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

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 (232 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, and World Health Organisation AMR priority 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 cells predisposed to survival in, and evolving resistance against, third generation cephalosporins (3GC). 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 (5111 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 (i.e. adapting to high levels of antimicrobial selection pressure). 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]. These STN lineages comprise frequently co-circulating polyphyletic lineages in an otherwise geographically associated population structure where isolates from diverse endemic settings (sampled from returning travellers) are phylogenetically proximate[9,10,13,14]. The antimicrobial use in STN is evidenced empirically (one report revealed 43% of shigellosis-affected MSM had received antimicrobials in the last 6 months[15]) and the convergent acquisition of AMR plasmids is frequently bystander resistance resulting from high intensity treatment for other traditional STIs (e.g. gonorrhoea, chlamydia)[9,10,16–19]. 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 to high selective pressure from antimicrobial use with a high potential for global spread.

In this study, we identify convergently evolving factors across 15 STN lineages among some 3,745 *Shigella sonnei* isolates from routine surveillance from 2004 to 2020 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 survivor 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 routine surveillance in the United Kingdom between 2004 and 2020 (Fig. 1a, Supplementary Table 1). 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, Fig. 1a, Table1, Extended Data Fig. 4, 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*[20]; ciprofloxacin (CIP), azithromycin (AZM), and ceftriaxone (CRO) (74% [n=765/1,028] compared with 37% [n=1011/2,717], P <0.001, Table 1, Supplementary Table 1). 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[10,17,18,21–23].

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*[24]. There were 229,559 unitigs with a minor allele frequency filter of 0.05 and had a distribution whereby the majority of unitigs (84%, n=192,832) had an absolute terminal association score of <0.56 (Fig. 1b, Supplementary Tables 2 and 3A, similar results with 'subsequent score' Extended Data Fig. 1). We then considered unitigs with scores over this threshold (n=36,727) as 'highly associated' with STN and explored their annotations. To do so, we mapped highly associated unitigs (n=36,727) to a concatenated reference containing the bacterial chromosome and the known AMR-plasmid pKSR100[17] and retrieved the annotations of mapping locations. Overall, 544 highly associated unitigs mapped to 50 chromosomal genes and 12 intergenic regions (Extended Data Fig. 2). The gene with the strongest support (i.e. most unitigs with the highest terminal score, Extended Data Fig. 2, Supplementary Data 3) was *metG.* This gene shared similar association scores with the AZM resistance genes *mphA* and *ermB* (Fig. 1b) which act as positive controls as they are known to have emerged globally on pKSR100 among *Shigella* in STN (Fig. 1a)[10,17,18,21–23]. The gene *metG* encodes methionine tRNA ligase, which plays a core role in translation, linking the amino acid methionine (a start codon amino acid) to its cognate tRNA molecule (and is thus also involved in chain elongation). In laboratory settings, disruption to *metG* alters growth characteristics and is considered a canonical mutational route to antimicrobial tolerance (i.e. short-term survival against killing concentrations of antimicrobials without a shift in AMR phenotype[25–28], which enables subsequent evolution of AMR[29]). For these reasons, we hypothesised that the auxiliary *metG* was assisting *Shigella* lineages during circulation in the high antimicrobial selection environment of STNs.

Initial bioinformatic analyses of coverage depth and variation in *metG* revealed both copy number and Single Nucleotide Polymorphisms (SNPs) variation (Extended Data Figs 5 & 6) which were challenging to interrogate with short-read data. Hence, we conducted complementary long-read sequencing of STN Lineage 12 strain SRR5005407 (Fig. 1a) which revealed that the extra copies of *metG* were carried on a 108,107bp circular sequence which we called pWPMR2 (GenBank accession: PQ180110). The pWPMR2 sequence was found to

contain both plasmid replication and phage structural genes (i.e. capsid, tail) indicating a phage-plasmid (Fig. 1c). Furthermore, pWPMR2, also contained a viral anti-defence system (RAD[30]) consistent with a phage lifestyle. Mapping of our data to the new SRR5005407 genome reference confirmed that many highly associated unitigs (36%, n=13,187/36,727) belonged to pWPMR2 (Extended Data Fig. 3). Mapping coverage across our dataset revealed variation in pWPMR2 copy number and coverage (Extended Data Fig. 5, Supplementary Table 1). We inferred that 650 *S. sonnei* isolates contained pWPMR2 (using frequency distribution-informed cut offs, see Extended Data Fig. 5) which mostly (87%, n=566/650, P <2.2e-16) belonged to STN lineages (Fig. 1a, Table 1). Furthermore, while most pWPMR2-containing isolates contained one to two copies of pWPMR2, there was a wide range of relative depth (Extended Data Fig. 5b), suggesting possible instability of pWPMR2 carriage. Although this could also potentially arise from a laboratory artefact, pWPMR2 instability would be consistent with previous studies showing that acquisition of essential genes can compromise plasmid inheritance[31]. The generalisation of *metG* encoding on pWPMR2 across isolates was confirmed by linear regression of the mapping depth of *metG* and pWPMR2 (R^2 = 0.91, P <2.2e-16, Extended Data Fig. 5c). These analyses also suggested the presence of a *metG*-bearing pWPMR2 variant (pWPMR2var, GenBank accession: PQ521219 from isolate SRR10313630, Extended Data Fig. 8) with reduced (~0.75) coverage that was mostly found outside of STN (Fig. 1a, Table 1, Supplementary Table 1). 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 Fig. 1a – see Supplementary Data for individual assignments

^ Genotyping performed according to [31,32]

165 [&] 53 of which were pWPMR2var
166
167 An analysis of SNPs across *metG* among the isolates revealed that the auxiliary *metG* on
168 pWPMR2 exhibited significant genetic variation relative to the essential chromosomal version.
169 Specifically, isolates without pWPMR2 (i.e. only chromosomal *metG*) showed high levels of
170 conservation, with only 12 of 3095 isolates containing SNPs across five sites in *metG*; four of
171 which introduced non-synonymous changes (found in non-STN Lineages not phylogenetically
172 proximate to STN Lineages), and one synonymous change in seven isolates of STN Lineage
173 15. Contrastingly, isolates that contained pWPMR2 had significant numbers of SNPs (0-118
174 per isolate, mean 66) affecting a total of 250 sites across the 2034bp gene, at minor allele
175 frequencies consistent with a low-copy number mobile genetic element, with SNPs generally
176 being phylogenetically grouped (Extended Data Fig. 6). A sliding window analysis of the
177 variation (density of SNPs) on *metG* showed relative sequence conservation in certain key
178 functional areas of the gene (such as the HIGH and KMSK amino acid motifs) and
179 comparatively higher variation in the N6-acetyllysine sites (Fig. 1d), which might alter *metG*
180 post-translational modification and abundance[32,33], with a knock-on effect on overall protein
181 production. While quantitating gene variation in multicopy genes using short read data is
182 complex, our complete genome sequencing confirmed that the *metG* borne by pWPMR2 in
183 our reference isolate SRR5005407 (see $metG_{PP}$ below) contained 115 SNPs, 6 of which were
184 non-synonymous. This highlights that the additional *metG* on pWPMR2 experiences
185 significant genetic variation relative to single chromosomal copies.
186
187 The origins of this increased genetic variation are unknown but may result from multiple drivers
188 e.g. increased replication/transmission of pWPMR2, increased mutation rates. One potential
189 source of variation is the possibility of multiple introductions of *metG* from an external
190 reservoir, which is supported by the presence of pWPMR2 and pWPMR2var. To examine the
191 possibility of multiple introductions, we phylogenetically reconstructed the evolutionary
192 relationships among *metG* located on pWPMR2/pWPMR2var in our surveillance dataset,
193 alongside those on plasmids of various bacterial species in a public plasmid database
194 (PLSDB, more below). This revealed that while there was evidence of multiple introductions
195 of *metG* into *S. sonnei*, it is likely that a single introduction, associated with pWPMR2, made
196 into *S. sonnei* circulating among sexual transmission networks which was followed by clonal
197 expansion (Extended Data Fig. 7, Supplementary Table 4). This analysis did, however,
198 suggest multiple potential introductions of pWPMR2var, highlighting the need for further
199 extensive study in *metG*-bearing mobile genetic elements (more below).

