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

Yi Ling Tam^[1], P. Malaka De Silva^[2], Clare R. Barker[3,10], Ruizhe Li[4], Gherard Batisti Biffignandi[2,6], Leanne Santos[5], 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, UK
2. Department of Genetics, University of Cambridge, Cambridge, CB2 3EH, UK
3. Gastrointestinal Infections & Food Safety (One Health), UK Health Security Agency, UK
4. Department of Engineering, University of Cambridge, Trumpington St, Cambridge, CB2 1PZ, UK
5. Department of Evolution, Ecology and Behaviour, Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, UK
6. European Bioinformatics Institute, European Molecular Biology Laboratory, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK
7. Department of Clinical Infection, Microbiology, and Immunology, Institute for Infection, Veterinary and Ecological Sciences, Liverpool, UK
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 (226 words)

Antimicrobial resistance (AMR) is a global public health crisis with few options for control. As such early identification of emerging bacterial strains capable of rapidly evolving AMR is key. Although antimicrobial tolerance and persistence are precursor phenotypes for AMR, little evidence exists to support their importance in real-world settings. Here we used bacterial genome wide association on 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). This revealed an association of an AMR trajectory with a multi-and highly variable second copy of *metG*, borne by a phage-plasmid we called pWPMR2. Further analyses revealed that pWPMR2 was present across clinically relevant enteric pathogens globally, including past and contemporary outbreaks, and that the additional-*metG* mechanism was present across multiple bacterial phyla. Functional microbiology, experimental evolution, and single-cell physiology studies confirmed that the expression of auxiliary *metG*, particularly the mutated version on pWPMR2, created a sub population of persister cells predisposed to survival in, and evolving resistance against, third generation cephalosporins. Thus, we demonstrate a novel mechanism of auxiliary *metG* carriage that predisposes bacteria to AMR with real world impacts. Furthermore, our approach is a timely example of using genomic epidemiology to rapidly guide functional microbiology studies in the era of routine genomic surveillance, and also highlights several deficiencies in current AMR surveillance practices.

Background (4091 words in main text)

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* ST131 [2–5], the carbapenemase associated *K. pneumoniae* ST258 [6], and Lineage III *S. sonnei* which has recently evolved extensive drug resistance (XDR) [7–9]. Once AMR has been acquired and disseminated in these successful clones, the resulting treatment complications make their control 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 an understanding of the precursors or 'stepping stones' that occur before AMR evolves, or is acquired.

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 established over the last two 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* and that these are subject to high levels of antimicrobial selection pressure [9,13,14]. 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 adaptations potentially associated with a trajectory to AMR that could rapidly spread.

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) to identify 'stepping stones' to AMR. We identified variant copies of methionine tRNA ligase (metG) carried on an SSU5-like phage-plasmid, which we called pWPMR2, as associated with STN Lineages (a proxy for high antimicrobial selection). Further *in silico* analyses showed that both pWPMR2 and other plasmid-borne *metG* are spread globally in other enteric bacteria and bacterial phyla. We demonstrate how auxiliary *metG* expression and variation create bacteria with persister phenotypes that predispose bacteria to evolve AMR against third generation cephalosporins.

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=3,745) 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). We 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* [17]; ciprofloxacin (CIP), azithromycin (AZM), and ceftriaxone (CRO) (74% [n=765/1,028] compared with 37% [n=1011/2,717], 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 *Shigella* circulating in STN are convergently evolving toward AMR, consistent with previous studies [10,14,16,18–20].

We then quantified the association of unique genetic sequences (i.e. unitigs from a compiled assembly graph of the 3,745 isolates) with membership of STN Lineages using treeWAS [21]. 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 [10,14,16,18–20] (Figure 1A). 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 [22–25], and predisposition to AMR [26]). Thus, we hypothesised that the auxiliary metG was playing a role in the adaptation of *Shigella* lineages to circulation in the high antimicrobial selection environment of STNs.

Bioinformatic analyses to explore the nature of the *metG* association revealed both copy number variation and Single Nucleotide Polymorphisms (SNPs, 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 copy number (relative depth of pWPMR2 to chromosomal mapping coverage) and coverage (proportion of pWMPR2 sites covered) of pWPMR2 among the *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 (Extended data), suggesting some instability in pWPMR2 carriage, consistent with the acquisition of an essential gene [27]. Mapping coverage further suggested the presence of a *metG*-bearing pWPMR2 variant (pWPMR2var) with reduced (~0.75) coverage (Figure 1A, Table 1, Extended data, GenBank accession: PQ521219 from isolate SRR10313630). Consistent with it being a phage-plasmid, pWPMR2 contained an anti-defence system, RAD, previously described as targeting retrons [28]. No known AMR or virulence genes were present on pWPMR2, so we continued with our GWAS-guided hypothesis that auxiliary *metG* expression was driving a phenotypic shift relevant for adaptation to antimicrobial selection pressure.

STN Lineage*	Isolates, n	pWPMR2-bearing isolates, n (%)	Dominant genotype^ (% belonging to 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 Lineages	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 Data for individual assignments
^ genotyping performed according to [20]
[&] 53 of which were pWPMR2var

An analysis of SNPs across *metG* among the 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 conservation, with only 12 of 3095 isolates containing SNPs across five sites in *metG*; four of which introduced non-synonymous changes (found in non-STN Lineages not phylogenetically 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, 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 relative sequence conservation in certain key functional areas of the gene (such as the HIGH and KMSK amino acid motifs) and comparatively higher variation in the N6-acetyllysine sites (Figure 1D), which might alter *metG* post-translational modification and abundance [29,30], with a knock-on effect on overall protein production. While quantitating gene variation in multicopy genes using short read data is complex our complete genome sequencing confirmed that, the *metG* borne by pWPMR2 in our reference isolate SRR5005407 (see $metG_{PP}$ below) contained 115 SNPs, 6 of which were non-synonymous. This highlights that the additional metG on pWPMR2 experiences significant mutation relative to single chromosomal copies.

