metG*, particularly the mutated version on pWPMR2, created a sub population of cells with persister phenotypes that predispose them to the evolution of resistance to third generation cephalosporins. This highlights that the provision of *metG* in trans predisposes bacteria to AMR with real world impacts, likely across a broad range of clinically relevant pathogens. As well as offering a warning sign for emerging AMR lineages, our approach is a timely exemplar of how genomic epidemiology frameworks can rapidly guide functional microbiology studies in the coming era of routine genomic surveillance.

Background

Antimicrobial resistance (AMR) is a global public health crisis, and microbial genomics has revealed that certain highly AMR bacterial lineages dominate epidemiologically. For example, *S. Typhimurium* sequence type 313 (ST313) in African invasive disease (1), the pandemic *E. coli* sequence type 131 (ST131) (2–5), the carbapenemase-associated *K. pneumoniae* sequence type 258 (6), and Lineage III *S. sonnei* of which extensively drug resistant (XDR) forms have emerged (7–9). Once AMR has been acquired and disseminated in these successful clones, the resulting treatment complications make their reduction challenging. In an ideal world, we would use the imminent tsunami of genomic surveillance data to move toward predictive frameworks of AMR surveillance in which clones that are predisposed to AMR could be identified for timely intervention. However, this requires us to understand the precursors or 'stepping stones' to the evolution and acquisition of AMR.

Here, we leveraged an epidemiological scenario of convergent evolution of AMR among lineages of a highly clonal pathogen to agnostically identify pathogen-factors associated with an AMR trajectory. Specifically, the severe diarrhoeal pathogen, *Shigella*, has emerged over the last few decades as a sexually transmissible illness (STI) among men who have sex with men (MSM) (10). This is particularly detectable in traditionally non-endemic high-income countries, such as the United States and the United Kingdom, where an over-representation of closely related isolates derived from males without a recent history of travel is used to identify lineages that are circulating in sexual transmission networks (STN) (11,12). Studies have shown that at a given time, there are multiple co-circulating STN Lineages of *Shigella* (9,13,14) and that these are subject to high levels of antimicrobial selection pressure. This pressure is evidenced by convergent acquisition of AMR plasmids and the development of bystander resistance, thought to result from high intensity treatment for other traditional STIs (9,10,14–16). Concerningly, it is also clear that these lineages spread globally over short time frames and share plasmids among themselves and with other pathogen groups. Thus, convergently evolving genetic traits among co-circulating STN Lineages of *Shigella* represent potential adaptations on a trajectory to AMR that have the potential to spread widely.

In this study, we identify convergently evolving factors across 15 STN lineages among some 3,745 *Shigella sonnei* isolates from routine surveillance in the United Kingdom using bacterial Genome Wide Association Study (bGWAS). This identified variant copies of methionine tRNA ligase carried on an SSU5-like phage-plasmid (which we called pWPMR2) as associated with STN Lineages (as a proxy for high AMU). We demonstrate how auxiliary *metG* expression and variation create bacteria with phenotypes that predispose bacteria to evolve AMR, particularly against third generation cephalosporins. Further *in silico* analyses showed that both pWPMR2 and other plasmid-borne *metG* are spread globally in other clinically relevant bacterial species, including contemporary outbreaks, and other bacterial families and phyla.

Phage-plasmid borne *metG* is associated with an AMR trajectory

To identify genetic features associated with AMR emergence, we first identified convergently evolving STN lineages of *S. sonnei*. To do this, we constructed a recombination-free maximum-likelihood phylogenetic tree of S. *sonnei* isolates (n=3745) from surveillance in the United Kingdom between 2004 and 2020 (Figure 1A, Extended data). Adapting validated approaches, we then used demographic data on patients' age, sex, and travel history for

ancestral state reconstruction to identify 15 lineages circulating in STN (hereafter STN lineages, Figure 1A, Table1, Extended Data, Methods). This identified 15 STN lineages of variable sizes which, consistent with the increased antimicrobial selection pressure in STN, had a significantly higher proportion of isolates containing genes conferring AMR against (any of) the three main clinically relevant antimicrobials for *Shigella* (Mason, L.C.E, personal communication); ciprofloxacin (CIP), azithromycin (AZM), and ceftriaxone (CRO) (74% [n=765/1028] compared with 37% [n=1011/2717], p-value <0.001, Table 1, Supplementary Data). Alongside the distribution of these resistances in five, six, and five of the 15 STN lineages respectively, this confirms that, consistent with previous studies, *Shigella* circulating in STN are convergently evolving toward AMR (10,14,16–19).

We then quantified the association of unique genetic sequences (i.e. unitigs from a compiled assembly graph of the 3745 isolates) with membership of STN Lineages using treeWAS (20). We identified a total of 229,603 unitigs which had varying degrees of association with STN lineages (Figure 1B). This approach was deemed valid as unitigs related to AZM resistance genes (*mphA*, n=157 and *ermB,* n=54) had some of the highest association scores (Figure 1B), and AZM resistance is known to have emerged globally among *Shigella* in STN (Figure 1A, (10,14,16–19). Among the highly associated unitigs (>0.66 association score, Extended data) we also found sequences (n=357) relating to *metG*, which encodes the methionine tRNA ligase, MetG (also known as methionine tRNA synthetase). MetG plays a core role in translation by linking the start codon amino acid, methionine, to its cognate tRNA molecule. Disruptions in *metG* have been studied in laboratory settings and are associated with altered growth characteristics and antimicrobial tolerance (i.e. short-term survival against killing concentrations of antimicrobials without a shift in AMR phenotype (21–24), and predisposition to AMR (25). Thus, we hypothesised that the auxiliary metG was playing a role in the adaptation of *Shigella* lineages to circulation in the high antimicrobial environment of STNs.

STN Lineage*	**Isolates (n)**	**pWPMR2-bearing n (%)**	**Dominant genotype^ (% of lineage)**
1	580	463 (80)	3.6.1.1.2 (100)
2	22	2 (9)	3.6.1.1.3.1 (77)
3	13	11 (85)	3.6.1.1 (100)
4	6	0 (0)	3.6.1.1 (100)
5	27	12 (44)	3.6.1.1.3.1 (85)
6	5	2 (40)	3.6.1 (100)
7	5	0 (0)	3.6.3 (100)
8	9	0 (0)	3.6.2 (100)
9	176	37 (21)	3.7.25 (100)
10	15	5 (33)	3.7.18 (100)
11	11	0 (0)	3.7.4 (100)
12	125	27 (22)	3.7.29.1.2 (66)
13	6	0 (0)	3.4.1 (100)
14	14	6 (43)	3.1 (100)
15	14	1 (7)	2.11.4 (71)
STN Lineages subtotal	1028	566 (55)	
Non STN Lineage	2717	84[&] (0.03)	

Total	3745	650 (17)

Table 1. Characteristics of *S. sonnei* lineages circulating in STNs
* arbitrarily named from top to bottom of tree in Figure 1A – see Supplementary Table for individual assignments
^ genotyping performed according to (19)
[&] 53 of which were pWPMR2var

Bioinformatic analyses to explore the association with *metG* revealed a combination of both copy number variation and single nucleotide polymorphisms (SNPs) for the gene (Extended data) which we explored in turn. Complementary long-read sequencing of STN Lineage 12 strain SRR5005407 (Figure 1A) revealed an extra copy of *metG* was carried on a 108,107bp circular sequence containing both plasmid, and phage sequences (Figure 1C), i.e. a phage-plasmid, which we called pWPMR2 (GenBank accession: PQ180110). Reference mapping to the chromosome and pWPMR2 then revealed variation in both copy number (relative depth of pWPMR2 to chromosomal mapping coverage) and coverage (proportion of pWMPR2 sites covered) of pWPMR2 among the 3,745 *S. sonnei* isolates (Extended data). By employing cut offs (0.22 relative depth, 0.7 coverage), we inferred that 650 *S. sonnei* isolates contained pWPMR2 which were mostly (87%, n=566/650, $p<2.2e-16$) associated with STN lineages (Figure 1A, Table 1, Extended data). While most pWPMR2-containing isolates contained one to two copies of pWPMR2, there was a wide range of relative depth, suggesting some instability in pWPMR2 carriage, consistent with the acquisition of an essential gene (26), (Extended data). Mapping coverage further suggested the presence of a *metG*-bearing pWPMR2 variant with reduced (~0.75) coverage (Figure 1A, Table 1, Extended data), which we sequenced from isolate SRR10313630 (GenBank accession: TBC). Consistent with it being a phage-plasmid, pWPMR2 contained an anti-defence system, RAD, previously described as targeting retrons (27). No known AMR or virulence genes were present on pWPMR2 that might explain its high prevalence, so we continued with our GWAS-guided hypothesis that auxiliary *metG* expression was driving a phenotypic shift relevant for adaptation to antimicrobial selection pressure.