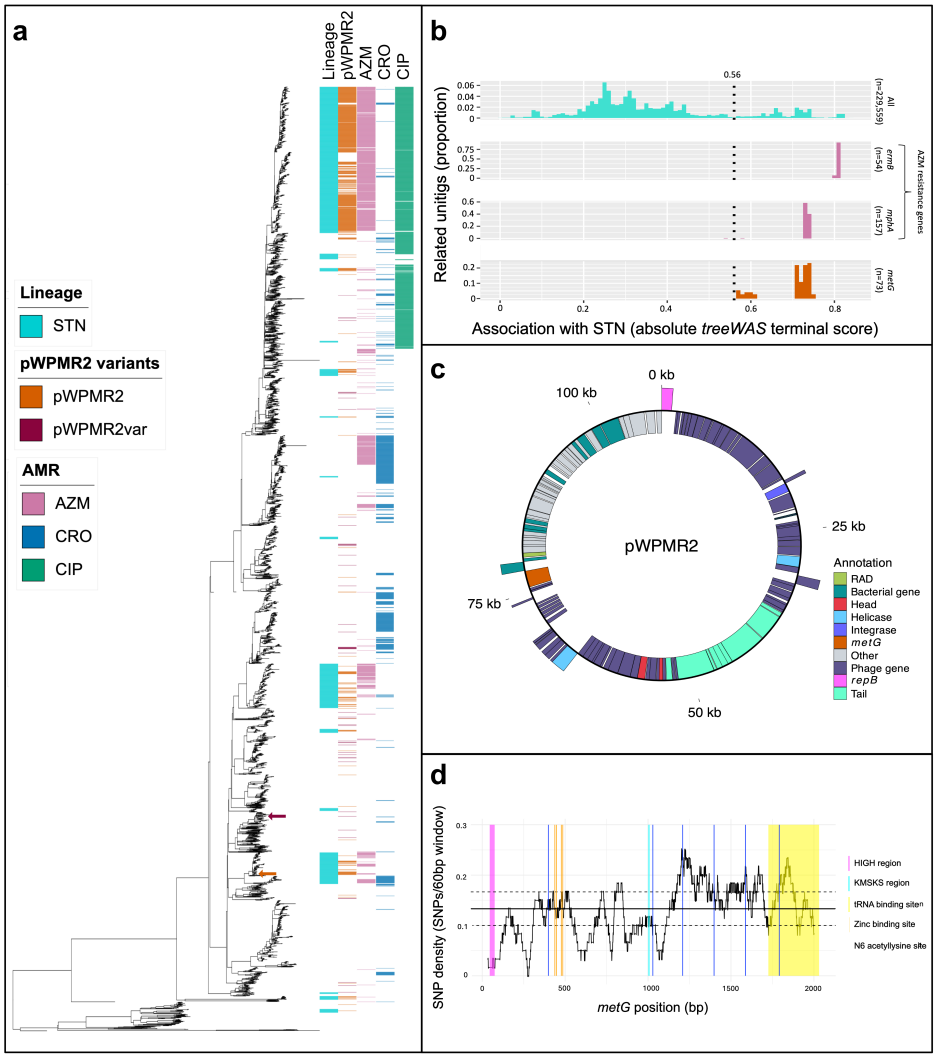

Fig. 1. Variable metG borne on pWPMR2 is associated with bacterial lineages on an AMR trajectory. a, A midpoint-rooted maximum likelihood phylogenetic tree constructed from 32,294 recombination-free polymorphic sites shows the evolutionary relationships among 3,745 routine genomic surveillance *S. sonnei* 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 proportional density of unitgs with different Genome Wide Association Scores (absolute *treeWAS* terminal score). Distributions are shown for all unitigs with ≥0.05 minor allele frequency among the isolates, and unitigs relating to the AZM resistance genes *mphA* and *ermB* (acting here as positive controls) and *metG*. Unitigs with association to STN lineages were defined as those with an association score higher than 0.56 as indicated by the vertical dotted line. **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*[33] and O117:H7 Shiga toxin-producing *E. coli* (STEC)[34], across a broad geographical range (Fig. 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[35], as well as an AZM-resistance bearing commensal *E. coli* strain (ST62) from a healthy college student in the USA[36]. Outside of *Shigella* and *E. coli* the closest relatives were found in *E. albertii* where multiple regions of pWPMR2 were absent (Extended Data Fig. 9). 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[37]. Comparison of its genome-wide sequence similarity with other phage genomes showed that pWPMR2 belonged within the pSLy3 group (Fig. 2b, Supplementary Table 5). 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[17,38].

		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.

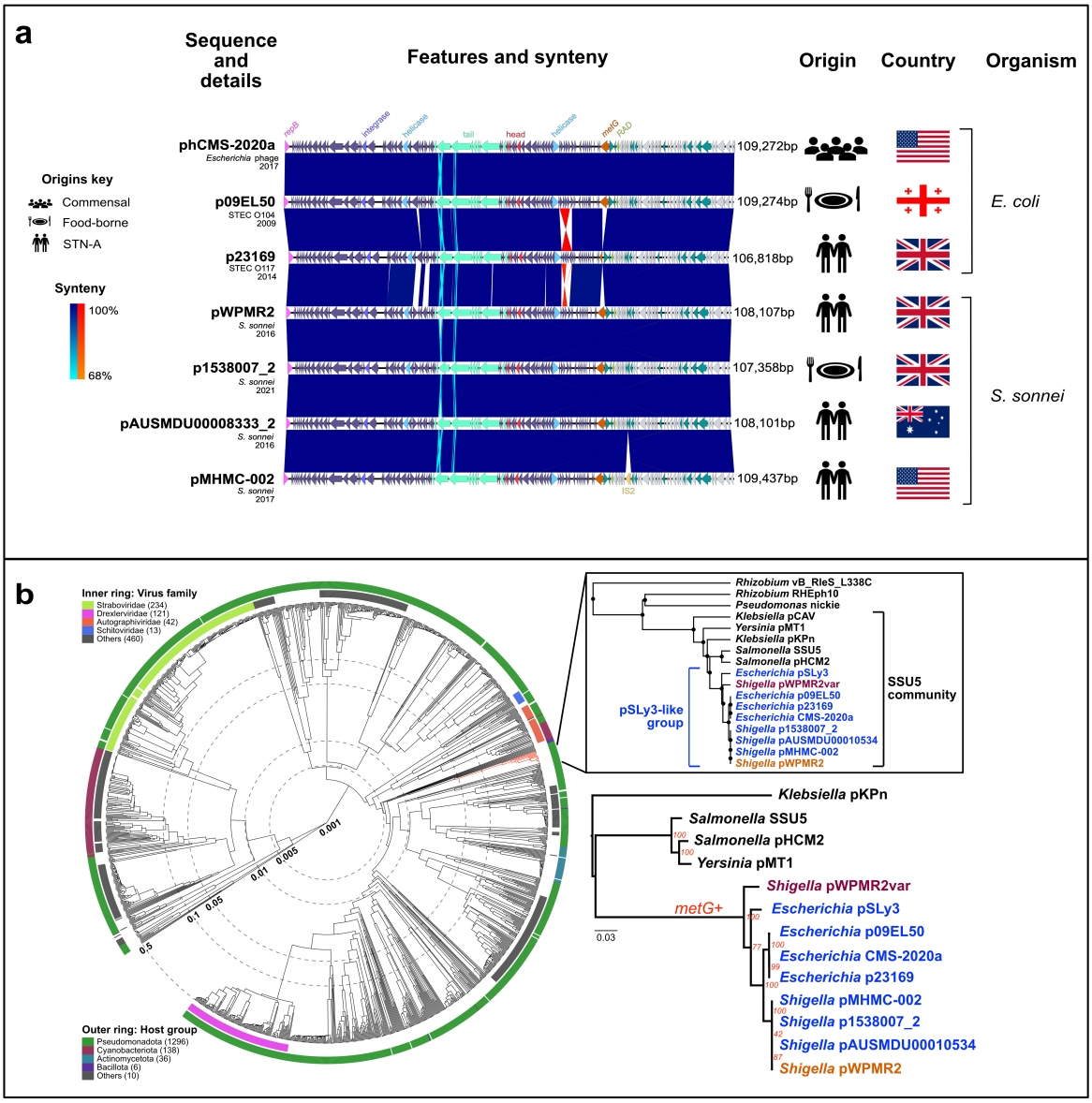

Fig. 2. Epidemiological and taxonomic relatives of pWPMR2. a, Comparison of pWPMR2 with its most similar publicly available sequences (labelled leftmost) and their epidemiological details. The sequence comparison is shown as linearised replicons starting with *repB* gene. Certain relevant gene features 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 ViPTree) 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 zoomed-in view of the clade containing SSU5 and its nearest relatives, highlighting the pSLy3 group (blue), containing pWPMR2 (orange) and pWPMR2var (claret). The lowest dendrogram is a maximum-likelihood phylogenetic tree of the SSU5 community, with bootstrap support values and *metG* acquisition indicated in red.