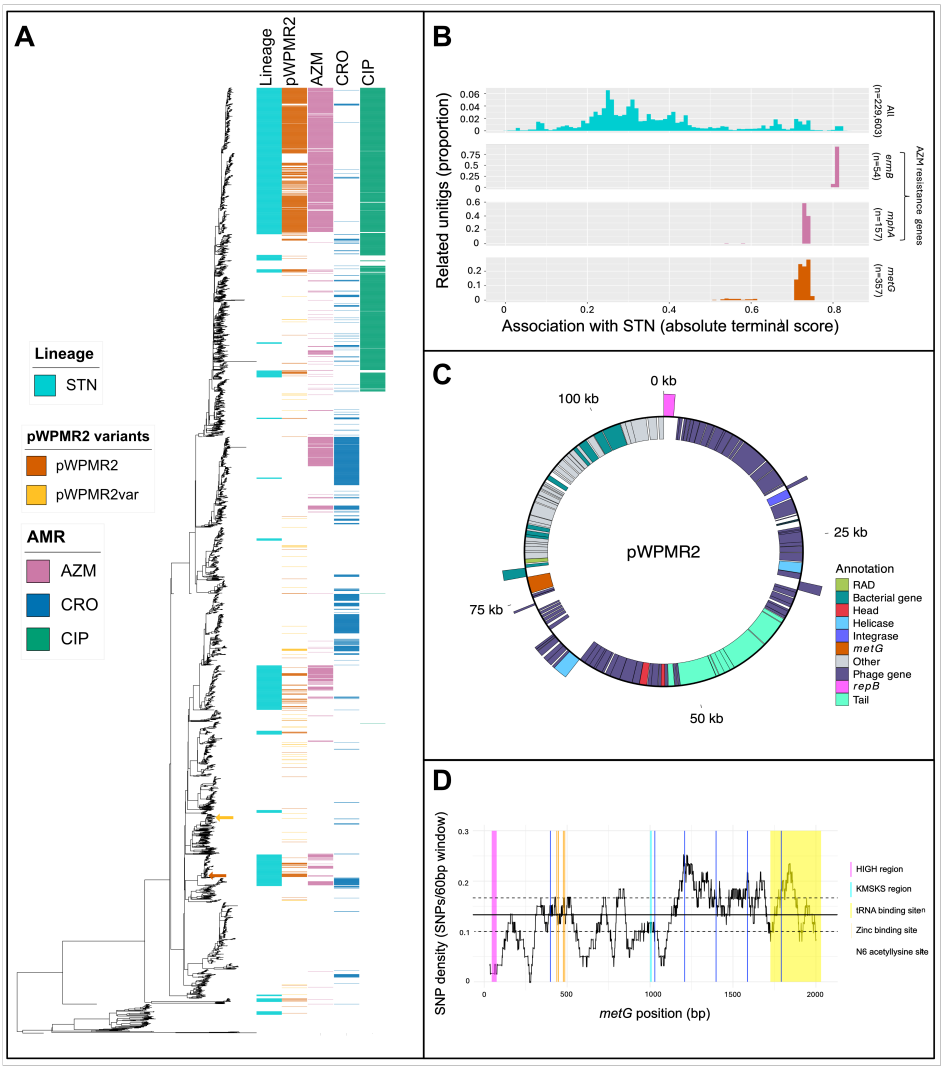

Figure 1. Variable metG borne on pWPMR2 is associated with bacterial lineages on an AMR trajectory. A. A midpoint-rooted maximum likelihood phylogenetic tree constructed from 32,294 recombination-free polymorphic sites shows the evolutionary relationships among 3,745 routine genomic surveillance *S. 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 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 showing forward (outer) and reverse (inner) open reading frames coloured by function according to the inlaid key. **D.** The variation in *metG* across the length of the gene among 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 (undulating line).

pWPMR2 and auxiliary metG are found globally in other clinically relevant organisms

Having associated a novel *metG*-bearing mobile genetic element with genomic lineages on an AMR trajectory, we then explored the extent to which pWPMR2 may have spread among clinically relevant organisms. A comparison of near-identical relatives of pWPMR2 in public databases revealed that it was present in 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 undescribed plasmids from MSM-associated outbreaks of *S. sonnei* [16,18] and O117:H7 Shiga toxin-producing *E. coli* (STEC) [31], across a broad geographical range (Figure 2A). A closely related element was also found to be present in an isolate belonging to an STEC O114:H4 outbreak in Georgia in 2009 [32], as well as an AZM-resistance bearing commensal *E. coli* strain (ST62) from a healthy college student in the USA [33]. 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 as well as 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 [34]. Comparison of its genome-wide sequence similarity with other phage genomes showed that pWPMR2 belonged within the pSLy3 group (Figure 2B, Extended data). The screen of clinical isolates from enteric infections 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%, Table 2). This indicates the mobilisation of pWPMR2 across multiple enteric pathogen species, consistent with previous observations for AMR plasmids [14,35].

		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 plasmid subtypes across enteric bacteria in the UK
[1] restricted to *Klebsiella* in GenBank [2] restricted to *Yersinia* in GenBank.

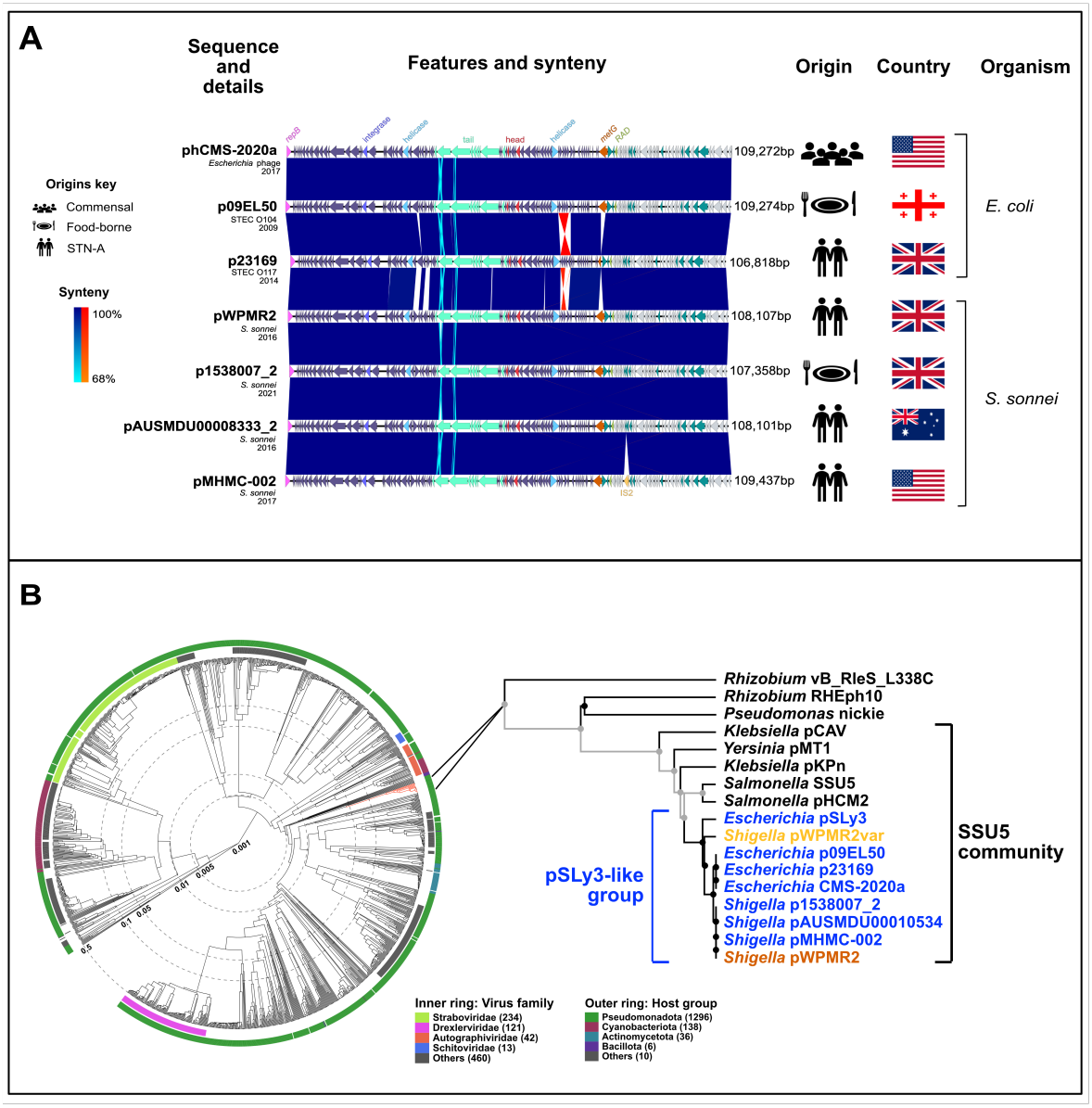

Figure 2. Epidemiological and taxonomic relatives of pWPMR2. A. Comparison of pWPMR2 with its most similar publicly available sequences (labelled leftmost) and their epidemiological details. The sequence comparison is shown as linearised replicons starting with *repB* gene. Certain relevant gene features 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 1,512 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, containing pWPMR2 (orange) and pWPMR2var (yellow).