An analysis of SNPs present across *metG* among the *S. sonnei* surveillance isolates revealed that the auxiliary *metG* on pWPMR2 exhibited significant genetic variation relative to the essential chromosomal version. Specifically, isolates without pWPMR2 (i.e. only chromosomal *metG*) showed high levels of gene conservation, with only 12 of 3095 isolates containing SNPs across five *metG* positions; four of which introduced non-synonymous changes in non-STN Lineages isolates (not proximate to STN Lineages), and one synonymous change in seven isolates of STN Lineage 15. Contrastingly, isolates that contained pWPMR2 had significant numbers of SNPs (0-118 per isolate, mean 66) affecting a total of 250 sites across the 2034bp gene, occurring at minor allele frequencies consistent with a low-copy number mobile genetic element (Extended data). A sliding window analysis of the variation (density of SNPs) on *metG* showed that variation was low in some key functional areas of the gene (such as the HIGH and KMSK amino acid motifs) and was comparatively higher in the last 400bp of the gene; the tRNA-binding region (Figure 1D). This pattern might suggest a protein that is able to bind the methionine substrate and less able to bind the cognate tRNA which may result in disrupted function and hamper protein production. Ultimately, exploring gene variation in multicopy genes with short read data was complex. However, complete genome sequencing confirmed that, *metG* borne by pWPMR2 contained 115 SNPs, 6 of which were non-synonymous. This highlights that the extra copy of the metG on pWPMR2 sustains a significant mutational burden relative to its normal state as an essential gene.

Mutated metG confers evolutionary advantages in the presence of ceftriaxone

To explore phenotypes associated with auxiliary *metG* expression and its reduced mutational constraint, we conducted laboratory experiments on growth characteristics and antimicrobial response phenotypes of DH5α *E. coli* mutants. Specifically, we constructed inducible vectors that expressed wildtype (i.e. *S. sonnei* chromosomal) *metG* (hereafter $metG_{WT}$) and the mutated version found on pWPMR2 (hereafter $metG_{PP}$). Using these constructs, we determined the impact of auxiliary *metG* expression (i.e. comparing $metG_{WT}$ with expressed induced and not induced) and *metG* variation (i.e. comparing induced $metG_{WT}$ with $metG_{PP}$) in various phenotypic assays.

Bulk growth experiments revealed that the induction of either $metG_{WT}$ or $metG_{PP}$ reduced bacterial growth (Figure 2A, 2B). Notably, while the expression of $metG_{WT}$ resulted in an increased lag time, consistent with a previously reported tolerance by lag phenotype (28), expression of the mutated version, $metG_{PP}$, generated diauxic growth.

We then proceeded to explore the impact of *metG* expression and variation on antimicrobial phenotypes. Expression of *metG* did not result in a change in antimicrobial resistance phenotype in that Minimum Inhibitory Concentrations (MIC) were consistent across strains (Extended data). To explore a potential benefit of in the presence of sub-inhibitory concentrations of antimicrobials, we compared the relative growth (as measured by ratios of Area Under the Curve, AUC) of constructs in a reduced concentration (0.5x MIC) of AZM, CIP and CRO. This revealed that $metG_{WT}$ expression posed no additional fitness cost in the presence of subinhibitory concentrations of CRO, despite being costly in the presence of subinhibitory concentrations AZM and CIP. Similar effects were also seen in comparing the impact of the variation found in *metG* i.e. the growth performance of the mutated *metG* on pWPMR2 ($metG_{PP}$) with the wildtype ($metG_{WT}$). This suggested that auxiliary *metG* expression, particularly the mutated form on pWPMR2, might provide a fitness advantage in coping with exposure to subinhibitory concentrations of CRO.

To explore the phenotypes conferred in tolerating higher concentrations of antimicrobials, we conducted an experimental evolution of 47 independent populations of *E. coli* expressing $metG_{WT}$ *and* $metG_{PP}$ for twelve days, doubling antimicrobial concentration daily (Methods, Figure 2C). Although all control (uninduced $metG_{WT}$) and induced $metG_{WT}$ populations died by the end of the experiment, expression of auxiliary *metG* adversely affected survival (Figure 2C top panel). However, this negative impact on survival was ameliorated in the case of expression of $metG_{PP}$ (Figure 2C bottom panel; Log-rank tests: AZM p=0.005, CIP, p=6e-6, CRO p=0.03). Similarly to the subinhibitory concentration growth experiments, the experimental evolution also suggested that $metG_{PP}$ was an advantageous trait in responding to exposure to CRO. Specifically, evolution of $metG_{PP}$ against CRO was the only condition under which surviving populations (n=4) were observed in the final exposure window (1024x MIC). This further supports the importance of $metG_{PP}$ in withstanding selective pressure conferred by CRO.

pWPMR2 and auxiliary metG are found globally other clinically relevant organisms

Having linked auxiliary *metG* expression with phenotypes relevant for the emergence of AMR, we then explored the extent to which pWPMR2 may be mobilised among clinically relevant organisms. A comparison of near-identical relatives of pWPMR2 in public databases revealed that it is associated with multiple other major outbreaks of *Shigella* and Shiga-toxigenic *E. coli* and may have links with *E. albertii.* Specifically, it was near-identical (>99% similarity) to several previously detected, though undescribed, plasmids from MSM-associated outbreaks of *S. sonnei* (16,17) and O117:H7 Shiga toxin-producing *E. coli* (STEC) (29) from a broad range of geographical sources including the UK, Australia, the United States (Figure 3A). A closely related element was also found to be present in an isolate belonging to an STEC O114:H4 outbreak in Georgia in 2009 (30), as well as an AZM-resistance bearing commensal *E. coli* strain (Sequence Type 62) from a healthy college student in the USA (31). Outside of *Shigella* and *E. coli* the closest relatives were found in *E. albertii* where multiple regions of pWPMR2 were absent (Extended data). This demonstrated that pWPMR2 is a widely distributed mobile genetic element associated with important outbreaks of clinical disease, but also detected in commensal organisms.

To further explore its association with clinically relevant organisms, we determined the taxonomy of pWPMR2 and screened all genomes of enteric infections in the UK isolated between 2016 and 2021 (n=66,929) for evidence of pWPMR2-like phage plasmids. Taxonomically, pWPMR2 was found to belong to the SSU5 super-community of phage-plasmids that are associated with species of the *Enterobacteriaceae* family (32), (Extended data). Comparison of its genome-wide sequence similarity with other phage genomes showed that pWPMR2 belonged within the pSLy3 group (Figure 3B, Extended data). The screen of clinical isolates from enteric infections (Table 2) revealed that while SSU5 phage-plasmids were present in isolates from all genera, the pSLy3-like group was primarily associated with *S. sonnei* (18% of isolates), *E. albertii* (8%), *E. coli* (2%), and other *Shigella* species (0.5%). This indicates the mobilisation of pWPMR2 across multiple enteric pathogen species, consistent with similar previous observations for AMR plasmids (14,33).

		SSU5 supercommunity phage plasmid group containing (n)					
Species	**Genomes (n)**	**pHCM2**	**pSLy3**	**pKpn[1]**	**pCAV[1]**	**pMT1[2]**	**Other**
S. enterica	47,495	539	0	0	0	0	2
E. coli	11,990	5	236	0	0	0	0
E. albertii	173	0	8	0	0	0	0
S. sonnei	3,427	2	628	0	0	0	0
S. flexneri	3,259	1	15	0	0	0	0
S. dysenteriae	196	0	0	0	0	0	0
S. boydii	289	0	4	0	0	0	0
Total	66,829	547	891	0	0	0	2

Table 2. The presence of SSU5-like phage plasmids across enteric bacteria in the UK
[1] restricted to *Klebsiella* in GenBank [2] restricted to *Yersinia* in GenBank.

Given this suggestion that pWPMR2 might be capable of spreading across pathogens, we sought to explore the temporospatial spread of pWPMR2. We first examined the temporal relationship of pWPMR2 acquisition with AMR across our 3,745 *S. sonnei* surveillance isolates (Figure 4A). This showed that pWPMR2 entered the *S. sonnei* population earlier than, and

then had similar temporal increases to, key antimicrobial resistance phenotypes (e.g. the plasmid-borne AZM and CRO resistances). This included stepwise acquisition in the internationally disseminated extensively drug-resistant Genotype 3.6.1.1.2 (STN Lineage 1 in Figure 1A, Supplementary table, Figure 1A (9)). This precursor acquisition is consistent with the work above showing that pWPMR2 confers phenotypes relevant to the emergence of AMR. To explore spatial spread, we screened for pWPMR2 in over 2 million bacterial genome sequences in the public domain (Methods), and found it was spread among some 2,028 *E. coli* and *Shigella* isolates from broad geographical areas, often in isolates derived from public health surveillance (Figure 4B, Extended data). This built further genomic epidemiological evidence that pWPMR2 is an emerging mobile genetic element carrying phenotypes of substantial public health relevance.