Given this evidence that pWPMR2 could spread across pathogens and our hypothesis that it was predisposing *Shigella* to AMR, 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 (Fig. 3a). This showed that pWPMR2 entered the *S. sonnei* population earlier than, and then had similar temporal increases to, key AMR profiles (e.g. the plasmid-borne AZM and CRO resistances, and CIP resistance). Furthermore, in a Genotype specific analysis we found that among the 20 (of 67 total) *S. sonnei* Genotypes containing pWPMR2 and an AMR determinant (i.e. against AZM, CRO, CIP) these were more often (n= 17/26) Genotypes where pWPMR2 acquisition preceded AMR acquisition in the than AMR acquisition preceding pWPMR2 acquisition (n=6/26, with 3/26 Genotypes acquiring both pWPMR2 and AMR in the same year, Fig. 3b). The stepwise acquisition was particularly clear in the internationally disseminated XDR Genotype 3.6.1.1.2 (see STN Lineage 1 in Fig. 1a, Supplementary Table 1[38]). 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 (Fig. 3c, Extended Data Fig. 10, Supplementary Table 6). This built further genomic epidemiological evidence that pWPMR2 is an emerging mobile genetic element capable of spreading among pathogenic organisms.

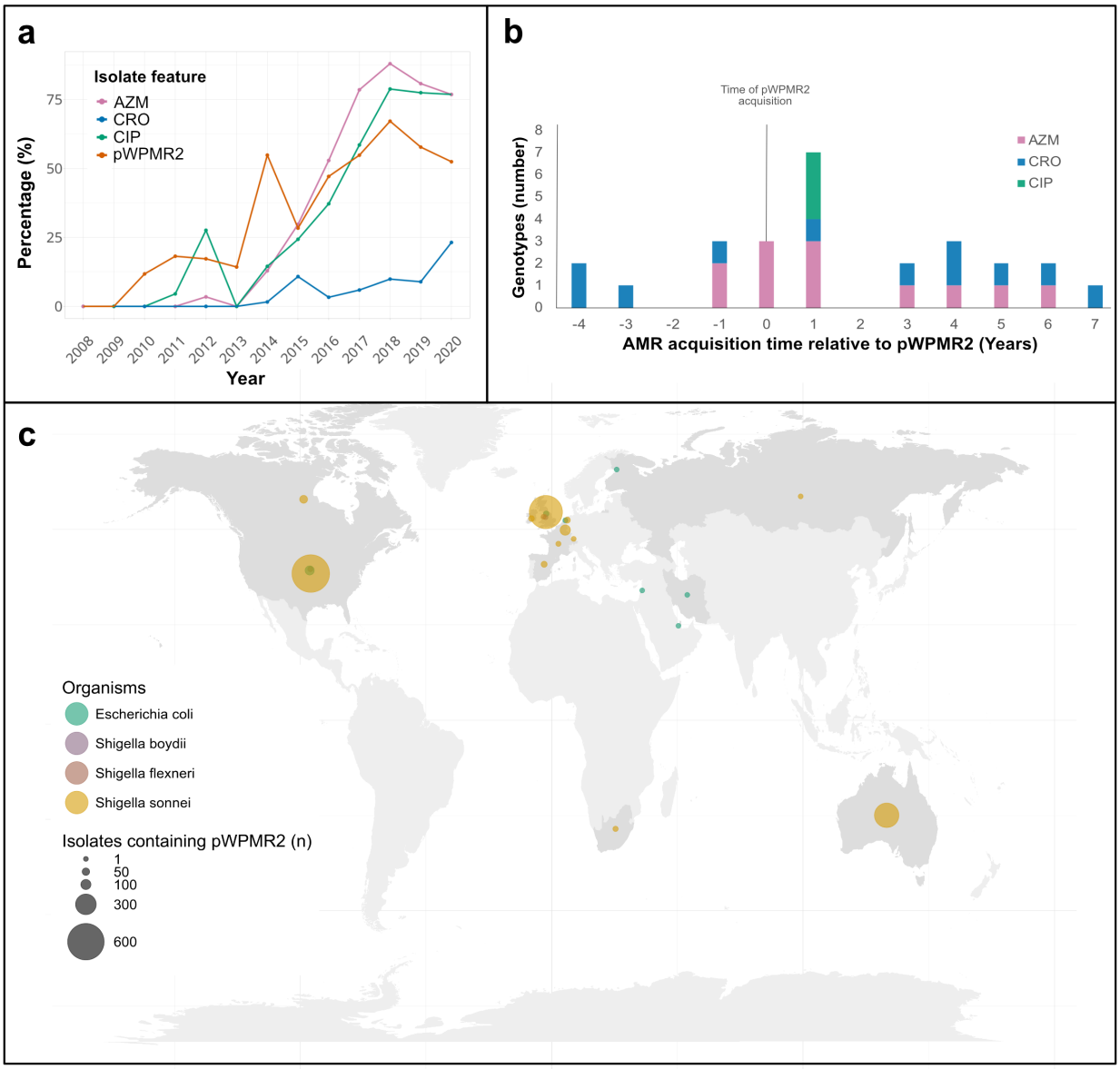

Fig. 3. Temporal and spatial patterns of pWPMR2. a, The growth in the percentage of isolates containing pWPMR2, and resistance markers against three key antimicrobials among 3,745 *S. sonnei* surveillance isolates is shown over time according to the inlaid key, highlighting the early ingress of pWPMR2 into the *S. sonnei* population. **b,** The temporal relationships of pWPMR2 and AMR acquisition within individual Genotypes of *S. sonnei.* The frequency histogram shows the time in years between acquisition of pWPMR2 and acquisition of AMR determinants, coloured according to the inlaid key, among 20 *S. sonnei* Genotypes that contained both pWPMR2 and AMR. **c,** 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[39] for *metG*, and other tRNA ligases. This revealed that tRNA ligase homologues were found at low levels on plasmids (n=513/59895, 0.86%, Supplementary Table 7). 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

307 frequently observed tRNA ligase, detected more often than all the other tRNA ligases put
308 together, and the only one found >5 times in >1 Family. Specifically, *metG* encoding plasmids
309 were predominantly found in Enterobacteriaceae (n=172/29242, 0.59%) and
310 Burkholderiaceae (*Ralstonia* spp., n=95/652, 14.57%). While the Enterobacteriaceae
311 plasmids were overwhelmingly (n=165/172) recorded in the database as Incompatibility group
312 F, B replicon type (FIB) phage-plasmids (consistent with pWPMR2 detection), the *Ralstonia*
313 plasmids were putative chromids (large, domesticated plasmids associated with
314 environmental adaptation) ranging from 1.78 Mb to 2.2 Mb ([40], Extended Data Fig. 11), and
315 their *metG* sequences were (generally) phylogenetically distinct from those of
316 *Enterobacteriaceae* (Extended Data Figure 7). As widespread plant pathogens however,
317 *Ralstonia* can experience various antimicrobial stresses, including molecules naturally
318 secreted from competing soil microorganisms, chemical agricultural inputs, and environmental
319 heavy metals, so they may also be served by contingent functional phenotypes conferred by
320 auxiliary *metG* that could enhance antimicrobial tolerance[41].
321
322 In contrast, when this screening was repeated for databases containing bacteriophage
323 genomes, *metG* was not detected in any genome other than the pWPMR2-like *Escherichia*
324 phage CMS-2020a (Fig. 2a, though some other tRNA ligases were more common
325 [Supplementary Table 8]). While phage-plasmids lacked representation in such curated
326 sequence databases[37,42,43], a recent large-scale computational identification of phage-
327 plasmids and analysis of their pangenomes found *metG* present only in the pSLy3 subgroup
328 of the SSU5 supercommunity[37], though other tRNA ligases are present in different phage-
329 plasmid groups[37,44]. This was reinforced by our screening of UK clinical enteric bacteria where
330 of the 1440 detected SSU5-like phage-plasmids (Table 2), only the members of the pSLy3
331 subgroup carried a copy of *metG*. Therefore, *metG* appears not to be a widespread feature
332 among either bacteriophages or phage-plasmids.