Given this evidence that pWPMR2 could spread across pathogens, we more fully characterised the temporospatial spread of pWPMR2. We first determined the temporal relationship of pWPMR2 acquisition with AMR across our 3,745 *S. sonnei* surveillance isolates

(Figure 3A). This showed that pWPMR2 entered the *S. sonnei* population earlier than, and then had similar temporal increases to, key AMR genotypes (e.g. the plasmid-borne AZM and CRO resistances). This stepwise acquisition is clear in the internationally disseminated XDR Genotype 3.6.1.1.2 (STN Lineage 1 in Figure 1A, Supplementary table, Figure 1A [9]). This pre-AMR acquisition is consistent with pWPMR2 conferring a phenotype 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 3B, Extended data). This built further genomic epidemiological evidence that pWPMR2 is an emerging mobile genetic element capable of spreading among pathogenic organisms.

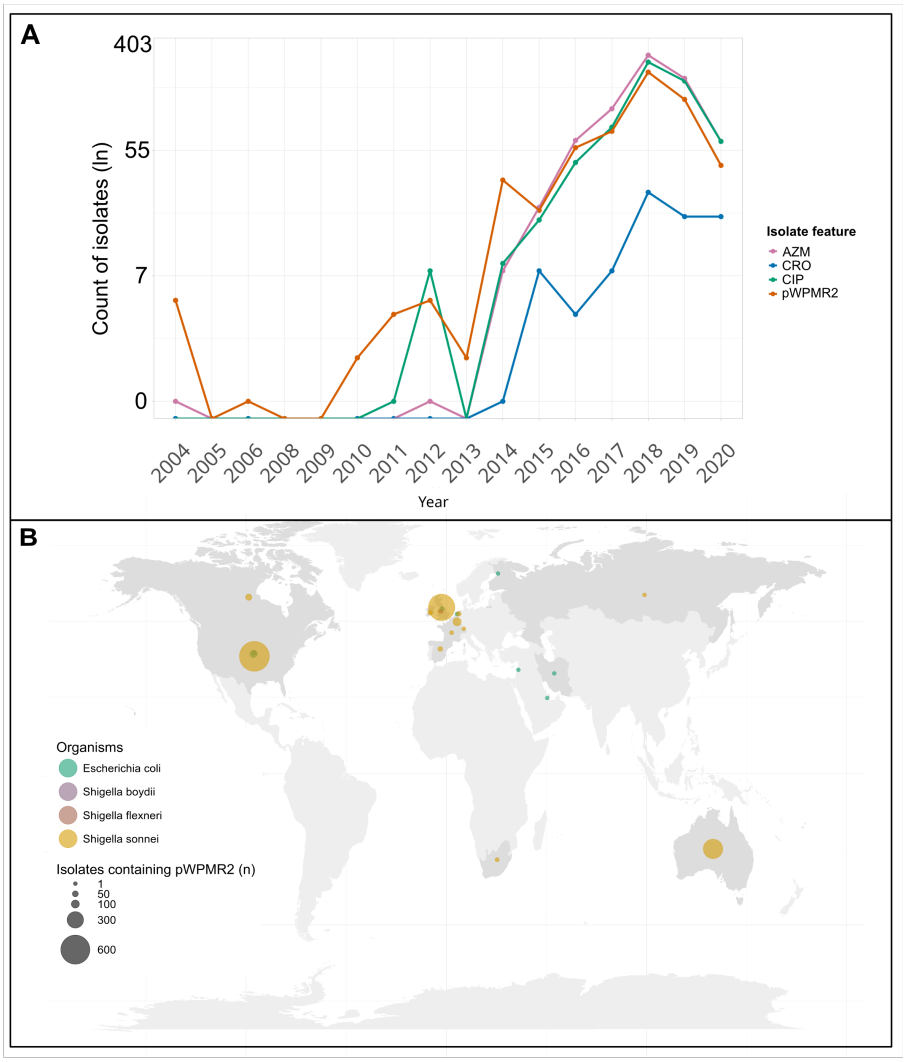

Figure 3. Temporal and spatial patterns of pWPMR2. A. The exponential growth in frequency of isolates containing pWPMR2, and resistance markers against three key antimicrobials among 3,745 *S. sonnei* surveillance isolates is shown over time according to the inlaid key, highlighting the early ingress 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.

To explore the relevance of auxiliary *metG* in the broader bacterial context, we screened sequences (n=59,895) from the PLSDB plasmid database [36] 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 (Methods), 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 ([37], 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 [38], so may be served by functional phenotypes that would be advantageous in adapting to significant antimicrobial selection pressure.

Mutated metG confers evolutionary advantages in the presence of ceftriaxone

Having determined that metG was borne on multiple mobile genetic elements, we sought to explore phenotypes associated with auxiliary *metG* expression and the reduced mutational constraint detected in pWPMR2. Specifically, we conducted laboratory experiments on growth characteristics and antimicrobial response phenotypes of DH5α *E. coli* mutants containing inducible vectors that expressed wildtype (i.e. *S. sonnei* chromosomal) *metG* (hereafter $metG_{WT}$) and the mutated version found on pWPMR2 (hereafter $metG_{PP}$). We first confirmed that *metG* expression did not result in a change in AMR phenotypes, finding Minimum Inhibitory Concentrations (MIC) consistent across strains (Extended data). Bulk growth experiments then revealed that induction of either $metG_{WT}$ or $metG_{PP}$ reduced bacterial growth, with $metG_{WT}$ introducing an increased lag time (consistent with a previously reported tolerance by lag phenotype [39]) and $metG_{PP}$ expression generating diauxic growth (Figure 4A). Using these constructs, we then determined the impact of auxiliary *metG* expression (by comparing $metG_{WT}$ with expression induced and uninduced) and *metG* variation (by comparing induced $metG_{WT}$ with induced $metG_{PP}$) in various phenotypic assays.

As MICs had not changed, we explored a potential benefit of auxiliary *metG* in the presence of sub-inhibitory concentrations of antimicrobials, comparing the relative growth (as measured by ratios of Area Under the Curve, AUC) of constructs in a reduced concentration (0.5 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 (Figure 4B). Similar effects were also seen in comparing the impact of the variation found in *metG* (i.e. the growth performance of $metG_{PP}$ relative to $metG_{WT}$, Figure 4B). 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 determine the role of metG expression and variation in the adaptation to supra-MIC concentrations, 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 4C). Although all *metG*$_{WT}$ populations (induced and uninduced) died by the end of the experiment, expression of auxiliary *metG* adversely affected survival (Figure 4C, top panel). However, this negative impact on survival was ameliorated in the case of *metG*$_{PP}$ expression (Figure 4C bottom panel; Log-rank tests: AZM p=0.005, CIP, p=6e-6, CRO p=0.03). Furthermore, evolution of *metG*$_{PP}$ against CRO was the only condition with surviving populations (n=4) from the final exposure at 1024X MIC. Thus, our results collectively indicate an advantage is offered by the mutations observed on *metG* in the presence of both sub- and supra- inhibitory concentrations of CRO.