To explore the relevance of auxiliary *metG* in the broader bacterial context, we screened sequences (n=59,895) from the PLSDB plasmid database (34) for *metG*, and other tRNA ligases. This revealed that tRNA ligase homologues were found at low levels on plasmids (n=513/59895, 0.86%, Extended data). Following filtration for potential chromosomal contamination, only five tRNA ligases were found on an appreciable number (≥20) of plasmids: *alaS*, *cysS*, *metG*, *serS*, *thrS*. Of these, *metG* was outstanding as the most frequently observed tRNA ligase, detected more often than all the other tRNA ligases put together, and the only one found >5 times in >1 Family. Specifically, *metG* encoding plasmids were predominantly found in Enterobacteriaceae (n=172/29242, 0.59%) and Burkholderiaceae (*Ralstonia* spp., n=95/652, 14.57%). While the Enterobacteriaceae plasmids were overwhelmingly (n=165/172) recorded in the database as an FIB phage-plasmid type (consistent with pWPMR2 detection), the *Ralstonia* plasmids were putative chromids (large, domesticated plasmids associated with environmental adaptation) ranging from 1.78 Mb to 2.2 Mb ((35), Extended data). As widespread plant pathogens, *Ralstonia* can experience both environmental (e.g. produced by niche competitors) and human (e.g. through agricultural practices) derived antimicrobial selection pressure (36), so may be served by the functional advantages conferred by auxiliary *metG* expression.

Epilogue: a real-time outbreak underlines the role of pWPMR2

During the period of this work, pWPMR2 was also involved in a new sexually transmitted outbreak of XDR *S. sonnei* that took place in England in 2023. The outbreak comprised 120 cases within a 10-SNP-surveillance-cluster that occurred from 2022 – 2023 (116/120 isolates are from 2023 (37). There were three isolates among this 10-SNP-surveillance-cluster in our pre-outbreak data isolates during 2019 and 2020 (Extended data). While the isolates in the 10-SNP-surveillance-cluster contained CRO resistance, only the large 2022 – 2023 outbreak also contained pWPMR2 further highlighting the role of this factor in driving outbreaks of highly drug-resistant bacteria (Extended data).

Conclusions

Here we have identified that auxiliary *metG* carried by the phage plasmid pWPMR2 is associated with evolution of bacterial lineages on an AMR trajectory (in this case *Shigella* sp. circulating in STN). Laboratory studies demonstrated that expression of auxiliary *metG* created cell subpopulations capable of surviving antimicrobial insult and predisposed to the evolution of AMR phenotypes (albeit at a significant fitness cost). It may be that these features

explain the nature of the *metG* carriage in bacterial populations. Specifically, that the pWPMR2 backbone on which *metG* is borne is a low, and even sometimes partial copy number (i.e. an infecting pool/colony would have a measure of 0 – 1 copy number per cell). This would act as an evolutionary hedge on behalf of the bacteria, in that many cells would avoid the significant fitness cost of *metG* expression, but that a small population would be able to survive antimicrobial treatment and regrow. Furthermore, the fact that this element lies on a phage-plasmid means that it may not need its bacterial host to survive antimicrobial insult to be passed on with the phage simply reinfecting a new host when able. To explore these possibilities, further work to explore the transfer dynamics of pWPMR2 and the transcriptional and translational regulation of mobile genetic element-borne *metG* is needed.

Our genomic epidemiology work then highlighted that mobile genetic element (MGE) borne *metG* and its commensurate impact on survival in the face of antimicrobials is not restricted to *S. sonnei*. Our work showed that pWPMR2 circulates among both clinically relevant bacteria and has been isolated from a commensal organism, and that other potentially mobilisable *metG* bearing mobile genetic elements exist. This is a pattern entirely consistent with other MGEs that bear advantageous phenotypes in the antibiotic era (e.g. most specifically AMR plasmids) and has been seen in sexually transmissible shigellosis for AZM and CRO resistance plasmids. Concerningly, and perhaps potentiating the potential service to AMR evolution, *metG* expression might induce a persister phenotype that could also contribute to chronic infections, which has been observed for some strains of sexually transmissible shigellosis (38,39). We also showed that pWPMR2 offers a *metG* copy that is not subject to evolutionary constraint, and some mutations have already occurred that ameliorate the fitness cost of the auxiliary *metG* expression. Hence, pWPMR2 and similar *metG*-bearing MGEs, may go on to become an increasingly common pathogen adaptation and contribute to growing other public health concerns by moving into other pathogens. For these reasons, copy number variations in *metG* or the detection of *metG*-bearing MGEs might act as an early warning sign that strains may be susceptible to becoming AMR.

Perhaps what is most concerning about pWPMR2 is we have (globally, as a public health community) been detecting evidence of this element since 2004. However, the significance remained undetected until now, and there are two important lessons from this for future microbiological surveillance in the genomic era. The first is that we have an overly narrow focus on certain phenotypes in facing the AMR crisis. While the scientific community has been talking of tolerance and persistence for many years, these phenotypes have never been actively screened for or formed part of surveillance practices (owing largely to a focus on individual level patient outcomes). Hence, this work highlights the need to think more broadly about what phenotypes drive lineages to epidemiological success and to do retrospective investigations such as this on the evolutionary events that led us to new disease phenomena. The second issue highlighted is an over-reliance on databases. Bacterial AMR is never going to stop evolving as we need to sustain life saving (though responsible) antimicrobial use. To this end, we are fortunate to now be in the genomic surveillance era featuring enhanced taxonomic nomenclature and well-maintained feature databases. However, if we are to keep up with our prokaryotic counterparts in this evolutionary arms race, we also need agile systems to detect, investigate, and respond to changes in pathogen genetic make-ups that we do not understand, and more effectively link pathogen evolution with public health outcomes.

Methods

Phylogenetic reconstruction

Short reads of 3745 *S. sonnei* isolates went through the following quality control procedures: raw read data were downloaded from the Sequence Read Archive (Bioproject: PRJNA248792, Supplementary Table). Reads were filtered using Trimmomatic v0.39, residual Illumina adapter sequences were removed, the first and last three base pairs in a read were trimmed, before a sliding window quality trimming of four bases with a quality threshold of 20 were applied, and reads of less than 36 bp (after trimming) were removed (ILLUMINACLIP:PE_All.fasta:2:30:10:2:keepBothReads SLIDINGWINDOW:4:20 LEADING:3 TRAILING:3 MINLEN:36) (40). The coverage of the resulting file was estimated against the size complete reference genome *S. sonnei* 53G (Accession: NC_016822.1) and down sampled to ~100 x coverage. Musket v1.1 (41) was applied for k-mer spectrum read correction using a k-mer size of 31bp. The processed reads were then mapped to *S. sonnei* reference genome 53G using snippy v4.6.0 (https://github. com/tseemann/snippy), resulting in a core genome alignment of 4,988,504bp in size. After filtering out polymorphic sites within recombined regions using Gubbins (42) with marginal ancestral state reconstruction implemented, 32,294 polymorphic sites unaffected by recombination were kept for the final phylogenetic tree reconstruction. A maximum likelihood phylogenetic tree reconstructed was performed using iqtree v2.3.6 with a General Time Reversible (GTR) substitution model with 4 gamma rate categories and 10,000 ultrafast bootstraps to provide approximately unbiased branch support values (43). A second phylogenetic reconstruction of the evolutionary relationships among 172 representative isolates from the 3745 *S. sonnei* isolates and 123 isolates from MSM-associated blaCTX-M-15 outbreak (37) short reads of 295 isolates went through quality control procedures as described in *Determining depth of coverage and variant calling* below. Then, a phylogenetic tree inferred similarly to above from 5,942bp polymorphic sites unaffected by recombination.

Identification of sexual transmission network lineages

Out of 3745 *S. sonnei* isolates, 1383 isolates with demographic data combination of male, adult and without travel history to high-risk countries were given presumptive MSM status (pMSM). Ancestral state reconstruction of each internal node in the phylogenetic tree based on the pMSM status of each isolate (tree tips) was performed using MrBayes (44) through R implementation, provided by the package MBASR (45). The pMSM status of each isolate was refined by the reconstructed ancestral status of their most adjacent nodes, as STN lineages are typically identified by an over-representation of isolates derived from pMSM (see references below). Through plotting the frequency distribution histogram of pMSM probabilities of the most adjacent node for each isolate, pMSM probabilities greater than 0.8 were defined as pMSM (Extended data). For defining STN lineages, all isolates descending from each internal node were defined as belonging to a lineage with the node being the most recent common ancestor. Then, iterating through each node in the tree, a STN lineage was defined when more than 75% of the isolates in the lineage were pMSM. This approach is an extension of, and in line with, previous work and validated surveillance practices (12). To avoid complete reconstruction of the tree as pMSM, and reflecting the recent emergence of sexually transmissible shigellosis, only the first and second nodes that were most adjacent to the tree tips were used in STN lineages definition.