333 **Mutated *metG* confers advantages in the presence of relevant antimicrobials**

334 Having determined that *metG* was borne on multiple mobile genetic elements, we sought to
335 explore phenotypes associated with auxiliary *metG* expression and the observed genetic
336 variation in *metG* on pWPMR2. Specifically, we conducted laboratory experiments on growth
337 characteristics and antimicrobial response phenotypes of *E. coli* SB8 (a derivative of MG1655,
338 see methods) strains containing inducible vectors that expressed wildtype (i.e. *S. sonnei*
339 chromosomal) *metG* (hereafter $metG_{WT}$) and the mutated version found on pWPMR2
340 (hereafter $metG_{PP}$). We first confirmed that *metG* expression did not result in a change in AMR
341 phenotypes, finding Minimum Inhibitory Concentrations (MIC) consistent across strains
342 (Extended Data Table 1). Bulk growth experiments then revealed that induction of either
343 $metG_{WT}$ or $metG_{PP}$ reduced overall bacterial growth, primarily through both $metG_{WT}$ and
344 $metG_{PP}$ introducing an increased lag time, consistent with a previously reported tolerance by
345 lag phenotype[45] (Fig. 4a). Using these constructs, we then determined the impact of auxiliary
346 *metG* expression (by comparing $metG_{WT}$ with expression induced and uninduced) and *metG*
347 variation (by comparing induced $metG_{WT}$ with induced $metG_{PP}$) in various phenotypic assays.
348
349 Although MICs had not changed, we hypothesised that alternative antimicrobial phenotypes
350 might be particularly relevant for evolution in surviving in STN given the high levels of
351 antimicrobial use as well as long half-life (e.g. 68 hours for azithromycin[46]), high plasma
352 concentrations (e.g. ceftriaxone 13-239μg/ml after first dose[47]) and widespread practice of

353 bolus dosing for STIs[48,49]. To explore this possibility, we first explored a potential benefit of
354 auxiliary *metG* in the presence of sub-inhibitory concentrations of antimicrobials (0.5x MIC).
355 Consistent with phenotypes previously associated with *metG* mutations, we saw an increase
356 in lag time with the expression of *metG* (Fig. 4b). Expression of $metG_{WT}$ in the absence of
357 antimicrobials extended lag time by more than two hours, and sub-MIC CIP and CRO
358 extended this further (P = 0.0370 and $P \leq 0.001$ respectively, Kruskal-Wallis tests, Dunn's
359 adjusted). Contrastingly, $metG_{PP}$ had lower lag time penalties compared to $metG_{WT}$, offered
360 increased variance in the no antimicrobial and CIP conditions, and selectively increased lag
361 in the presence of CRO ($P \leq 0.001$, Kruskal-Wallis test, Dunn's adjusted, Fig. 4b). Collectively,
362 these results suggest that $metG_{PP}$ might offer a tolerance by lag phenotype specific to CRO
363 and an increased variance in population lag times which might offer survival advantages under
364 antimicrobial exposure.
365
366 We then determined the role of *metG* expression and variation in the adaptation to supra-MIC
367 concentrations by conducting experimental evolution of 47 independent populations of *E. coli*
368 expressing $metG_{WT}$ and $metG_{PP}$ for eleven days, starting at 0.5x MIC and doubling daily to
369 512x MIC. None of the populations survived until the end for any antimicrobial condition (Fig.
370 4c) and expression of either *metG* had an adverse effect on the number of surviving
371 populations (Fig. 4c, Extended Data Fig. 12). However, we also saw a statistically significant
372 increase in the number of surviving populations as a result of the variation found on $metG_{PP}$
373 (relative to $metG_{WT}$) in the presence of CRO (Fig. 4c). This is consistent with the results above
374 suggesting a potentially 'special' effect for pWPMR2 in tolerating CRO. In fact, when we
375 truncated our experimental data to clinically relevant concentrations of CRO[47] ($\geq$ 16x MIC), we
376 found no statistically significant survival difference among the induced and uninduced
377 populations (Fig. 4c, Extended Data Fig. 13). This further highlighted the important role that
378 $metG_{PP}$ could be playing in bacterial populations adapting to CRO exposure and motivated us
379 to investigate the effect at single cell resolution.

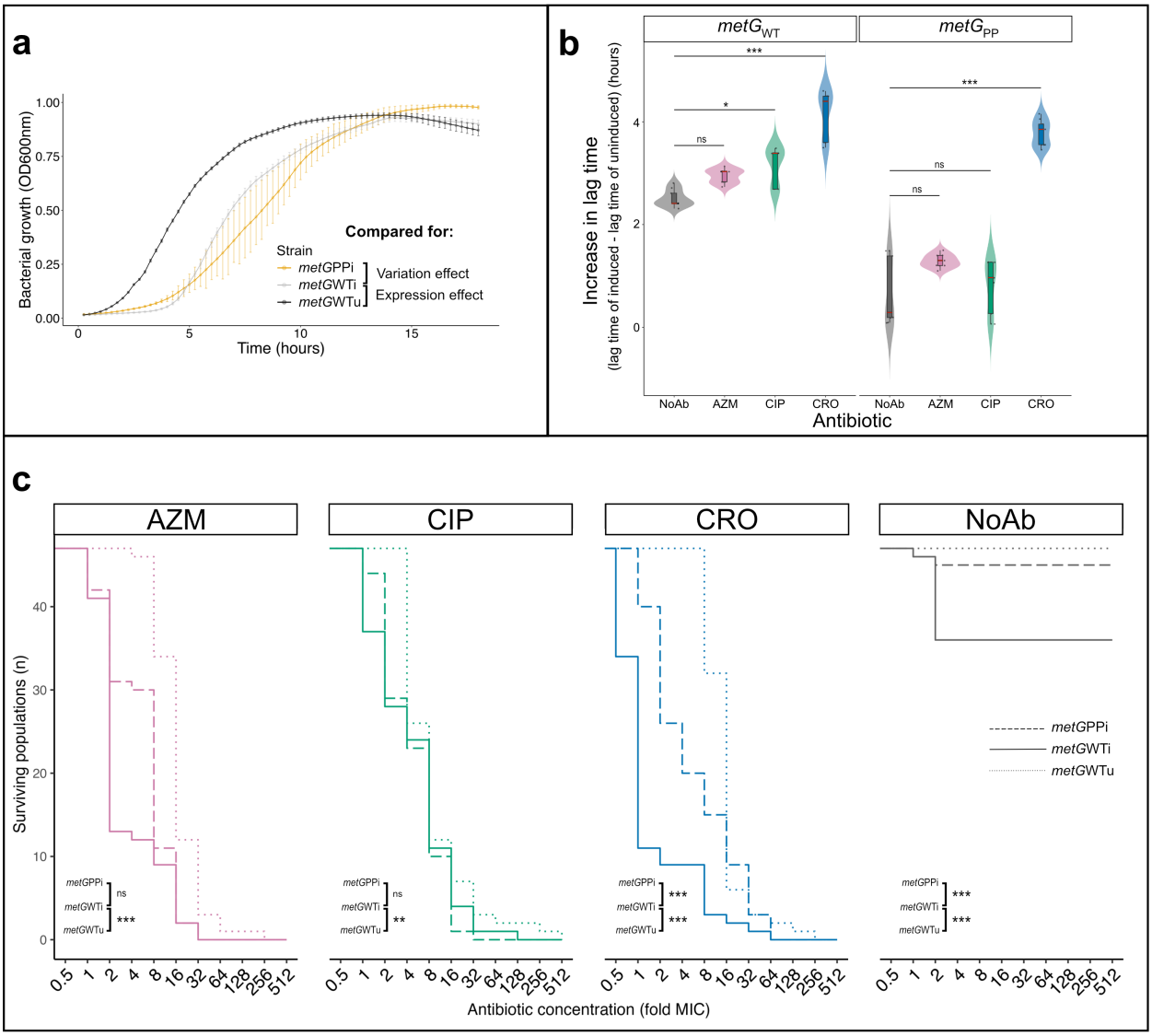

Fig. 4. Auxiliary *metG* expression and variation generates altered growth profiles and antimicrobial response phenotypes. a, Growth over time for *E. coli* SB8 constructs with auxiliary $metG_{WT}$ expression induced (*metG*WTi) and the respective (i.e. uninduced) control (metGWTu) along with the induced pWPMR2 phage-plasmid version of *metG* (*metG*PPi). The error bars show the 95% confidence limits of the mean (marker) of 9 data points (see methods) **b,** Increase in lag time of *E. coli* SB8 strains expressing *metG* constructs (WT and PP) in the presence of subinhibitory concentrations of key treatment antimicrobials (0.5x MIC) compared to the respective uninduced versions. The leftmost plot shows the increase in lag time in $metG_{WT}$ bearing strains in the presence of sub-MIC concentrations of the indicated antimicrobials while the rightmost plot shows the same for the $metG_{PP}$ bearing strains. Box and whisker plots within the violin plots show the median (centre line in red) with the whiskers indicating the 25th and 75th percentile values. Individual data points are also shown within the violin plots. Asterisks denote the statistical significance from pairwise comparisons calculated using Kruskal-Wallis test and corrected by Dunn's test (Benjamini-Hochberg correction) where * denotes P <0.05, *** denotes P <0.001 and "ns" denotes no statistical significance. **c,** Population survival curves of *E. coli* SB8 carrying plasmid constructs of $metG_{WT}$ uninduced (dotted), $metG_{WT}$ induced (solid) and $metG_{PP}$ (dashed) in the presence of indicated antimicrobials starting at 0.5x MIC and going up to 512x MIC. The asterisks for the relevant comparisons are shown in the sub panels where ** denotes P <0.01 and *** denotes P <0.001 and ns denotes no statistical significance as determined by log-rank test. Full statistical analyses including Hazard Ratios are included in Supplementary Table 9. CRO = ceftriaxone, AZM = azithromycin, CIP = ciprofloxacin