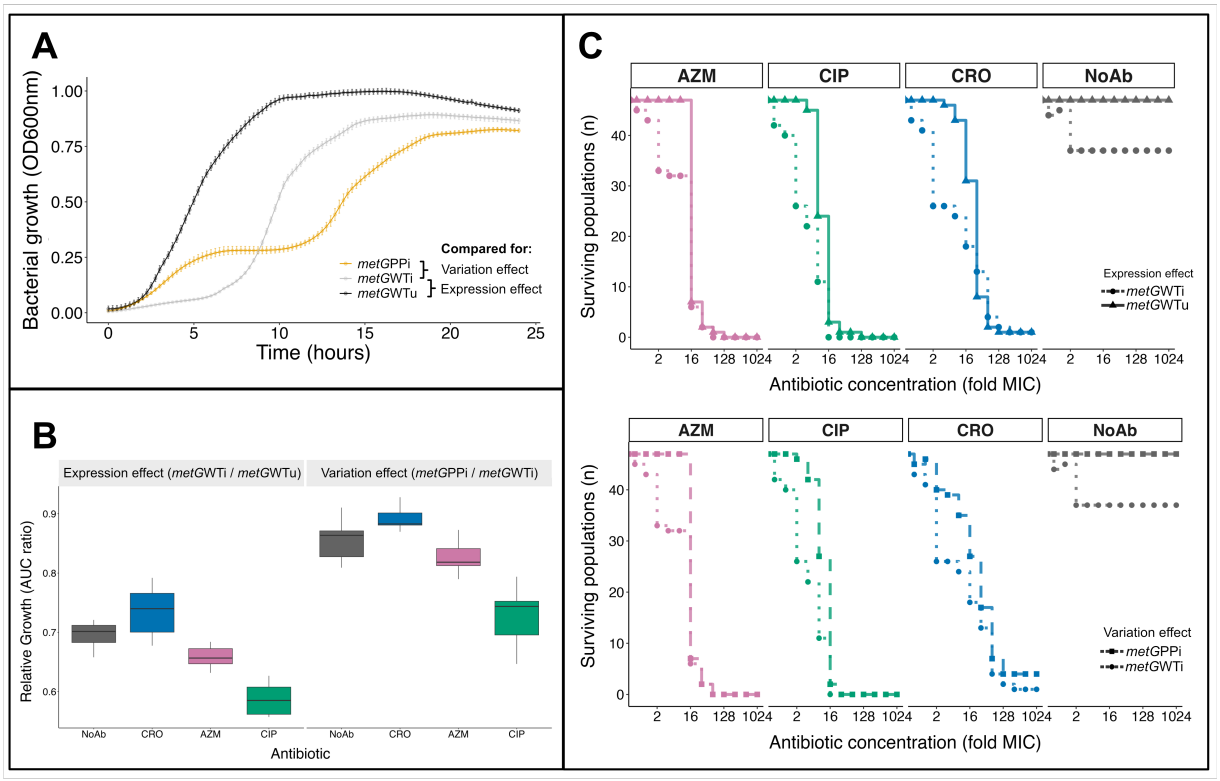

Figure 4. Auxiliary *metG* expression and variation generates altered growth profiles and antimicrobial response phenotypes. A. Growth over time for *E. coli* DH5α constructs with auxiliary *metG*$_{WT}$ expression induced (*metG*WTi) and the respective (i.e. uninduced) control (metGWTu) and the induced pWPMR2 version of *metG* (*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 panel]. CRO = ceftriaxone, AZM = azithromycin, CIP = ciprofloxacin AUC = area under the curve.

metG$_{PP}$ facilitated survival in CRO is mediated by multiple phenotypes

To further resolve the observed advantages in response to CRO, we conducted time-resolved single-cell analysis of the CRO response in strains carrying inducible *metG*$_{WT}$ or *metG*$_{PP}$ plasmids. Briefly, plasmids were introduced into *E. coli* strains carrying a constitutively expressed (under the PrpsL promoter), chromosomally integrated Cyan or Yellow Fluorescence Protein [40] to distinguish pooled strains, and a plasmid-free background strain,

constitutively expressing Red Fluorescence Protein (Figure 5A). The mixed population of three strains was loaded into two separate lanes of a high-throughput microfluidic device [41] such that cells were loaded into trenches randomly within each lane to eliminate systematic biases. In this setup, cells in one lane were exposed to media containing the *metG* inducer, while cells in the other lane received media without the inducer. The narrow trench design of the microfluidic device ensured that the progeny of the cell type occupying the closed end of a trench (the MOTHER cell) replaced cells below, resulting in a pure genotype per trench over time (example of three trenches with randomised loading and eventual purification are shown in Figure 5B). Thus, mixed fluorescence populations, and use of multiple specifically designed microfluidic devices allowed us to compare the effects of *metG* auxiliary expression, similarly as above, but at a single cell level.

We then used this set up to precisely track and analyse the growth and response phenotypes of each genotype towards high-dose CRO treatment (50X MIC for 2 hours from stationary phase). An example time series of cell length from an untreated cell is shown in Figure 5C to indicate variables at baseline in contrast with the diverse response phenotypes observed during and after CRO treatment (Figure 5D). These diverse response phenotypes were classified into four distinct categories including two terminal phenotypes: 1. Cells that visibly lysed or 2. Cells that form long filaments; 3. Viable, But Non Culturable (VBNC), cells that did not grow during or after treatment; and 4. Recovered cells capable of normal growth after treatment. Recovered cells were further divided into two subcategories: Persisters that exited dormancy, and resumed growth, after CRO was removed and Short filament formers that formed short filaments during treatment and resolved into planktonic cells once the antibiotic was removed through a series of asymmetric divisions. Critically, in the absence of the inducer, strains carrying both $metG_{WT}$ or $metG_{PP}$ plasmids showed similar CRO response profiles (see comparable phenotype proportions in Figure 5E) highlighting their utility as controls.

The effect of expressing auxiliary *metG* was then captured through $metG_{WT}$ induction which resulted in a drastic reduction of VBNCs. All cells exited dormancy and resumed growth during CRO treatment with entirely lethal outcomes (i.e. the entire population either lysed or formed long filaments, Figure 5E). This dormancy exit was facilitated by a reduced lag time (defined as the time to reach 1.5X the dormant cell size; Figure 5C). The comparison of lag time distributions between uninduced and induced $metG_{WT}$ populations showed that expression caused all cells to exit dormancy during treatment leaving them susceptible to CRO-induced killing (Figure 5F). This finding is consistent with prior results demonstrating that *metG* is essential for dormancy exit; an extra copy accelerates this process, while a mutated version slows it down [42]. Additionally, the induction of $metG_{WT}$ increased the probability of dormancy exit, evidenced by the reduced fraction of VBNC cells. However, this also reduced overall survival, aligning with bulk observations (Figure 4C, top panel).