Bacterial genome wide association analysis

Unitig-based bGWAS was run on 3,745 *S. sonnei* isolates with the 1,028 isolates belonging to MSMA lineages being a binary phenotype. Unitigs counting was performed using unitig-caller (https://github.com/bacpop/unitig-caller). The presence-absence matrix of the unitigs was associated with the binary phenotype belonging to MSMA lineages through computing terminal association scores in treeWAS v1.1 (20). Unitigs with the highest absolute association scores showed presence-absence patterns that were the most significantly associated with the phenotype. A minor allele frequency filter of 0.05 was applied and associated unitigs with association score of greater than 0.66 were examined by the investigative team for further follow up.

Determining depth of coverage and variant calling

Illumina reads for 3745 *S. sonnei* isolates were trimmed using fastp v0.23.4 (46), using the following options, in the order as appear: auto-detection for adapter for pair-end data (--detect_adapter_for_pe), move a sliding window of size 4bp from front (5') to tail, drop the bases in the window if its mean quality is less than 20, stop otherwise (--cut_front), move a sliding window of size 4bp from tail (3') to front, drop the bases in the window if its mean quality is less than 20, stop otherwise (--cut_tail), global trimming at front and tail for 15 bp (--trim_front1 15, --trim_front2 15), detect and trim polyG in read tails in minimum length of 10 bases (--trim_poly_g), detect and trim polyX in read tails in minimum length of 10 bases (--trim_poly_x).

The relative depth of pWPMR2 coverage was determined by mapping trimmed short reads using Bowtie2 (47) to a concatenated reference that contained the following components: SRR5005407 chromosome (Nanopore contig 1, 4 819 329bp), pSRK100 (contig 8, 65 728bp) and P-P (contig 9, 108 107bp), separated by strings of 500xN. Final destination of any read with multiple equally good alignments was randomly reported as default setting. The relative depth of coverage for each isolate was obtained from the average read depth for the pWPMR2, normalised by that of the chromosome using Samtools v 1.19 (48). PCR duplicates were marked using Samtools v 1.19 (48).

For copy number determination and variant calling in *metG,* trimmed short reads were mapped using GEM3 (49) to SRR5005407 chromosome (contig 1) only, followed by marking PCR duplicates. For *metG* variant calling, the number of reads supporting each allele in each variant position in the *metG* gene region were extracted using bcftools v1.19 mpileup command. They were then converted into .bcf file using bcftools call command, with the following options applied: output all alternate alleles in the alignments regardless of if they appear in any of the final genotype (-A), haploid (--ploidy 1), and alternative model for multiallelic and rare-variant calling (-m). For each variant position in each sample, the presence of an alternative allele was only defined when it was supported by at least 40 reads. Alternative allele proportion was calculated by the number of reads supporting the alternative allele divided by the total read depth of the position in a sample. Copy number of *metG* was obtained based on the same mapping, by computing its average read depth for each isolate, normalised by that of chromosome using Samtools v1.19. PCR duplicates were marked using Samtools (48).

Genetic analyses of pWPMR2

The circular topology of pWPMR2 (Nanopore contig 10) and pWMPR2var (Nanopore contig 8) was confirmed by assembling the Nanopore reads of SRR5005407 and SRR13013630 respectively using Flye v2.9.5-b1801 (50). Assembled contigs were visualised using Bandage v0.9.0. (51). Prior assembly, adapter sequences were trimmed and reads with internal adapter were discarded using Porechop v0.2.4 (https://github.com/ rrwick/Porechop). Quality score and length of remaining reads were visualised using NanoPlot v1.43.0 (52), then reads of less than 150bp and with quality score less than 10 were filtered out using NanoFilt v2.8.0 (53).

Detection of phage sequences, plasmid specific sequences and AMR genes were performed by running pWPMR2 through PHASTEST v3.0 (54), PlasmidFinder v2.1 (55,56) and AMRfinderPlus v3.12.8 (57). We determined the taxonomy of pWPMR2 using the ViPTree server (58) to generate an all-against-all tBLASTx similarity proteomic tree against 1,512 related dsDNA prokaryotic viruses from the Virus-Host DB, which utilises complete genomes from NCBI RefSeq and GenBank (59), also including examples from each of the five defined phage-plasmid groups within the SSU5 community (32). To further explore its association with clinically relevant organisms we downloaded the genomes of its closest relatives, as found through a BLASTn search against the NCBI nucleotide collection (nt) database; selecting sequences with >95% coverage and >98% identity for direct comparison.

Five *E. albertii* plasmids were chosen for comparison with pWPMR2 as they represented the top non-*Shigella* and non-*E. coli* hits in the online nBLAST database showing highest percentage identity. To screen 66,929 genomes of enteric bacteria available from clinical cases in the UK for evidence of SSU5-like, pSLy3-like and pWPMR2-like phage plasmids and additional *metG*, a mapping approach as described previously (60) was used. Briefly, target genes including *metG* (CP104415.1:54597..56630), *repB* (CP104415.1:22858..24072), and *repA* (NC_018843.1:59428..60492) were screened in clinical enteric bacteria derived from UK reference laboratory routine genomic surveillance, available under BioProjects PRJNA315192 (*E. coli* and *Shigella* spp.) and PRJNA248792 (*Salmonella* spp.) (58,60).

Structural variants between STN-lineage associated pWPMR2 and pWPMR2var, comparison of near-identical relatives of pWPMR2 in public databases, pWPMR2 and *E. albertii* plasmid comparison as well as was visualised using EasyFig genome comparison visualiser (61) or Artemis and ACT (62). The pWPMR2 sequence was additionally screened against ~2M bacterial genomes in the AllTheBacteria database v0.3 (63) using LexicMap tool v0.4 (64), with a 70% coverage and 70% identity threshold. Metadata of associated hits were then retrieved using Entrez Direct (https://www.ncbi.nlm.nih.gov/books/NBK25501/).

Long read sequencing of selected isolates

Isolates were grown in TSB overnight with 200 rpm shaking at 37˚C and 500µl of the overnight culture was then used in DNA extraction using either MasterPure total DNA/RNA extraction kit (Biosearch Technologies) or FireMonkey DNA extraction kit (Revolugen) following manufacturers' recommendations. Extracted DNA was then quantified using a Qubit before sending off to be sequenced, analysed and annotated by commercially available Plasmidsaurus service (https://www.plasmidsaurus.com/) using Oxford Nanopore Technologies (ONT) platform.

Construction of E. coli strains

metG$_{WT}$ and *metG*$_{PP}$ sequences from *S. sonnei* obtained from SRR5005407 were used to synthesise the respective genes using GeneArt custom synthesis platform (ThermoScientific). The synthesised genes were then cloned into the inducible pTrcHis2A vector using the same service and the two resulting plasmids were then transformed into *E. coli* NEB5α (New England Biolabs, UK) and selected using the ampicillin marker on the vector. Transformed *E. coli* strains were used in the downstream experiments with 1mM IPTG (Millipore Sigma, UK) as the inducer to induce the expression of the *metG* genes.

Sub-MIC growth curves

Minimum Inhibitory Concentrations (MIC) for each of the *E. coli* strains carrying the two different versions of *metG* genes on pTrcHis2A vector were determined using E-strips (Liofilchem, Italy) or broth dilution method for azithromycin (0.25µg/ml), ciprofloxacin (0.008 µg/ml) and ceftriaxone (0.008µg/ml). Those strains were then grown in biological and technical triplicates in TSB containing half of the MIC concentrations of each of the three antibiotics with and without induction of the *metG* variants. The OD_{600nm} values were recorded every 15 minutes using a BioTek Synergy H1 multi-mode plate reader. The resulting data were then analysed using gcplyr R package (65) and graphs were generated using ggplot2 R package (66).

Experimental evolution study

E. coli NEB5α containing *metG*$_{WT}$ and *metG*$_{PP}$ on a pTrcHis2a plasmid backbone were streaked out on TSA plates containing 100µg/ml ampicillin and incubated overnight at 37˚C to obtain isolated colonies of each of the *metG* variant carrying *E. coli*. These isolated colonies were then inoculated into individual wells of 96 well plates containing either no antibiotics or AZM, CIP or CRO at a concentration of ½ MIC value for each and treated as an individual population. Each 96 well plate contained 47 wells with TSB with IPTG at a final concentration of 1mM and a control well, 47 wells with no IPTG and a control well. These 96 well plates were then incubated at 37˚C with shaking at 200 rpm for 24 hours before transferring 1% (v/v) into a fresh 96 well plate containing twice the concentration of each of the antibiotics from the overnight plate. This was carried on for 12 days with the final concentration of each antibiotic reaching 1024x MIC. At the end of each of the 24-hour incubation periods, each of the wells were scored for growth and the number of populations survived was plotted against the antibiotic concentration using the ggplot2 R package. Statistical analysis for the survival of each of the populations was performed using the survival R package (https://CRAN.R-project.org/package=survival).