404 ***metG*$_{PP}$ facilitated survival in CRO is mediated by multiple phenotypes**
405 To further resolve the observed advantages in response to CRO, we conducted time-resolved
406 single-cell analysis of the CRO response in strains carrying inducible *metG*$_{WT}$ or *metG*$_{PP}$
407 plasmids. Briefly, plasmids were introduced into *E. coli* strains carrying a constitutively
408 expressed (under the PrpsL promoter), chromosomally integrated Cyan or Yellow
409 Fluorescence Protein[50] to distinguish pooled strains, and a plasmid-free background strain,
410 constitutively expressing Red Fluorescence Protein (Fig. 5a). The mixed population of three
411 strains was loaded into two separate lanes of a high-throughput microfluidic device[51] such that
412 cells were loaded into trenches randomly within each lane to eliminate systematic biases
413 (Extended Data Fig. 14). In this setup, cells in one lane were exposed to media containing the
414 *metG* inducer, while cells in the other lane received media without the inducer. Notably, in the
415 absence of antimicrobials, there was no impact on cell growth upon induction of *metG*$_{PP}$
416 (Extended Data Fig. 15). The narrow trench design of the microfluidic device ensured that the
417 progeny of the cell type occupying the closed end of a trench (the MOTHER cell) replaced
418 cells below, resulting in a pure genotype per trench over time (example of three trenches with
419 randomised loading and eventual purification are shown in Fig. 5b). Thus, mixed fluorescence
420 populations, and use of multiple specifically designed microfluidic devices (Extended Data Fig.
421 14 and 15c, d) allowed us to compare the effects of *metG* auxiliary expression, similarly as
422 above, but at a single cell level.
423
424 We then used this set up to precisely track and analyse the growth and response phenotypes
425 of each genotype towards high-dose CRO treatment (6x MIC for 2 hours from stationary
426 phase). An example time series of cell length from an untreated cell is shown in Fig. 5c to
427 indicate variables at baseline in contrast with the diverse response phenotypes observed
428 during and after CRO treatment (Fig. 5d). Diverse response phenotypes among cells after
429 CRO treatment were classified into four distinct categories including two terminal phenotypes:
430 1. Cells that visibly lysed or 2. Cells that form long filaments; 3. Viable, But Non Culturable
431 (VBNC), cells that did not grow during or after treatment; and 4. Recovered cells capable of
432 normal growth after treatment. Recovered cells were further divided into two subcategories:
433 Survivors that exited dormancy, and resumed growth, after CRO was removed and Short
434 filament formers that formed short filaments during treatment and resolved into planktonic cells
435 once the antimicrobial was removed through a series of asymmetric divisions. Critically, in the
436 absence of the inducer, strains carrying both *metG*$_{WT}$ or *metG*$_{PP}$ plasmids showed similar CRO
437 response profiles (see comparable phenotype proportions in Fig. 5e, Extended Data Fig. 16)
438 highlighting their utility as controls.
439
440 The effect of expressing auxiliary *metG* was then captured through *metG*$_{WT}$ induction which
441 resulted in a drastic reduction of VBNCs (Fig. 5e, Extended Data Fig. 16). All cells exited
442 dormancy and resumed growth during CRO treatment with entirely lethal outcomes (i.e. the
443 entire population either lysed or formed long filaments, Fig. 5e). This dormancy exit was
444 facilitated by a reduced lag time (defined as the time to reach 1.5X the dormant cell size; Fig.
445 5c). The comparison of lag time distributions between uninduced and induced *metG*$_{WT}$
446 populations showed that expression caused all cells to exit dormancy during treatment leaving
447 them susceptible to CRO-induced killing (Fig. 5f). This finding is consistent with prior results
448 demonstrating that *metG* is essential for dormancy exit; an extra copy accelerates this
449 process, while a mutated version slows it down[51,52]. Additionally, the induction of *metG*$_{WT}$
450 increased the probability of dormancy exit, evidenced by the reduced fraction of VBNC cells

(Extended Data Fig. 16). However, this also reduced overall survival, aligning with bulk observations (Fig. 4c).

Turning then to the effect of *metG* variation, while induction of $metG_{PP}$ also reduced VBNC fractions (though to a lesser extent than induction of $metG_{WT}$), it increased the fraction of recovered cells (Fig. 5e). Lag time distributions revealed that $metG_{PP}$ expression led to a higher proportion of survivor 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, Fig. 5g). This distinction was further supported by cell length distributions of recovered cells immediately after CRO removal (Fig. 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 survive through treatment, and shorter lag times enable early exit from dormancy, leading to filament formation during treatment with recovery to normal division post-treatment.

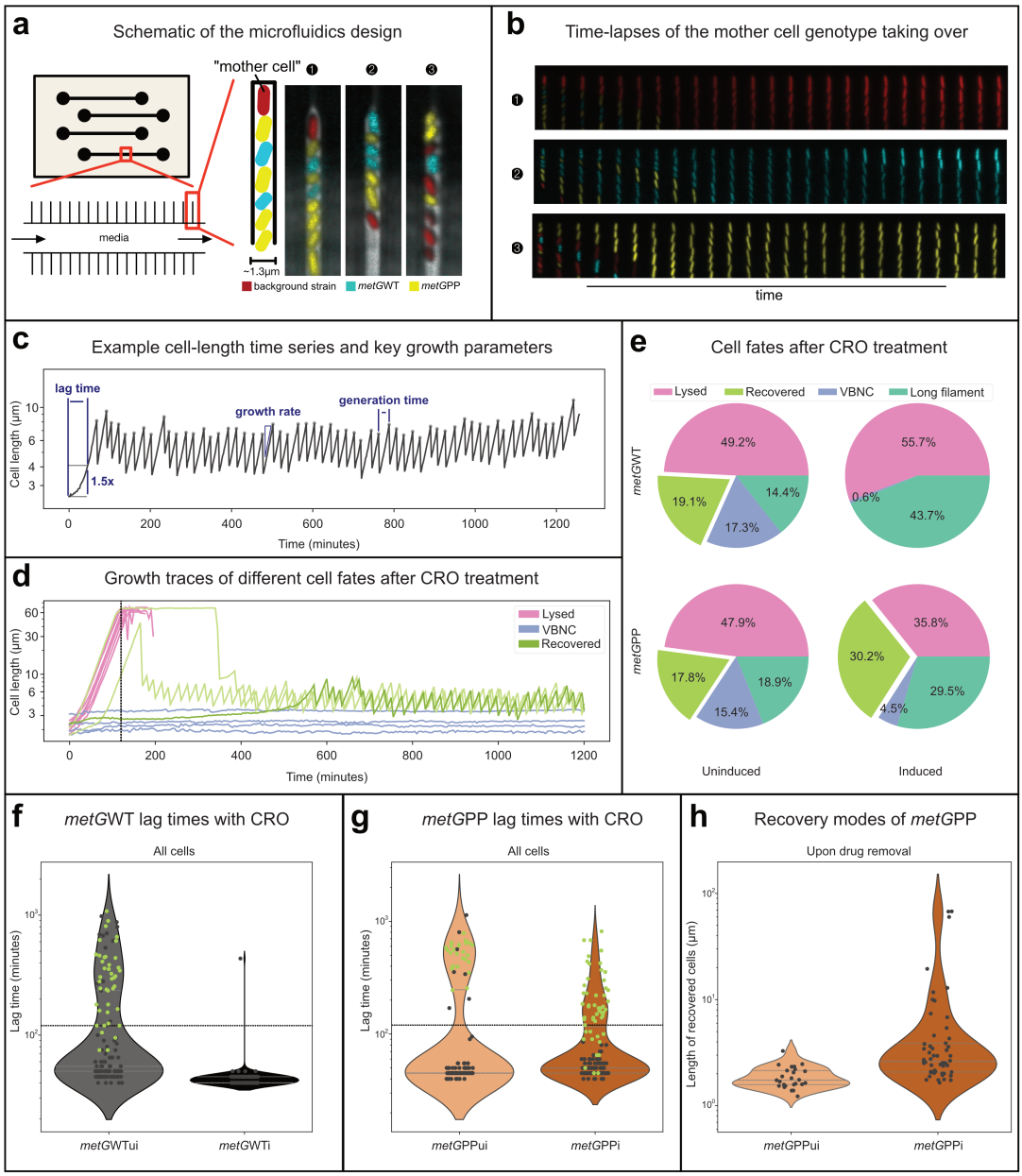

Fig. 5: Single-cell analysis of CRO response of *metG*$_{WT}$ and *metG*$_{PP}$ carrying cells