Turning then to the effect of *metG* variation, while induction of $metG_{PP}$ also reduced VBNC fractions (though to a lesser extent that induction of $metG_{WT}$), it increased the fraction of recovered cells (Figure 5E). Lag time distributions revealed that $metG_{PP}$ expression led to a higher proportion of persister cells (i.e. those with lag times longer than the treatment period) and cells that formed short filaments during treatment but resolved after CRO was removed (i.e. still recovered, but with lag times shorter than the treatment period, Figure 5G). This distinction was further supported by cell length distributions of recovered cells immediately

after CRO removal (Figure 5H). Specifically, induction of *metG*$_{PP}$ resulted in a greater fraction of longer cells at the point of CRO removal, which subsequently resolved into regular-sized cells, suggesting that *metG*$_{PP}$ introduces greater variability into lag times compared to the wild-type version. This variability facilitates a bet-hedging strategy, where longer lag times allow cells to persist through treatment, and shorter lag times enable early exit from dormancy, leading to filament formation during treatment with recovery to normal division post-treatment.

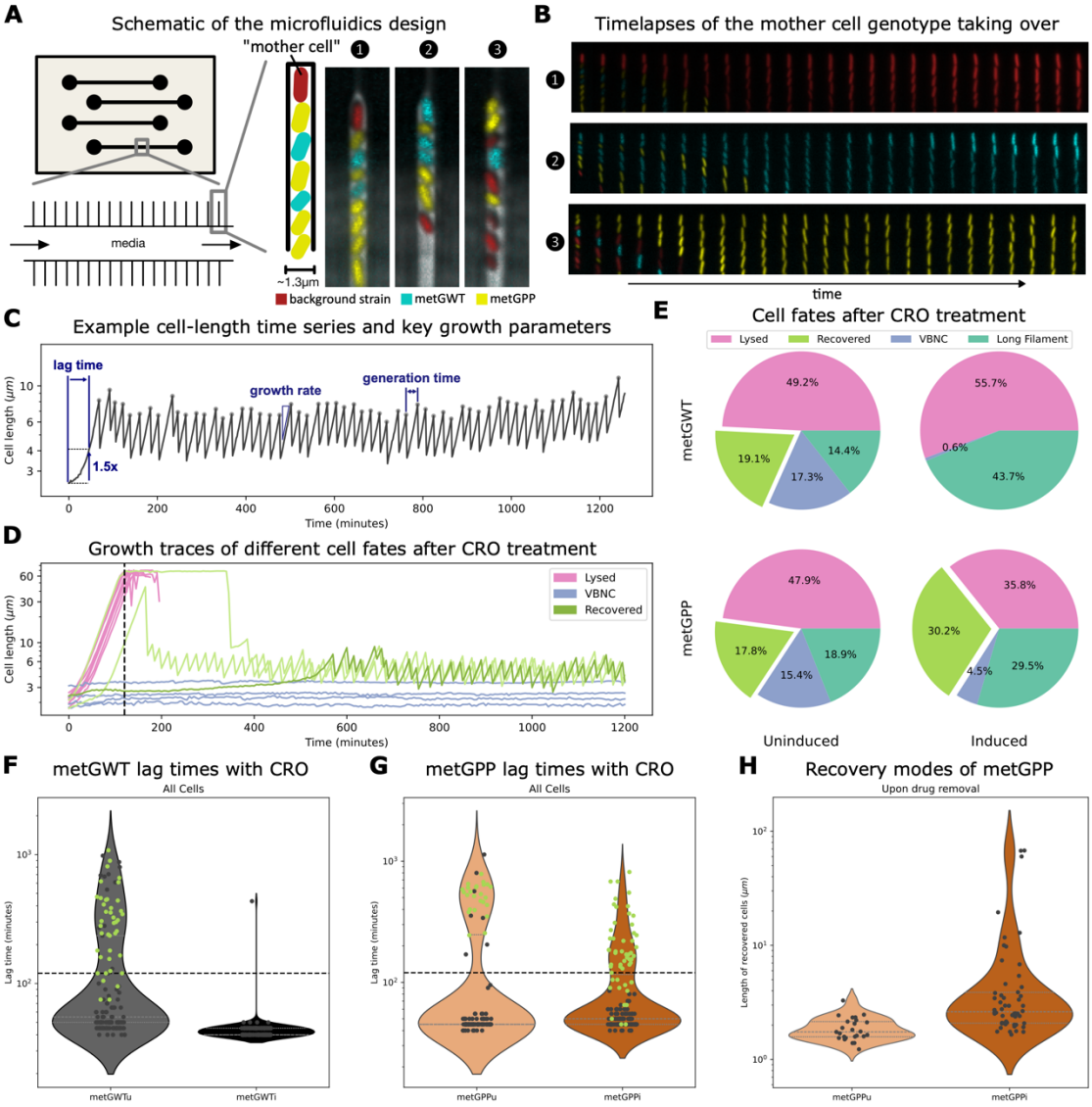

Figure 5: Single-cell analysis of CRO response of *metG*$_{WT}$ and *metG*$_{PP}$ carrying cells

A. A schematic of the high throughput microfluidics device for pooled single cell time lapse experiment. Individual lanes are used to load and treat cells under induced or uninduced conditions. Each lane contains narrow trenches (1.3µm wide) to house isolated cell lineages of cells from a randomised pool (*E. coli* carrying *metG*$_{WT}$ and *metG*$_{PP}$, and their background strain). Strains randomly occupy positions within individual trenches, but the progenies of the cell at the closed end (referred to as the mother cell) eventually occupy the trench, causing each trench to have a pure genotype. Individual cell lineages from all strains are randomly distributed across trenches and share a homogeneous treatment and growth condition from media flowing in the lane throughout specified time. Three example trench images including mixed strains at t=0, where different genotypes occupy the mother position, are shown. Each trench was identified by the "mother cell" located at the closed end of the trench. **B.** Three example timelapses of each trench shown in (A) are shown overtime. The trenches were occupied by the progenies of the mother cell as they were grown from t=0. **C.** Example cell-length time series and key growth parameters measured. The cell-length time series were obtained by segmentation and tracking of individual cells from stationary phase to

exponential growth. From these traces, key single cell growth parameters were determined including lag time, growth rate, and generation time. **D.** Example cell-length traces leading to different fates from CRO treatment (50X MIC, 2h). A fraction of cells lyse during treatment (Pink) or form long filaments which are both considered terminal phenotypes. Blue traces are categorised as VBNC cells which show no visible signs of growth. Green traces represent cells that can recover normal growth after treatment, including cells that are dormant throughout treatment and resuscitate after the antibiotic is removed (Persisters) and cells that form short filaments and resolve into regular sized cells after the antibiotic is removed (Short filament formers). **E.** Proportion of cell fates after CRO treatment shifted by $metG_{WT}$ and $metG_{PP}$ expression. Fractions were obtained by bootstrapping approximately 200 cells for each strain. Inducing $metG_{WT}$ caused most cells to lyse and reduce the fraction of VBNC, while inducing $metG_{PP}$ led to more cells recovered from the treatment, i.e., reduced lysis and VBNC population. **F.** Lag time distribution of *metG*WTu and *metG*WTi cells. Recovered cells are shown as green dots with drug removal indicated by a dashed line. Most *metG*WTi cells have exited dormancy before the drug is removed. **G.** Lag time distribution of *metG*PPu and *metG*PPi cells. Recovered cells are shown as green dots with drug removal indicated by a dashed line. The *metG*PPi were able to recover cells that had exited from a wide range of time points. All the cells that exited dormancy after drug removal were able to recover. **H.** Cell size of $metG_{PP}$ populations just before treatment recovery; many *metG*PPi cells resolved from filamentation through a sequence of asymmetric divisions.