Analysis of plasmid databases

PLSDB database version 2023_11_23 was downloaded from https://ccb-microbe.cs.uni-saarland.de/plsdb/plasmids/download/ and annotated using Bakta version 1.8.2 (67) with the 'full' database. A list of COG categories for tRNA synthetases was downloaded from the COG database and used as search terms in the Bakta output .gff files. To avoid potentially misclassified chromosomal sequences in PLSDB, sequences encoding ≥3 tRNA synthetases were removed from analyses of tRNA synthetase-encoding plasmids. Plasmid types were extracted from PLSDB metadata. Analyses were conducted in R using tidyverse (68).

Acknowledgements
This work was funded by BBSRC (BB/V009184/1) and MRC (MR/X000648/1) project grants. JPJH is supported by an MRC Career Development Award (MR/W02666X/1). KSB, CJ, and LCEM are affiliated to the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Gastrointestinal Infections at University of Liverpool in partnership with the United Kingdom Health Security Agency, in collaboration with University of Warwick. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health and Social Care or UKHSA. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission. SN is affiliated with the HPRU GED in partnership with the UK Health Security Agency.

Author contributions

Conceptualization – KSB, JPJH
Data curation – KSB, LC
Formal analysis – CRB, YLT, JPJH, LS, GBB, MDS, CEC
Funding acquisition – KSB, JPJH
Investigation – KSB, LC, YLT, JPJH, LS, SN, MDS, CEC
Methodology – KSB, LC, YLT
Project administration – KSB
Resources – KSB
Supervision – KSB, LC, JPJH
Visualisation – KSB, CRB, YLT, GBB, MDS, CEC
Writing – original draft – KSB, YLT, JPJH, GBB, MDS, CEC
Writing – review & editing – KSB, LCEM, LC, CRB, YLT, JPJH, SN, MDS, CEC

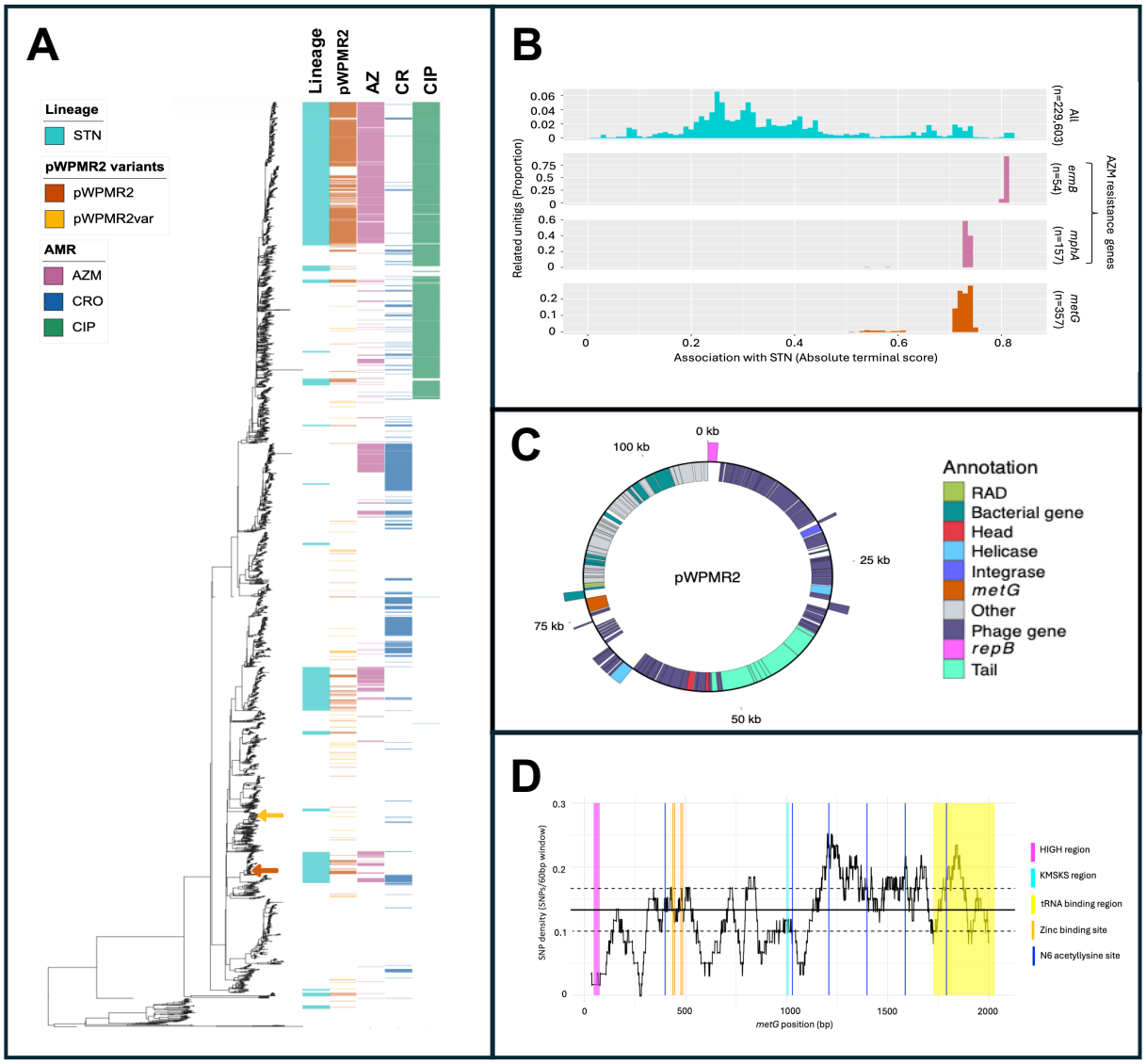

Figure 1: Variable metG borne on pWPMR2 is associated with bacterial lineages on an AMR trajectory. A. A midpoint-rooted maximum likelihood phylogenetic tree constructed from 32,294 recombination-free polymorphic sites shows the evolutionary relationships among 3,745 routine genomic surveillance *S. sonnei* isolates. The adjacent metadata columns show (from left to right): STN lineages (i.e. circulating in sexual transmission networks); the presence of pWPMR2 or its variant pWPMR2var; and the presence of genotypically predicted resistance against three key antimicrobial treatments; azithromycin (AZM), ceftriaxone (CRO), and ciprofloxacin (CIP). The two isolates used to nanopore sequence pWPMR2 and pWPMR2var are indicated by arrows. **B.** Frequency histograms show the distribution of Genome Wide Association Scores. Distributions are shown for all unitigs ≥0.05 minor allele frequency (n=229,560) among the isolates, and unitigs relating to the AZM resistance genes *mphA* and *ermB* (acting here as positive controls) and *metG*. **C.** A genetic map of the phage-plasmid pWPMR2 forward and reverse open reading frames are shown as outer and inner colour blocks which are coloured according to the inlaid key. **D.** The variation in *metG* across the length of the gene among some 650 pWPMR2/pWPMR2var-containing isolates. Median (solid) and interquartile-range quantiles (dashed) of SNP-density across the gene are indicated by horizontal lines and a sliding 60bp window of SNP-density reveals higher variation in the C-terminal, including the tRNA binding region, compared with the conserved amino acid motifs (HIGH and KMSKS).