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 (6x 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 antimicrobial is removed (survivors) and cells that form short filaments and resolve into regular sized cells after the antimicrobial is removed (Short filament formers). **e,** Proportion of fates for cells 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 ($P < 0.001$ between *metG*WTu and *metG*WTi, t-test), while inducing *metG*$_{PP}$ led to more cells recovered from the treatment, i.e., reduced lysis and VBNC population ($P < 0.001$ between metGPPu and metGPPi, t-test). **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. (For figure subsections **f,g** and **h**, the violin plots depict the distribution of data points with the horizontal lines showing the 25th percentile, median, and 75th percentile from bottom to the top while the edges of the violin plots show the upper and lower limits of the data points)

Epilogue: further real-time outbreaks underline the role of pWPMR2

During the period of this work, pWPMR2 was also involved in several new multi- and extensively drug-resistant outbreaks of *S. sonnei* that took place in high income nations, specifically, England, Spain, and Canada in 2023[52,53]. Concerningly, the Canadian outbreak was associated with shigellosis moving into a new demographic of endemic transmission, people living with homelessness, as well as an increased severity of disease[52]. The UK outbreak comprised of 120 cases within a 10-SNP-surveillance-cluster that occurred from 2022 – 2023 with most (n=116/120 isolates) being from 2023[54]). There were three isolates among this 10-SNP-surveillance-cluster in the pre-outbreak data analysed in Fig. 1 (Extended Data Fig. 17). 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. Similarly, isolates from an MSM-CRO-resistant associated outbreak in Spain in October 2023, and in 90% (n=82/91) isolates from an outbreak among people experiencing homelessness in Vancouver 2021 – 2023 also contained pWPMR2 (Extended Data Fig. 17). We also probed whether there was further evidence of pWPMR2 in lower to middle income nations by exploring data from patients who had recently travelled (an informal remote surveillance mechanism)[10,14,55] and found 11 isolates from: Europe, Morocco, Saudi Arabia, Mexico, and Kenya which contained pWPMR2 or pWPMR2var.

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 cost. This is consistent with the apparent bet-hedging ecology of *metG* borne pWPMR2 in bacterial populations. Specifically, pWPMR2 is borne at low, potentially even partial copy numbers (i.e. where infecting pool would have cells with a mixture of 0 or 1 pWPMR2 copies per cell), and high variance in lag time, which would facilitate many cells avoiding the significant lag time increase, while leaving a small population able to survive antimicrobial treatment. The mutated version of *metG* borne by pWPMR2 (*metG*$_{PP}$) not only reduced the increase in lag time, but introduced similarly stochastic population dynamics through single-cell variability in dormancy exit and cell growth under antimicrobial exposure

536 that was not seen for the wildtype (*metG*$_{WT}$). Thus, although conferring resistance-relevant
537 phenotypes, pWPMR2 does not carry AMR genes, as described for some previously
538 described phage-plasmids[56]. And finally, as a phage plasmid, pWPMR2, may not actually
539 need its bacterial host to survive antimicrobial treatment to be passed on, with phage particles
540 simply reinfecting new hosts once available. Collectively, this indicates that *metG*-bearing
541 pWPMR2 has several ecological hedges to offer substantial phenotypic advantages under
542 antimicrobial selection whilst minimising its high collateral effects.
543
544 These mechanistic insights are entirely consistent with 'nature's experiment' captured in the
545 temporospatial dynamics of pWPMR2. Specifically, pWPMR2 conferred advantages against
546 antimicrobials that were clinically relevant for the community under study (i.e. azithromycin
547 and ceftriaxone are used to treat many of the co-morbid STIs in MSM including gonorrhoea
548 and chlamydia[48,49,57]). Furthermore, it's presence preceded AMR emergence and being
549 distributed among other clinically relevant and commensal organisms, as well as the detection
550 of other potentially mobilisable *metG* bearing mobile genetic elements in other bacteria.
551 Concerningly, and perhaps potentiating any services to the evolution of AMR, is that the
552 survivor phenotype induced by auxiliary *metG*$_{PP}$ expression might also act in other ways to
553 advantage AMR organisms. For example, to subvert human immunological responses and
554 contribute to chronic infections, which have been observed for some strains of sexually
555 transmissible shigellosis[58,59] and to subvert bacterial host defences as other phage-encoded
556 tRNA functions have previously been described to do[60]. These alternative functions are an
557 important avenue for future research, particularly as pWPMR2 and other tRNA-ligase
558 encoding MGEs with the potential to contribute to AMR emergence may increasingly emerge
559 as pathogen adaptations that persist and mobilise among microbial communities.
560
561 In the face of that possibility, there are several important avenues of future work here.
562 Uncovering the fundamental biology of pWPMR2 and other tRNA ligase-bearing mobile
563 genetic elements in their hosts should be explored. These studies should include aspects of
564 transmission, evolution, and population ecology, as well as mechanistic pathways around
565 gene regulation, expression, and functional impacts for the host. The natural distribution of
566 genetic features that have been detected here should be explored more fully (e.g. *metG*, and
567 other tRNA ligases on MGEs, and pWPMR2 and pWPMR2var across larger databases). For
568 example, the screen for the distribution of pWPMR2 (Fig. 3b) did not detect pWPMR2var
569 (Supplementary Table 6). Hence, the distribution of pWPMR2var, other *metG* (and other tRNA
570 ligase) bearing mobile genetic elements detected (Supplementary Tables 7 and 8), alongside
571 more targeted analyses screening datasets of other known successful AMR bacterial
572 pathogen sublineages for copy number variation in *metG* and other tRNA ligases is warranted.
573 These works will determine the extent to which auxiliary tRNA ligases are employed as an
574 evolutionary adaptation in bacterial populations and provide critical information on the
575 underpinning biological pathways.
576
577 To end, perhaps the biggest lesson from this work is the value of data driven biology for
578 naturally evolving pathogen populations. We have, as a global public health community,
579 sequenced >2000 pathogenic strains of bacteria containing pWPMR2 since 2004 without
580 understanding its significance. There are two important lessons from this for future
581 microbiological surveillance in the genomic era. The first is that we have an overly narrow
582 focus on certain phenotypes in facing the AMR crisis. While the scientific community has
583 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.

592 **Methods**

593 *Phylogenetic reconstruction*

594 For this study, 3,745 *S. sonnei* isolates sampled in the UK for routine microbiological
595 surveillance, initially sequenced as academic collaborations (isolates representative of routine
596 surveillance from 2004 – 2014, n=329 ERR accessions) and then from routine genomic
597 surveillance (2015 – 2020, n=3416, SRR accessions) were included (Supplementary Table
598 1,[14,61]). Short reads of the 3,745 *S. sonnei* isolates went through the following quality control
599 procedures: raw read data were downloaded from the Sequence Read Archive (Bioproject:
600 PRJNA248792, Supplementary Table 1). Reads were filtered using Trimmomatic v0.39,
601 residual Illumina adapter sequences were removed, the first and last three base pairs in a
602 read were trimmed, before a sliding window quality trimming of four bases with a quality
603 threshold of 20 were applied, and reads of less than 36 bp (after trimming) were removed
604 (ILLUMINACLIP:PE_All.fasta:2:30:10:2:keepBothReads SLIDINGWINDOW:4:20 LEADING:3
605 TRAILING:3 MINLEN:36)[62]. The coverage of the resulting file was estimated against the size
606 of complete reference genome *S. sonnei* 53G (Accession: NC_016822.1) and down sampled
607 to ~100 x coverage. Musket v1.1[63] was applied for k-mer spectrum read correction using a k-
608 mer size of 31bp. The processed reads were then mapped to *S. sonnei* reference genome
609 53G using snippy v4.6.0 (https://github. com/tseemann/snippy), resulting in a core genome
610 alignment of 4,988,504bp in size. After filtering out polymorphic sites within recombined
611 regions using Gubbins[64] with marginal ancestral state reconstruction implemented, 32,294
612 polymorphic sites unaffected by recombination were kept for the final phylogenetic tree
613 reconstruction. A maximum likelihood phylogenetic tree reconstructed was performed using
614 iqtree v2.3.6 with a General Time Reversible (GTR) substitution model with 4 gamma rate
615 categories and 10,000 ultrafast bootstraps to provide approximately unbiased branch support
616 values[65]. A second phylogenetic reconstruction of the evolutionary relationships among 172
617 representative isolates from the 3745 *S. sonnei* isolates and 123 isolates from MSM-
618 associated $bla_{\text{CTX-M-15}}$ outbreak[54] short reads of 295 isolates went through quality control
619 procedures as described in *Determining depth of coverage and variant calling* below. Then, a
620 phylogenetic tree inferred similarly to above from 5,942bp polymorphic sites unaffected by
621 recombination.