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 with most (n=116/120 isolates) being from 2023 [43]). There were three isolates among this 10-SNP-surveillance-cluster in the pre-outbreak data analysed in Figure 1 (Extended data). These isolates (from 2019 and 2020) contained CRO resistance but did not contain pWPMR2. However, the 120 outbreak isolates from 2022 – 2023 contain pWPMR2 further highlighting the association of pWPMR2 in new 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* facilitated survival of antimicrobial insult and predisposed to the evolution of AMR, but at a significant fitness cost. This substantial fitness cost is consistent with the apparent bet-hedging ecology of *metG* borne pWPMR2 in bacterial populations. Specifically, pWPMR2 is borne at low, sometimes even partial copy numbers (i.e. where infecting pool would have cells with a mixture of 0 or 1 pWPMR2 copies per cell) which would facilitate many cells avoiding the significant fitness cost, while leaving a small population able to survive antimicrobial treatment. The mutated version of *metG* borne by pWPMR2 ($metG_{PP}$) not only reduced the fitness cost, but introduced similarly stochastic population dynamics through single-cell variability in dormancy exit and cell growth under antimicrobial exposure that was not seen for the wildtype ($metG_{WT}$). And finally, as a phage plasmid, pWPMR2, may not actually need its bacterial host to survive antimicrobial treatment to be passed on, with phage particles simply reinfecting new hosts once available. Collectively, this indicates that *metG*-bearing pWPMR2 has several ecological hedges to offer substantial phenotypic advantages under antimicrobial selection whilst minimising its high fitness cost.

These mechanistic insights are entirely consistent with ‘nature’s experiment’ captured in the temporospatial dynamics of pWPMR2. Specifically, pWPMR2 preceding AMR emergence and being distributed among other clinically relevant and commensal organisms, as well as the

detection of other potentially mobilisable *metG* bearing mobile genetic elements in other bacteria. Concerningly, and perhaps potentiating any services to the evolution of AMR, is that the persister phenotype induced by auxiliary $metG_{PP}$ expression might also act to also subvert immunological responses and contribute to chronic infections, which have been observed for some strains of sexually transmissible shigellosis [44,45] and many other AMR pathogens. Hence, pWPMR2 and similar 'hopeful monster' MGEs may be, or may increasingly emerge as, common pathogen adaptations that persist in microbial communities and mobilise among pathogens.

In the face of that possibility, there are several important avenues of future work here. Uncovering the fundamental biology of pWPMR2 and other tRNA ligase-bearing mobile genetic elements in their hosts should be explored. These studies should include aspects of transmission, evolution, and population ecology, as well as mechanistic pathways around gene regulation, expression, and functional impacts for the host. The natural distribution of genetic features that have been detected here should be explored more fully. For example, the screen for the distribution of pWPMR2 (Fig3B) did not detect pWPMR2var (Extended data). Hence, the distribution of pWPMR2var, other *metG* (and other tRNA ligase) bearing mobile genetic elements detected (Extended data), alongside more targeted analyses screening datasets of other known successful AMR bacterial pathogen sublineages for copy number variation in *metG* and other tRNA ligases. These works will determine the extent to which auxiliary tRNA ligases are employed as an evolutionary adaptation in bacterial populations and provide critical information on the underpinning biological pathways.

To end, perhaps the biggest lesson from this work is the value of data driven biology for naturally evolving pathogen populations. We have, as a global public health community, sequenced >2000 pathogenic strains of bacteria containing pWPMR2 since 2004 without understanding its significance. 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 known of tolerance and persistence for decades, these phenotypes are not part of routine surveillance (owing partly to a focus on patient-level outcomes). Our work highlights the need to consider more broadly what phenotypes drive bacterial lineages to epidemiological success. The second issue highlighted is an over-reliance on databases. Bacterial AMR will not 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 facilitating enhanced taxonomic nomenclature and well-maintained feature databases. However, we also need agile systems to detect and investigate changes in pathogen genetic make-ups that we do not understand if we are to better connect pathogen evolution with public health outcomes.

Methods (2295 words)

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) [46]. 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 [47] 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 [48] 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 [49]. A second phylogenetic reconstruction of the evolutionary relationships among 172 representative isolates from the 3745 *S. sonnei* isolates and 123 isolates from MSM-associated $bla_{\text{CTX-M-15}}$ outbreak [43] 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 [50] through R implementation, provided by the package MBASR [51]. 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 [21]. 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 [52], 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 [53] 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 [54]. PCR duplicates were marked using Samtools v 1.19 [54].

For copy number determination and variant calling in *metG,* trimmed short reads were mapped using GEM3 [55] 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 [54].

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 [56]. Assembled contigs were visualised using Bandage v0.9.0. [57]. 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 [58], then reads of less than 150bp and with quality score less than 10 were filtered out using NanoFilt v2.8.0 [59].

Detection of phage sequences, plasmid specific sequences, AMR genes and defence systems were performed by running pWPMR2 through PHASTEST v3.0 [60], PlasmidFinder v2.1 [61,62], AMRfinderPlus v3.12.8 [63], and DefenseFinder [64,65]. We determined the taxonomy of pWPMR2 using the ViPTree server [66] 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 [67], also including examples from each of the five defined phage-plasmid groups within the SSU5 community [34]. 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 [68] 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.) [66,68].

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 [69] or Artemis and ACT [70]. The pWPMR2 sequence was additionally screened against ~2M bacterial genomes in the AllTheBacteria database v0.3 [71] using LexicMap tool v0.4 [72], 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.

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 [73] 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 [74].

Construction of E. coli strains

metG$_{WT}$ sequence obtained from the chromosomal copy of *metG* from *S. sonnei* SRR5005407 and the *metG*$_{PP}$ sequences obtained from the pWPMR2 of *S. sonnei* SRR5005407 were used to synthesise the respective versions of *metG* 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 [75] and graphs were generated using ggplot2 R package [76].

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).

Microfluidic single-cell experiment

The microfluidic devices were fabricated following the protocol described in Wedd et al [41]. On the day before the experiment, the bacterial strains (*metG*$_{WT}$, *metG*$_{PP}$, and background) were inoculated into TSB and incubated overnight with 200 rpm shaking at 37˚C for 18 hours. After incubation, the cultures were pooled in equal proportions and loaded into the device. Individual strains were identified based on their chromosomally integrated fluorescence protein markers, which produced bright signals in dedicated fluorescence channels. Just before the image acquisition, the two syringes with flow medium (with and without inducer) were prepared, as well as with and without CRO (50x MIC, 0.4µg/ml). These media were connected to designated flow lanes in the microfluidic device and switched during the experiment. Time-lapse multi-colour imaging was conducted to monitor the exit dynamics and post-exit growth and division of the cells. Images were captured using a Nikon Eclipse Ti2 inverted microscope with a Hamamatsu C14440-20UP camera. All images were saved in Nikon's ND2 file format and processed to extract single-cell growth traces using the pipeline described by Wedd et al [41], which uses a combination of synthetic training images and machine-learning models for single-cell segmentation [77]. Cells were tracked across frames using a custom script (https://github.com/erezli/MMLineageTracking). The extracted time-series data were analysed to determine key growth parameters, such as lag time, using custom Python scripts.