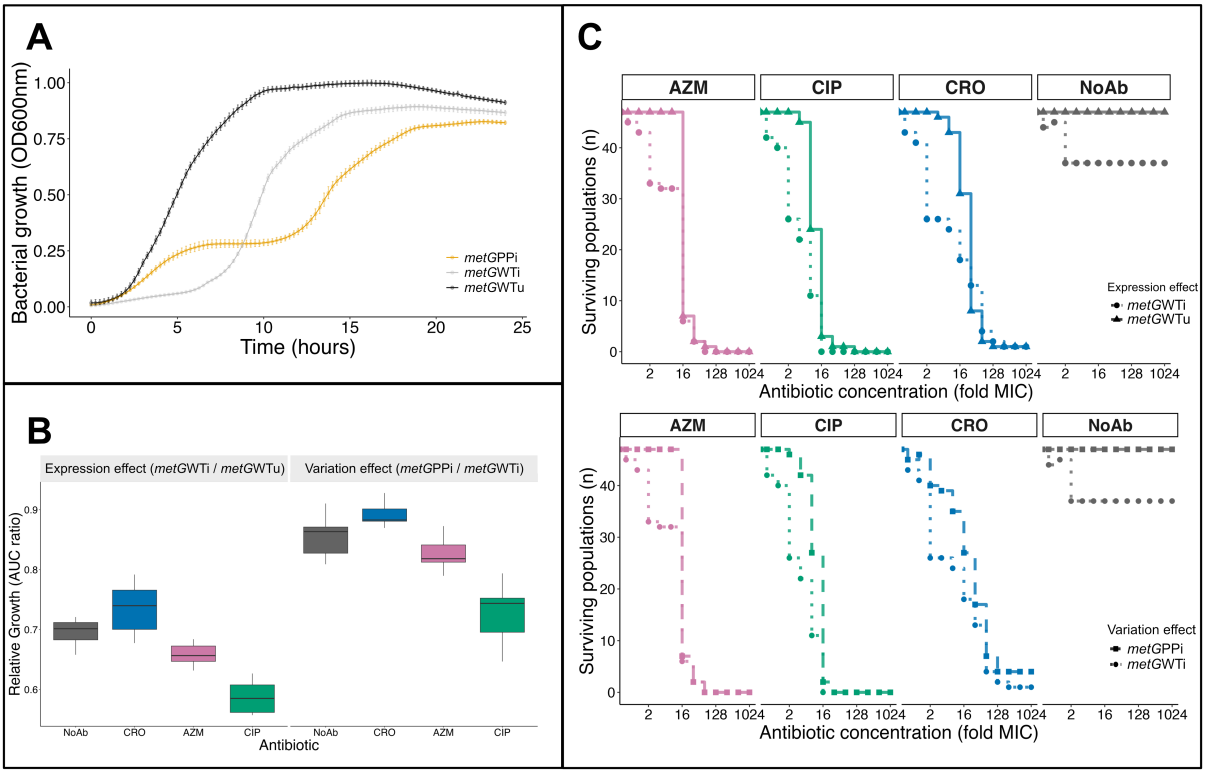

Figure 2: Auxiliary *metG* expression and variation generates altered growth profiles and antimicrobial response phenotypes. A. Growth over time for *E. coli* DH5α constructs with auxiliary $metG_{WT}$ expression induced (*metG*WTi) and the respective (i.e. not induced) control (metGWTu) and the pWPMR2 version of *metG* ($metG_{PP}$) induced (*metG*PPi). **B.** Relative growth of constructs in the presence of subinhibitory concentrations of key treatment antimicrobials (0.5x MIC) shown for auxiliary *metG* expression (left, *metG*WTi relative to *metG*WTu) and variation (right, *metG*PPi relative to *metG*WTi). **C.** Population survival curves of *E. coli* constructs induced to express auxiliary $metG_{WT}$ (*metG*WTi) and uninduced (*metG*WTu) [upper] and *metG*WTi relative to the pWPMR2 version (*metG*PPi) [lower]. (CRO = ceftriaxone, AZM = azithromycin, CIP = ciprofloxacin AUC = area under the curve).

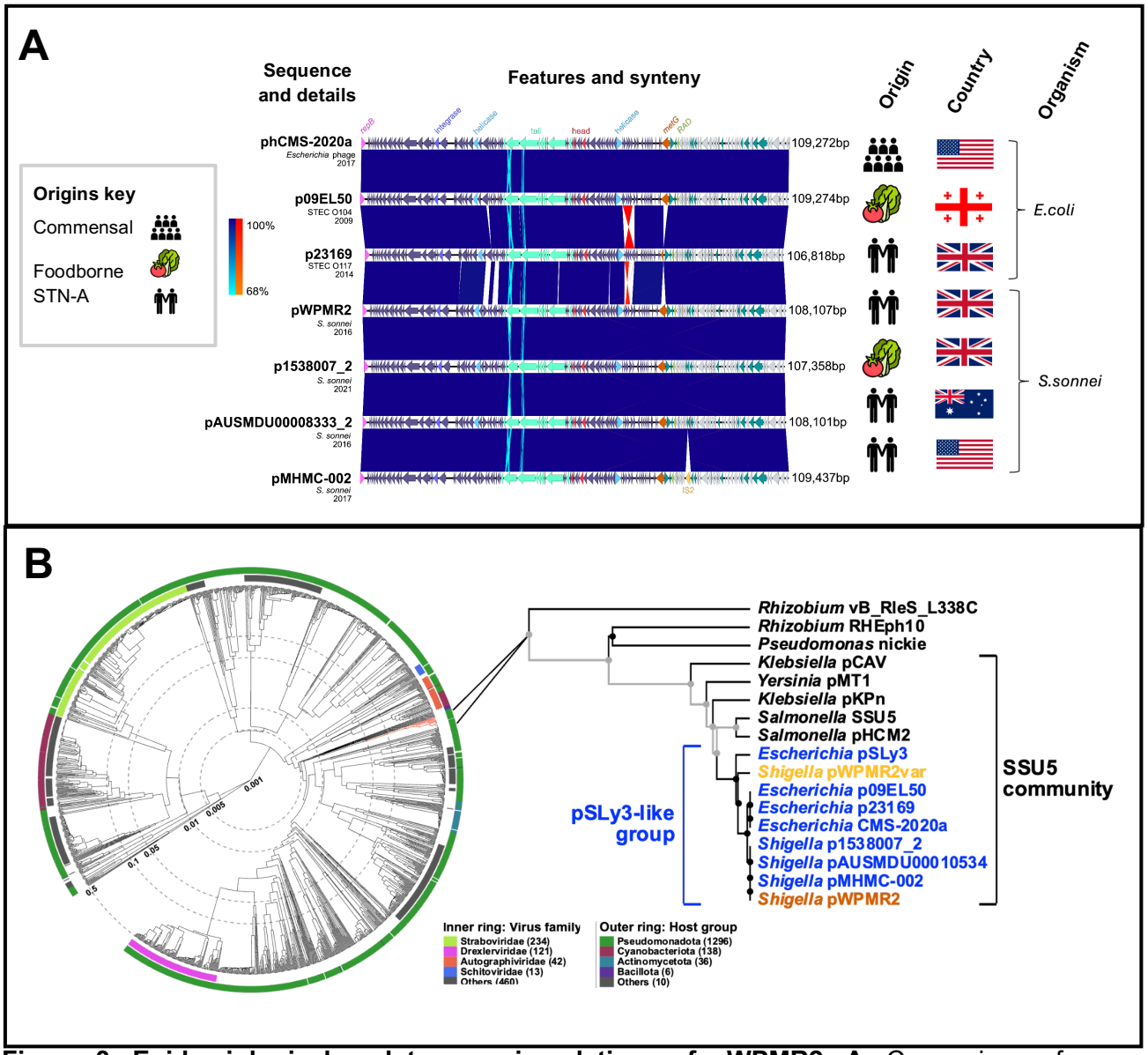

Figure 3. Epidemiological and taxonomic relatives of pWPMR2. A. Comparison of pWPMR2 with its most similar publicly available sequences (labelled leftmost) and their epidemiological details. The sequence comparison is shown as linearised replicons starting with *repB* gene. Certain relevant gene features are shown in colour and labelled uppermost. Colour blocks intervening between sequences showing colinear (blue scale) and inverted (red scale) synteny according to the inlaid coloured gradient keys. The origin of the sequence, country and organism are shown to the right of sequences. The symbol key for origin is shown leftmost. Other available details for the sequences (e.g. subtype, year of isolation) are shown under the sequence label. STEC = Shiga toxigenic *Escherichia coli*. STN-A = Sexual Transmission Network associated **B.** The circular cladogram shows comparison among 1512 related reference viral sequences (see methods). The branch lengths show genetic relatedness based on tBLASTx comparison (embedded in VIP tree) on a non-linear scale. The inner and outer rings show virus family and host group according to the inlaid keys, and the SSU5 community is highlighted as a red branch. The pop out shows a SSU5 community cladogram, highlighting the pSLy3 group in blue, and pWPMR2 and its variant in orange and yellow, respectively.