622

623 *Identification of sexual transmission network lineages*

624 Out of 3745 *S. sonnei* isolates, 1383 isolates with demographic data combination of male,
625 adult and without travel history to high-risk countries were given presumptive MSM status
626 (pMSM). Ancestral state reconstruction of each internal node in the phylogenetic tree based
627 on the pMSM status of each isolate (tree tips) was performed using MrBayes[66] through R
628 implementation, provided by the package MBASR[67]. The pMSM status of each isolate was
629 refined by the reconstructed ancestral status of their most adjacent nodes, as STN lineages
630 are typically identified by an over-representation of isolates derived from pMSM (see
631 references below). Through plotting the frequency distribution histogram of pMSM
632 probabilities of the most adjacent node for each isolate, pMSM probabilities greater than 0.8
633 were defined as pMSM (Extended Data Fig. 4). For defining STN lineages, all isolates
634 descending from each internal node were defined as belonging to a lineage with the node
635 being the most recent common ancestor. Then, iterating through each node in the tree, a STN
636 lineage was defined when more than 75% of the isolates in the lineage were pMSM. This
637 approach is an extension of, and in line with, previous work and validated surveillance
638 practices[24]. To avoid complete reconstruction of the tree as pMSM, and reflecting the recent

639 emergence of sexually transmissible shigellosis, only the first and second nodes that were
640 most adjacent to the tree tips were used in STN lineages definition.
641
642 *Bacterial genome wide association analysis*

643 Unitig-based bGWAS was run on 3,745 *S. sonnei* isolates with the 1,028 isolates belonging
644 to MSMA lineages being a binary phenotype. Unitigs counting was performed using unitig-
645 caller (https://github.com/bacpop/unitig-caller). The presence-absence matrix of the unitigs
646 was associated with the binary phenotype belonging to MSMA lineages through computing
647 terminal association scores in *treeWAS* v1.1[24]. Unitigs with the highest absolute association
648 scores showed presence-absence patterns that were the most significantly associated with
649 the phenotype. A minor allele frequency filter of 0.05 was applied and the unitigs with greater
650 than 0.56 *treeWAS* terminal and subsequent association score (informed by frequency
651 distribution histogram) were examined for further follow up.

652 Terminal score evaluates how a trait is distributed across the tips (leaves) of a phylogenetic
653 tree. For binary traits, it works by tallying the four possible combinations of presence or
654 absence of both the phenotype and the genotype among the samples. Whereas subsequent
655 score aims at identifying associations that persist in a broad or scattered pattern across the
656 phylogenetic tree[67]. Annotations of these unitigs were identified by mapping them to a
657 concatenated reference consisted of the chromosome and pKSR100 plasmid of *S. sonnei*
658 strain SRR5005407 using bwa mem v0.7.17-r1188[68]. Unitigs belonging to azithromycin
659 resistance genes *mphA* and *ermB* were annotated using BLASTn v2.14.1+ with minimum
660 coverage and percentage identity of 70% and 95% respectively. Mapping of the unitigs to the
661 nine Nanopore sequenced contigs of the same reference strain was performed the same way
662 using bwa as described above.

663 *Determining depth of coverage and variant calling*

664 Illumina reads for 3745 *S. sonnei* isolates were trimmed using fastp v0.23.4[69], using the
665 following options, in the order as appear: auto-detection for adapter for pair-end data (--
666 detect_adapter_for_pe), move a sliding window of size 4bp from front (5') to tail, drop the
667 bases in the window if its mean quality is less than 20, stop otherwise (--cut_front), move a
668 sliding window of size 4bp from tail (3') to front, drop the bases in the window if its mean quality
669 is less than 20, stop otherwise (--cut_tail), global trimming at front and tail for 15 bp (--
670 trim_front1 15, --trim_front2 15), detect and trim polyG in read tails in minimum length of 10
671 bases (--trim_poly_g), detect and trim polyX in read tails in minimum length of 10 bases (--
672 trim_poly_x).

673 The relative depth of pWPMR2 coverage was determined by mapping trimmed short reads
674 using Bowtie2[70] to a concatenated reference that contained the following components:
675 SRR5005407 chromosome (Nanopore contig 1, 4 819 329bp), pKSR100 (contig 8, 65 728bp)
676 and P-P (contig 9, 108 107bp), separated by strings of 500xN. Final destination of any read
677 with multiple equally good alignments was randomly reported as default setting. The relative
678 depth of coverage for each isolate was obtained from the average read depth for the

679 pWPMR2, normalised by that of the chromosome using Samtools v 1.19[71]. PCR duplicates
680 were marked using Samtools v 1.19[71].

681 For copy number determination and variant calling in *metG,* trimmed short reads were mapped
682 using GEM3[72] to SRR5005407 chromosome (contig 1) only, followed by marking PCR
683 duplicates. For *metG* variant calling, the number of reads supporting each allele in each variant
684 position in the *metG* gene region were extracted using bcftools v1.19 mpileup command. They
685 were then converted into .bcf file using bcftools call command, with the following options
686 applied: output all alternate alleles in the alignments regardless of if they appear in any of the
687 final genotype (-A), haploid (--ploidy 1), and alternative model for multiallelic and rare-variant
688 calling (-m). For each variant position in each sample, the presence of an alternative allele
689 was only defined when it was supported by at least 40 reads. Alternative allele proportion was
690 calculated by the number of reads supporting the alternative allele divided by the total read
691 depth of the position in a sample. Copy number of *metG* was obtained based on the same
692 mapping, by computing its average read depth for each isolate, normalised by that of
693 chromosome using Samtools v1.19. PCR duplicates were marked using Samtools[71].

694 *Genetic analyses of pWPMR2*

695 The circular topology of pWPMR2 (Nanopore contig 10) and pWMPR2var (Nanopore contig
696 8) was confirmed by assembling the Nanopore reads of SRR5005407 and SRR13013630
697 respectively using Flye v2.9.5-b1801[73]. Assembled contigs were visualised using Bandage
698 v0.9.0.[74]. Prior assembly, adapter sequences were trimmed and reads with internal adapter
699 were discarded using Porechop v0.2.4 (https://github.com/ rrwick/Porechop). Quality score
700 and length of remaining reads were visualised using NanoPlot v1.43.0[75], then reads of less
701 than 150bp and with quality score less than 10 were filtered out using NanoFilt v2.8.0[76].

702 Detection of phage sequences, plasmid specific sequences, AMR genes and defence systems
703 were performed by running pWPMR2 through: PHASTEST v3.0[77] using default parameters,
704 and phage region information was obtained from detail.txt and summary.txt output files;
705 PlasmidFinder v2.1[78,79] with Enterobacteriales database, a minimum percentage identity of
706 95% and minimum coverage of 60%; and AMRfinderPlus v3.12.8[80], DefenseFinder and
707 AntiDefense Finder using default settings[81,82] We determined the taxonomy of pWPMR2 using
708 the ViPTree server[83] to generate an all-against-all tBLASTx similarity proteomic tree against
709 1,512 related dsDNA prokaryotic viruses from the Virus-Host DB, which utilises complete
710 genomes from NCBI RefSeq and GenBank[84], also including examples from each of the five
711 defined phage-plasmid groups within the SSU5 community[37]. For the detailed SSU5
712 supercommunity phylogeny, Panaroo v1.5.2[85] was used to analyse the pangenomes
713 (annotated using Bakta v1.8.2[86]) of the relatives detected in the NCBI nt database, as well as
714 selected sequences identified in the literature[37]. An alignment of 39,267bp, representing 43
715 filtered shell genes present in 95% of sequences, was passed to IQTree v3.0.0[65] to generate
716 a maximum likelihood phylogenetic tree using the GTR+F+G4 model and 10,000 ultrafast
717 bootstraps. To further explore its association with clinically relevant organisms we downloaded
718 the genomes of its closest relatives, as found through a BLASTn search against the NCBI
719 nucleotide collection (nt) database, selecting sequences with >95% coverage and >98%
720 identity for direct comparison.