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). The research in the Bakshi lab was supported by the Wellcome Trust Award (grant number RG89305), a University Startup Award for Lectureship in Synthetic Biology (grant number NKXY ISSF3/46), an EPSRC New Investigator Award (EP/W032813/1) and a seed fund from the School of Technology at University of Cambridge. CJ, CRB, KSB, LCEM and PR 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 (UKHSA), in collaboration with University of Warwick. SN and PR are affiliated with the NIHR HPRU in Genomics and Enabling Data at University of Warwick in partnership with the UKHSA, in collaboration with University of Cambridge and University of Oxford. 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. PR acknowledges the Research/Scientific Computing teams at The James Hutton Institute and NIAB for providing computational resources and technical support for the UK's Crop Diversity Bioinformatics HPC (BBSRC grant BB/S019669/1), use of which has contributed to the results reported within this article. The authors would like to thank Professor Wei Shen for his generous support regarding installing and running AllTheBacteria database and Lexicmap tool. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

706

707 **Author contributions**

708 Conceptualization – KSB, JPJH, SB, RL
709 Data curation – KSB, LC, SCB, CRB, CJ, RL
710 Formal analysis – CRB, YLT, JPJH, LS, GBB, MDS, CEC, SCB, PR, SB, RL
711 Funding acquisition – KSB, JPJH, CJ, MDS, LC, SB
712 Investigation – KSB, LC, YLT, JPJH, LS, SN, MDS, CEC, SCB, PR, SB, RL
713 Methodology – KSB, LC, YLT, PR, SB. RL
714 Project administration – KSB, MDS
715 Resources – KSB, CJ
716 Supervision – KSB, LC, JPJH, SCB, CJ, SB
717 Visualisation – KSB, CRB, YLT, GBB, MDS, CEC, LS, JPJH, SB, RL
718 Writing – original draft – KSB, YLT, JPJH, GBB, MDS, CEC, SB, RL
719 Writing – review & editing – KSB, LCEM, LC, CRB, YLT, JPJH, SN, MDS, CEC, SCB, PR, CJ,
720 SB, RL, GBB, LS

References

1. Pulford, C. V. *et al.* Stepwise evolution of Salmonella Typhimurium ST313 causing bloodstream infection in Africa. *Nat. Microbiol.* **6**, 327–338 (2021).
2. Johnson, J. R., Johnston, B., Clabots, C., Kuskowski, M. A. & Castanheira, M. *Escherichia coli* sequence type ST131 as the major cause of serious multidrug-resistant *E. coli* infections in the United States. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* vol. 51 286–294 (2010).
3. Banerjee, R. & Johnson, J. R. A new clone sweeps clean: The enigmatic emergence of *Escherichia coli* sequence type 131. *Antimicrob. Agents Chemother.* **58**, 4997–5004 (2014).
4. Edquist, P. J., Makitalo, B., Olsson-Liljequist, B., Soderblom, T. & Wisell, K. T. Epidemiology extended-spectrum beta-lactamase-producing *Escherichia coli* Sweden 2007–2011. *Clin Microbiol Infect* **20**, O344–O352 (2007).
5. Whitmer, G. R., Moorthy, G. & Arshad, M. The pandemic *Escherichia coli* sequence type 131 strain is acquired even in the absence of antibiotic exposure. *PLoS Pathog.* **15**, e1008162 (2019).
6. García-Fernández, A. *et al. Klebsiella pneumoniae* ST258 producing KPC-3 identified in italy carries novel plasmids and OmpK36/OmpK35 porin variants. *Antimicrob. Agents Chemother.* **56**, 2143–2145 (2012).
7. Thompson, C. N., Duy, P. T. & Baker, S. The rising dominance of *Shigella sonnei:* An intercontinental shift in the etiology of bacillary dysentery. *PLoS Negl. Trop. Dis.* **9**, e0003708 (2015).
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, L. C. E. *et al.* The evolution and international spread of extensively drug resistant *Shigella sonnei*. *Nat. Commun.* **14**, 1983 (2023).

10. Baker, K. S. *et al.* Intercontinental dissemination of azithromycin-resistant shigellosis through sexual transmission: a cross-sectional study. *Lancet Infect. Dis.* **15**, 913–921 (2015).
11. Mitchell, H. D. *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.* **9**, e0121321 (2021).
12. Mitchell, H. D. *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.* **5**, (2019).
13. Lefèvre, S. *et al.* Rapid emergence of extensively drug-resistant *Shigella sonnei* in France. *Nat. Commun.* **14**, 462 (2023).
14. Baker, K. S. *et al.* Horizontal antimicrobial resistance transfer drives epidemics of multiple *Shigella* species. *Nat. Commun.* **9**, 1462 (2018).
15. Gilbart, V. L. *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.* **91**, 598–602 (2015).
16. Ingle, D. J. *et al.* Co-circulation of multidrug-resistant *Shigella* among men who have sex with men in Australia. *Clin. Infect. Dis.* **69**, 1535–1544 (2019).
17. Mason, L. C. E. *et al.* Using demographics of patients to inform treatment of shigellosis in England. *Lancet Microbe* 101026 (2024).
18. Worley, J. N. *et al.* Genomic drivers of multidrug-resistant *Shigella* affecting vulnerable patient populations in the United States and abroad. *MBio* **12**, (2021).
19. Bardsley, M. *et al.* Persistent transmission of shigellosis in England is associated with a recently emerged multidrug-resistant strain of *Shigella sonnei*. *J. Clin. Microbiol.* **58**, (2020).
20. Hawkey, J. *et al.* Global population structure and genotyping framework for genomic surveillance of the major dysentery pathogen, *Shigella sonnei*. *Nat. Commun.* **12**, 2684 (2021).

21. 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.* **14**, e1005958 (2018).
22. Jarvest, R. L. *et al.* Nanomolar inhibitors of *Staphylococcus aureus* methionyl tRNA synthetase with potent antibacterial activity against gram-positive pathogens. *J. Med. Chem.* **45**, 1959–1962 (2002).
23. Jarvest, R. L. *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.* **13**, 1265–1268 (2003).
24. Girgis, H. S., Harris, K. & Tavazoie, S. Large mutational target size for rapid emergence of bacterial persistence. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 12740–12745 (2012).
25. Khare, A. & Tavazoie, S. Extreme antibiotic persistence via heterogeneity-generating mutations targeting translation. *mSystems* **5**, (2020).
26. Levin-Reisman, I. *et al.* Antibiotic tolerance facilitates the evolution of resistance. *Science* **355**, 826–830 (2017).
27. Wein, T. *et al.* Essential gene acquisition destabilizes plasmid inheritance. *PLoS Genet.* **17**, e1009656 (2021).
28. 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).
29. Lozano-Terol, G. *et al.* Relative impact of three growth conditions on the *Escherichia coli* protein acetylome. *iScience* **27**, 109017 (2024).
30. Kremer, M. *et al.* Bacteria employ lysine acetylation of transcriptional regulators to adapt gene expression to cellular metabolism. *Nat. Commun.* **15**, 1674 (2024).
31. Baker, K. S., Dallman, T. J., Thomson, N. R. & Jenkins, C. An outbreak of a rare Shiga-toxin-producing *Escherichia coli* serotype (O117:H7) among men who have sex with men. *Microb. Genom.* **4**, (2018).