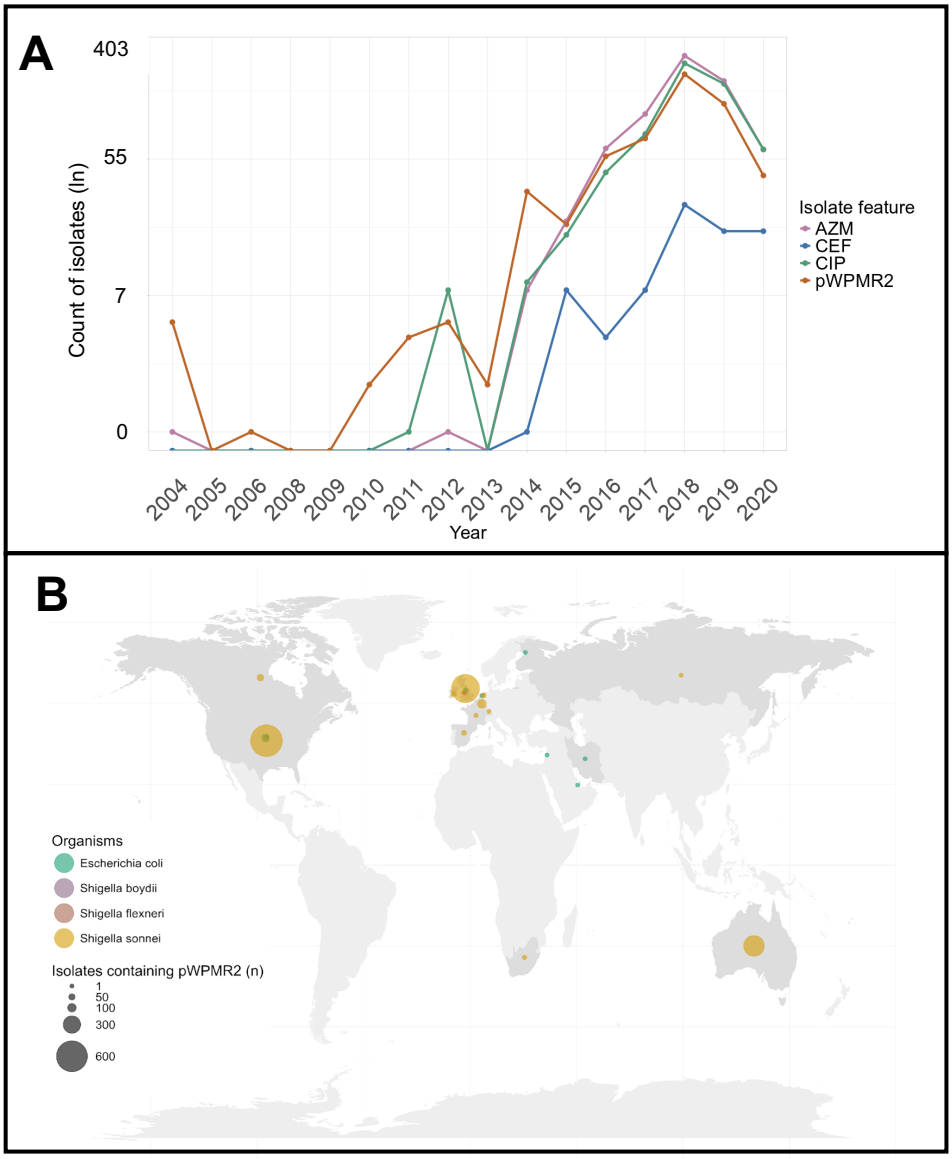

Figure 4. Temporal and spatial spread of pWPMR2. A. The exponential growth in frequency of isolates containing pWPMR2, and resistance markers against three key antimicrobials among 3,745 *S. sonnei* surveillance isolates is shown over time according to the inlaid key, highlighting the early ingress of pWPMR2 into the *S. sonnei* population. **B.** The global distribution of 2,028 pWPMR2-containing isolates in public databases is shown with countries containing sequences highlighted in darker grey. Bubbles over country centroids are shown scaled by the number of isolates and coloured by the organism of origin according to the inlaid keys

References

1. Pulford CV, Perez-Sepulveda BM, Canals R, Bevington JA, Bengtsson RJ, Wenner N, et al. Stepwise evolution of Salmonella Typhimurium ST313 causing bloodstream infection in Africa. Nat Microbiol. 2021 Mar;6(3):327–38.

2. Johnson JR, Johnston B, Clabots C, Kuskowski MA, Castanheira M. Escherichia coli sequence type ST131 as the major cause of serious multidrug-resistant E. coli infections in the United States. Clin Infect Dis. 2010 Aug 1;51(3):286–94.

3. Banerjee R, Johnson JR. A new clone sweeps clean: The enigmatic emergence of Escherichia coli sequence type 131. Antimicrob Agents Chemother. 2014 Sep;58(9):4997–5004.

4. Edquist PJ, Makitalo B, Olsson-Liljequist B, Soderblom T, Wisell KT. Epidemiology extended-spectrum beta-lactamase-producing Escherichia coli Sweden 2007–2011. Clin Microbiol Infect. 2007;20(6):O344–52.

5. Whitmer GR, Moorthy G, Arshad M. The pandemic Escherichia coli sequence type 131 strain is acquired even in the absence of antibiotic exposure. PLoS Pathog. 2019 Dec;15(12):e1008162.

6. García-Fernández A, Villa L, Carta C, Venditti C, Giordano A, Venditti M, et al. Klebsiella pneumoniae ST258 producing KPC-3 identified in italy carries novel plasmids and OmpK36/OmpK35 porin variants. Antimicrob Agents Chemother. 2012 Apr;56(4):2143–5.

7. Thompson CN, Duy PT, Baker S. The rising dominance of Shigella sonnei: An intercontinental shift in the etiology of bacillary dysentery. PLoS Negl Trop Dis. 2015 Jun 11;9(6):e0003708.

8. Bennett, R J ,De Silva, P. M. , Bengtsson, R. J. , Horsburgh, M. J. , Blower, T. R. and Baker, K. S. Temporal GWAS identifies widely distributed putative adhesin contributing pathogen success Shigella spp. bioRxivorg. 2022;

9. Mason LCE, Greig DR, Cowley LA, Partridge SR, Martinez E, Blackwell GA, et al. The evolution and international spread of extensively drug resistant Shigella sonnei. Nat Commun. 2023 Apr 8;14(1):1983.

10. Baker KS, Dallman TJ, Ashton PM, Day M, Hughes G, Crook PD, et al. Intercontinental dissemination of azithromycin-resistant shigellosis through sexual transmission: a cross-sectional study. Lancet Infect Dis. 2015 Aug;15(8):913–21.

11. Mitchell HD, Thomson NR, Jenkins C, Dallman TJ, Painset A, Kirwan P, et al. Linkage of whole genome sequencing, epidemiological, and clinical data to understand the genetic diversity and clinical outcomes of Shigella flexneri among men who have sex with men in England. Microbiol Spectr. 2021 Dec 22;9(3):e0121321.

12. Mitchell HD, Mikhail AFW, Painset A, Dallman TJ, Jenkins C, Thomson NR, et al. Use of whole-genome sequencing to identify clusters of Shigella flexneri associated with sexual transmission in men who have sex with men in England: a validation study using linked behavioural data. Microb Genom [Internet]. 2019 Nov;5(11). Available from: http://dx.doi.org/10.1099/mgen.0.000311

13. Lefèvre S, Njamkepo E, Feldman S, Ruckly C, Carle I, Lejay-Collin M, et al. Rapid emergence of extensively drug-resistant Shigella sonnei in France. Nat Commun. 2023 Jan 28;14(1):462.

14. Baker KS, Dallman TJ, Field N, Childs T, Mitchell H, Day M, et al. Horizontal antimicrobial resistance transfer drives epidemics of multiple Shigella species. Nat Commun. 2018 Apr 13;9(1):1462.

15. Gilbart VL, Simms I, Jenkins C, Furegato M, Gobin M, Oliver I, et al. Sex, drugs and smart phone applications: findings from semistructured interviews with men who have sex with men diagnosed with Shigella flexneri 3a in England and Wales. Sex Transm Infect. 2015 Dec;91(8):598–602.

16. Ingle DJ, Easton M, Valcanis M, Seemann T, Kwong JC, Stephens N, et al. Co-circulation of multidrug-resistant Shigella among men who have sex with men in Australia. Clin Infect Dis. 2019 Oct 15;69(9):1535–44.

17. Worley JN, Javkar K, Hoffmann M, Hysell K, Garcia-Williams A, Tagg K, et al. Genomic drivers of multidrug-resistant Shigella affecting vulnerable patient populations in the United States and abroad. MBio [Internet]. 2021 Jan 26;12(1). Available from: http://dx.doi.org/10.1128/mBio.03188-20

18. Bardsley M, Jenkins C, Mitchell HD, Mikhail AFW, Baker KS, Foster K, et al. Persistent transmission of shigellosis in England is associated with a recently emerged multidrug-resistant strain of Shigella sonnei. J Clin Microbiol [Internet]. 2020 Mar 25;58(4). Available from: http://dx.doi.org/10.1128/JCM.01692-19

19. Hawkey J, Paranagama K, Baker KS, Bengtsson RJ, Weill F-X, Thomson NR, et al. Global population structure and genotyping framework for genomic surveillance of the major dysentery pathogen, Shigella sonnei. Nat Commun. 2021 May 11;12(1):2684.

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

21. Jarvest RL, Berge JM, Berry V, Boyd HF, Brown MJ, Elder JS, et al. Nanomolar inhibitors of Staphylococcus aureus methionyl tRNA synthetase with potent antibacterial activity against gram-positive pathogens. J Med Chem. 2002 May 9;45(10):1959–62.