32. Ahmed, S. A. *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* **7**, e48228 (2012).
33. Stephens, C. *et al.* F plasmids are the major carriers of antibiotic resistance genes in human-associated commensal *Escherichia coli*. *mSphere* **5**, (2020).
34. Pfeifer, E., Moura de Sousa, J. A., Touchon, M. & Rocha, E. P. C. Bacteria have numerous distinctive groups of phage-plasmids with conserved phage and variable plasmid gene repertoires. *Nucleic Acids Res.* **49**, 2655–2673 (2021).
35. Thorley, K. *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.* **23**, 732–739 (2023).
36. Schmartz, G. P. *et al.* PLSDB: advancing a comprehensive database of bacterial plasmids. *Nucleic Acids Res.* **50**, D273–D278 (2022).
37. diCenzo, G. C., Mengoni, A. & Perrin, E. Chromids aid genome expansion and functional diversification in the family Burkholderiaceae. *Mol. Biol. Evol.* **36**, 562–574 (2019).
38. Clough, S. E., Elphinstone, J. G. & Friman, V.-P. Plant pathogenic bacterium *Ralstonia solanacearum* can rapidly evolve tolerance to antimicrobials produced by *Pseudomonas* biocontrol bacteria. *J. Evol. Biol.* **37**, 225–237 (2024).
39. Levin-Reisman, I., Brauner, A., Ronin, I. & Balaban, N. Q. Epistasis between antibiotic tolerance, persistence, and resistance mutations. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 14734–14739 (2019).
40. Zaslaver, A. *et al.* A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nat. Methods* **3**, 623–628 (2006).
41. Wedd, C. *et al.* Single-cell imaging of the lytic phage life cycle in bacteria. *bioRxiv* (2024) doi:10.1101/2024.04.11.588870.
42. Blattman, S. B. *et al.* Identification and genetic dissection of convergent persister cell states. *Nature* (2024) doi:10.1038/s41586-024-08124-2.

833 43. Charles H., Greig D.R., Swift C., Olonade I., Simms I., Sinka K., Baker K.S., Godbole G.,
834 Jenkins C. Outbreak of sexually transmitted *S. sonnei* $bla_{CTX\text{-}M\text{-}15}$ in England: an
835 epidemiological and genomic investigation. *medRxiv* (2024)
836 doi:10.1101/2024.10.14.24314996.
837 44. Allen, H. *et al.* Evidence for re-infection and persistent carriage of *Shigella* species in adult
838 males reporting domestically acquired infection in England. *Clin. Microbiol. Infect.* **27**,
839 126.e7-126.e13 (2021).
840 45. Bengtsson, R. J. *et al.* Accessory genome dynamics and structural variation of *Shigella*
841 from persistent infections. *MBio* **12**, (2021).
842 46. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina
843 sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
844 47. Liu, Y., Schröder, J. & Schmidt, B. Musket: a multistage k-mer spectrum-based error
845 corrector for Illumina sequence data. *Bioinformatics* **29**, 308–315 (2013).
846 48. Croucher, N. J. *et al.* Rapid phylogenetic analysis of large samples of recombinant
847 bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* **43**, e15 (2015).
848 49. Minh, B. Q. *et al.* IQ-TREE 2: New models and efficient methods for phylogenetic
849 inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
850 50. Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model
851 choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
852 51. Heritage, S. MBASR: Workflow-simplified ancestral state reconstruction of discrete traits
853 with MrBayes in the R environment. *bioRxiv* (2021) doi:10.1101/2021.01.10.426107.
854 52. Chen, S. Ultrafast one-pass FASTQ data preprocessing, quality control, and
855 deduplication using fastp. *Imeta* **2**, e107 (2023).
856 53. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods*
857 **9**, 357–359 (2012).
858 54. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, (2021).
859 55. Marco-Sola, S., Sammeth, M., Guigó, R. & Ribeca, P. The GEM mapper: fast, accurate
860 and versatile alignment by filtration. *Nat. Methods* **9**, 1185–1188 (2012).

56. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
57. Wick, R. R., Schultz, M. B., Zobel, J. & Holt, K. E. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* **31**, 3350–3352 (2015).
58. De Coster, W. & Rademakers, R. NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics* **39**, (2023).
59. De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M. & Van Broeckhoven, C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* **34**, 2666–2669 (2018).
60. Wishart, D. S. *et al.* PHASTEST: faster than PHASTER, better than PHAST. *Nucleic Acids Res.* **51**, W443–W450 (2023).
61. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
62. Carattoli, A. & Hasman, H. PlasmidFinder and in silico pMLST: Identification and typing of Plasmid replicons in whole-genome sequencing (WGS). *Methods Mol. Biol.* **2075**, 285–294 (2020).
63. Feldgarden, M. *et al.* AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci. Rep.* **11**, 12728 (2021).
64. Tesson, F. *et al.* Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nat. Commun.* **13**, 2561 (2022).
65. Tesson, F. *et al.* Exploring the diversity of anti-defense systems across prokaryotes, phages and mobile genetic elements. *Nucleic Acids Res.* (2024) doi:10.1093/nar/gkae1171.
66. Nishimura, Y. *et al.* ViPTree: the viral proteomic tree server. *Bioinformatics* **33**, 2379–2380 (2017).
67. Mihara, T. *et al.* Linking virus genomes with host taxonomy. *Viruses* **8**, 66 (2016).

68. Nair, S. *et al.* Presence of phage-plasmids in multiple serovars of *Salmonella* enterica. *Microb. Genom.* **10**, (2024).
69. Sullivan, M. J., Petty, N. K. & Beatson, S. A. Easyfig: a genome comparison visualizer. *Bioinformatics* **27**, 1009–1010 (2011).
70. Carver, T. J. *et al.* ACT: The Artemis Comparison Tool. *Bioinformatics* **21**, 3422–3423 (2005).
71. Hunt, M., Lima, L., Shen, W., Lees, J. & Iqbal, Z. AllTheBacteria - all bacterial genomes assembled, available and searchable. (2024) doi:10.1101/2024.03.08.584059.
72. Shen, W. & Iqbal, Z. LexicMap: efficient sequence alignment against millions of prokaryotic genomes. *bioRxiv* (2024) doi:10.1101/2024.08.30.610459.
73. Schwengers, O. *et al.* Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microb. Genom.* **7**, (2021).
74. Wickham, H. *et al.* Welcome to the tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
75. Blazanin, M. gcplyr: an R package for microbial growth curve data analysis. *BMC Bioinformatics* **25**, 232 (2024).
76. Wickham, H. Ggplot2. *Wiley Interdiscip. Rev. Comput. Stat.* **3**, 180–185 (2011).
77. Hardo, G., Noka, M. & Bakshi, S. Synthetic Micrographs of Bacteria (SyMBac) allows accurate segmentation of bacterial cells using deep neural networks. *BMC Biol.* **20**, 263 (2022).