22. Jarvest RL, Berge JM, Brown P, Houge-Frydrych CSV, O'Hanlon PJ, McNair DJ, et al. Conformational restriction of methionyl tRNA synthetase inhibitors leading to analogues with potent inhibition and excellent gram-positive antibacterial activity. Bioorg Med Chem Lett. 2003 Apr 7;13(7):1265–8.

23. Girgis HS, Harris K, Tavazoie S. Large mutational target size for rapid emergence of bacterial persistence. Proc Natl Acad Sci U S A. 2012 Jul 31;109(31):12740–5.

24. Khare A, Tavazoie S. Extreme antibiotic persistence via heterogeneity-generating mutations targeting translation. mSystems [Internet]. 2020 Jan 21;5(1). Available from: http://dx.doi.org/10.1128/mSystems.00847-19

25. Levin-Reisman I, Ronin I, Gefen O, Braniss I, Shoresh N, Balaban NQ. Antibiotic tolerance facilitates the evolution of resistance. Science. 2017 Feb 24;355(6327):826–30.

26. Wein T, Wang Y, Barz M, Stücker FT, Hammerschmidt K, Dagan T. Essential gene acquisition destabilizes plasmid inheritance. PLoS Genet. 2021 Jul;17(7):e1009656.

27. Azam, A.H., Chihara, K., Kondo, K., Nakamura, T., Ojima, S., Tamura, A., Yamashita, W., Cui, L., Takahashi, Y., Watashi, K., & Kiga, K. Viruses encode tRNA and anti-retron to evade bacterial immunity. bioRxivorg. 2023;

28. Levin-Reisman I, Brauner A, Ronin I, Balaban NQ. Epistasis between antibiotic tolerance, persistence, and resistance mutations. Proc Natl Acad Sci U S A. 2019 Jul 16;116(29):14734–9.

29. Baker KS, Dallman TJ, Thomson NR, Jenkins C. An outbreak of a rare Shiga-toxin-producing Escherichia coli serotype (O117:H7) among men who have sex with men. Microb Genom [Internet]. 2018 Jul;4(7). Available from: http://dx.doi.org/10.1099/mgen.0.000181

30. Ahmed SA, Awosika J, Baldwin C, Bishop-Lilly KA, Biswas B, Broomall S, et al. Genomic comparison of Escherichia coli O104:H4 isolates from 2009 and 2011 reveals plasmid, and prophage heterogeneity, including shiga toxin encoding phage stx2. PLoS One. 2012 Nov 1;7(11):e48228.

31. Stephens C, Arismendi T, Wright M, Hartman A, Gonzalez A, Gill M, et al. F plasmids are the major carriers of antibiotic resistance genes in human-associated commensal Escherichia coli. mSphere [Internet]. 2020 Aug 5;5(4). Available from: http://dx.doi.org/10.1128/mSphere.00709-20

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

33. Thorley K, Charles H, Greig DR, Prochazka M, Mason LCE, Baker KS, et al. Emergence of extensively drug-resistant and multidrug-resistant Shigella flexneri serotype 2a associated with sexual transmission among gay, bisexual, and other men who have sex with men, in England: a descriptive epidemiological study. Lancet Infect Dis. 2023 Jun;23(6):732–9.

34. Schmartz GP, Hartung A, Hirsch P, Kern F, Fehlmann T, Müller R, et al. PLSDB: advancing a comprehensive database of bacterial plasmids. Nucleic Acids Res. 2022 Jan 7;50(D1):D273–8.

35. diCenzo GC, Mengoni A, Perrin E. Chromids aid genome expansion and functional diversification in the family Burkholderiaceae. Mol Biol Evol. 2019 Mar 1;36(3):562–74.

36. Clough SE, Elphinstone JG, Friman V-P. Plant pathogenic bacterium Ralstonia solanacearum can rapidly evolve tolerance to antimicrobials produced by Pseudomonas biocontrol bacteria. J Evol Biol. 2024 Feb 14;37(2):225–37.

37. Charles H., Greig D.R., Swift C., Olonade I., Simms I., Sinka K., Baker K.S., Godbole G., Jenkins C. Outbreak of sexually transmitted *S. sonnei* $bla_{\text{CTX-M-15}}$ in England: an epidemiological and genomic investigation. medRxiv [Internet]. 2024; Available from: http://dx.doi.org/10.1101/2024.10.14.24314996

38. Allen H, Mitchell HD, Simms I, Baker KS, Foster K, Hughes G, et al. Evidence for re-infection and persistent carriage of Shigella species in adult males reporting domestically acquired infection in England. Clin Microbiol Infect. 2021 Jan;27(1):126.e7-126.e13.

39. Bengtsson RJ, Dallman TJ, Allen H, De Silva PM, Stenhouse G, Pulford CV, et al. Accessory genome dynamics and structural variation of Shigella from persistent

infections. MBio [Internet]. 2021 Apr 27;12(2). Available from: http://dx.doi.org/10.1128/mBio.00254-21

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

41. Liu Y, Schröder J, Schmidt B. Musket: a multistage k-mer spectrum-based error corrector for Illumina sequence data. Bioinformatics. 2013 Feb 1;29(3):308–15.

42. 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 Res. 2015 Feb 18;43(3):e15.

43. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020 May 1;37(5):1530–4.

44. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012 May;61(3):539–42.

45. Heritage S. MBASR: Workflow-simplified ancestral state reconstruction of discrete traits with MrBayes in the R environment [Internet]. bioRxiv. bioRxiv; 2021. Available from: http://dx.doi.org/10.1101/2021.01.10.426107

46. Chen S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. Imeta. 2023 May;2(2):e107.

47. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Mar 4;9(4):357–9.

48. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience [Internet]. 2021 Feb 16;10(2). Available from: http://dx.doi.org/10.1093/gigascience/giab008

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

50. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019 May;37(5):540–6.

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

52. De Coster W, Rademakers R. NanoPack2: population-scale evaluation of long-read sequencing data. Bioinformatics [Internet]. 2023 May 4;39(5). Available from: http://dx.doi.org/10.1093/bioinformatics/btad311

53. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics. 2018 Aug 1;34(15):2666–9.

54. Wishart DS, Han S, Saha S, Oler E, Peters H, Grant JR, et al. PHASTEST: faster than PHASTER, better than PHAST. Nucleic Acids Res. 2023 Jul 5;51(W1):W443–50.

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

56. Carattoli A, Hasman H. PlasmidFinder and in silico pMLST: Identification and typing of Plasmid replicons in whole-genome sequencing (WGS). Methods Mol Biol. 2020;2075:285–94.

57. Feldgarden M, Brover V, Gonzalez-Escalona N, Frye JG, Haendiges J, Haft DH, et al. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. Sci Rep. 2021 Jun 16;11(1):12728.

58. Nishimura Y, Yoshida T, Kuronishi M, Uehara H, Ogata H, Goto S. ViPTree: the viral proteomic tree server. Bioinformatics. 2017 Aug 1;33(15):2379–80.

59. Mihara T, Nishimura Y, Shimizu Y, Nishiyama H, Yoshikawa G, Uehara H, et al. Linking virus genomes with host taxonomy. Viruses. 2016 Mar 1;8(3):66.

60. Nair S, Barker CR, Bird M, Greig DR, Collins C, Painset A, et al. Presence of phage-plasmids in multiple serovars of Salmonella enterica. Microb Genom [Internet]. 2024 May;10(5). Available from: http://dx.doi.org/10.1099/mgen.0.001247

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

62. Carver TJ, Rutherford KM, Berriman M, Rajandream M-A, Barrell BG, Parkhill J. ACT: The Artemis Comparison Tool. Bioinformatics. 2005 Aug 15;21(16):3422–3.

63. Hunt M, Lima L, Shen W, Lees J, Iqbal Z. AllTheBacteria - all bacterial genomes assembled, available and searchable [Internet]. 2024. Available from: http://dx.doi.org/10.1101/2024.03.08.584059

64. Shen W, Iqbal Z. LexicMap: efficient sequence alignment against millions of prokaryotic genomes [Internet]. bioRxiv. 2024. Available from: http://dx.doi.org/10.1101/2024.08.30.610459

65. Blazanin M. gcplyr: an R package for microbial growth curve data analysis. BMC Bioinformatics. 2024 Jul 9;25(1):232.

66. Wickham H. Ggplot2. Wiley Interdiscip Rev Comput Stat. 2011 Mar;3(2):180–5.

67. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. Microb Genom [Internet]. 2021 Nov;7(11). Available from: http://dx.doi.org/10.1099/mgen.0.000685

68. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the tidyverse. J Open Source Softw. 2019 Nov 21;4(43):1686.