721 Five *E. albertii* plasmids were chosen for comparison with pWPMR2 as they represented the
722 top non-*Shigella* and non-*E. coli* hits in the online nBLAST database showing highest
723 percentage identity. To screen 66,929 genomes of enteric bacteria available from clinical
724 cases in the UK for evidence of SSU5-like, pSLy3-like and pWPMR2-like phage plasmids and
725 additional *metG*, a mapping approach as described previously[42] was used. Briefly, target
726 genes including *metG* (CP104415.1:54597..56630), *repB* (CP104415.1:22858..24072), and
727 *repA* (NC_018843.1:59428..60492) were screened in clinical enteric bacteria derived from UK
728 reference laboratory routine genomic surveillance, available under BioProjects PRJNA315192
729 (*E. coli* and *Shigella* spp.) and PRJNA248792 (*Salmonella* spp.)[42,83].

730 Available data on patient travel was linked with isolates containing pWPMR2 and revealed 11
731 isolates (n=4 pWPMR2, n=7 pWPMR2var) with confirmed recent travel history among the
732 surveillance data set. The relevant destinations were Belgium, Hungary, and two isolates from
733 Spain for pWPMR2, and one from each of Morocco, Saudi Arabia, Mexico, Kenya, Bosnia and
734 Herzegovina, and 'Africa' for pWPMR2var.

735 Structural variants between STN-lineage associated pWPMR2 and pWPMR2var, comparison
736 of near-identical relatives of pWPMR2 in public databases, pWPMR2 and *E. albertii* plasmid
737 comparison as well as was visualised using EasyFig genome comparison visualiser[87] or
738 Artemis and ACT[88]. The pWPMR2 sequence was additionally screened against ~2M bacterial
739 genomes in the AllTheBacteria database v0.3[89] using LexicMap tool v0.4[90], with a 70%
740 coverage and 70% identity threshold. Metadata of associated hits were then retrieved using
741 Entrez Direct (https://www.ncbi.nlm.nih.gov/books/NBK25501/). Genotype AMR analysis was
742 done using sonneityper Genotypes (see Table 1) and the earliest year of identification of either
743 pWPMR2 or AZM, CRO, and CIP AMR determinants was identified. Subtracting the year of
744 AMR acquisition from the year of pWPMR2 acquisition was then used to identify the relevant
745 time interval.

746 *Long read sequencing of selected isolates*
747 Isolates were grown in TSB overnight with 200 rpm shaking at 37˚C and 500µl of the overnight
748 culture was then used in DNA extraction using either MasterPure total DNA/RNA extraction
749 kit (Biosearch Technologies) or FireMonkey DNA extraction kit (Revolugen) following
750 manufacturers' recommendations. Extracted DNA was then quantified using a Qubit before
751 sending off to be sequenced, analysed and annotated by commercially available
752 Plasmidsaurus service (https://www.plasmidsaurus.com/) using Oxford Nanopore
753 Technologies (ONT) platform.
754
755 *Analysis of metG introductions*
756 For constructing the maximum likelihood phylogenetic tree that explored introductions of
757 *metG*, reads from *S. sonnei* isolates containing pWPMR2 or pWPMR2var (n=650) went
758 through quality control as described in "*Determining depth of coverage and variant calling*"
759 above. Then were mapped to the (full length) chromosome of SRR5005407, followed by
760 variant calling performed as for copy number determination and variant calling in *metG* (see
761 below). Consensus sequences of *metG* in each sample were obtained using bcftools v1.19[71]
762 consensus command. Gene *metG* gene sequences from 285 *metG* containing plasmids from
763 various bacterial species from PLSDB (Supplementary Data 4) were then extracted. Five
764 plasmids with duplicate metG sequences were removed (Accessions: CP016905.1,

765 CP022791.1, CP088831.1, NZ_CP093325.1, NZ_CP115947.1) to streamline analysis. For
766 facilitating alignment, *metG* sequences from the 650 *S. sonnei* surveillance isolates were
767 deduplicated using seqkit rmdup command[91], resulting in 388 unique *metG* sequences.
768 Together with the 285 *metG* sequences from PLSDB, 673 *metG* sequences were aligned
769 using Kalign3 v3.4.0[92], followed by phylogenetic reconstruction using iqtree v3.0.1 with
770 parameters as described as above[65]. The number of variant sites from the alignment was
771 obtained using SNP-sites v 2.5.1[93].
772
773 *Analysis of plasmid and bacteriophage databases*
774 PLSDB database version 2023_11_23 was downloaded from https://ccb-microbe.cs.uni-
775 saarland.de/plsdb/plasmids/download/ and annotated using Bakta version 1.8.2[86] with the 'full'
776 database. A list of COG categories for tRNA synthetases was downloaded from the COG
777 database and used as search terms in the Bakta output .gff files. To avoid potentially
778 misclassified chromosomal sequences in PLSDB, sequences encoding ≥3 tRNA synthetases
779 were removed from analyses of tRNA synthetase-encoding plasmids. Plasmid types were
780 extracted from PLSDB metadata. Analyses were conducted in R using tidyverse[94]. This
781 analysis was repeated with a curated bacterial prophage database: prokaryotic dsDNA virus
782 sequences downloaded from Virus-Host DB[83] (n=5631). In addition, the annotations provided
783 with the INPHARED phage database[95] (n=28665 excluding RefSeq genomes, downloaded
784 from https://millardlab.org/2025/03/06/phage-genomes-march-2025/) were searched for the
785 presence of tRNA synthetase/ligase products (Supplementary Table 8).
786
787 *Construction of E. coli strains*
788 $metG_{WT}$ sequence obtained from the chromosomal copy of *metG* from *S. sonnei* SRR5005407
789 and the $metG_{PP}$ sequences obtained from the pWPMR2 of *S. sonnei* SRR5005407 were used
790 to synthesise the respective versions of *metG* using GeneArt custom synthesis platform
791 (ThermoScientific). The synthesised genes were then cloned into the inducible pTrcHis2A
792 vector using the same service and the two resulting plasmids were then transformed into *E.*
793 *coli* SB8 (a *motA* deletion variant of *E. coli* MG1655[96]) and selected using the ampicillin marker
794 (100µg/ml) on the vector. Transformed *E. coli* strains were used in the downstream
795 experiments with 1mM IPTG (Millipore Sigma, UK) as the inducer to induce the expression of
796 the *metG* genes.
797
798 *Sub-MIC growth curves*
799 Minimum Inhibitory Concentrations (MIC) for each of the *E. coli* SB8 strains carrying the two
800 different versions of *metG* genes on pTrcHis2A vector were determined using E-strips
801 (Liofilchem, Italy) and/or broth dilution method for azithromycin (0.25µg/ml), ciprofloxacin
802 (0.0078125 µg/ml) and ceftriaxone (0.0625µg/ml). Those strains were then grown in biological
803 and technical triplicates in TSB containing half of the MIC concentrations of each of the three
804 antimicrobials with and without induction of the *metG* variants. The OD_{600nm} values were
805 recorded every 15 minutes using a BioTek Synergy H1 multi-mode plate reader. The resulting
806 data were then analysed using gcplyr R package[97] or the shiny-based version of miLAG[98] and
807 graphs were generated using ggplot2 R package[99].
808
809 *Experimental evolution study*
810 *E. coli* SB8 containing $metG_{WT}$ and $metG_{PP}$ on a pTrcHis2a plasmid backbone and the parent
811 strain without any plasmids were streaked out on TSA plates containing 100µg/ml ampicillin

812 for the plasmid carrying strains and without antimicrobials for the parent strain and incubated
813 overnight at 37˚C to obtain isolated colonies of each of the *metG* variant carrying *E. coli* and
814 the parent strain. These isolated colonies were then inoculated into individual wells of 96 well
815 plates containing either no antimicrobials or AZM, CIP or CRO at a concentration of ½ MIC
816 value for each and treated as an individual population. Each 96 well plate contained 47 wells
817 with TSB with IPTG at a final concentration of 1mM and a control well, 47 wells with no IPTG
818 and a control well. These 96 well plates were then incubated at 37˚C with shaking at 200 rpm
819 for 24 hours before transferring 1% (v/v) into a fresh 96 well plate containing twice the
820 concentration of each of the antimicrobials from the overnight plate. This was carried on for
821 11 days with the final concentration of each antimicrobial reaching 512x MIC. At the end of
822 each of the 24-hour incubation periods, each of the wells were scored for growth using the
823 OD_{600nm} values obtained by end point reads from a BioTek Synergy H1 plate reader and the
824 number of populations survived was plotted against the antimicrobial concentration using the
825 ggplot2 R package. Statistical analysis for the survival of each of the populations was
826 performed using the survival R package (https://CRAN.R-project.org/package=survival).
827
828 *Microfluidic single-cell experiment*
829 The microfluidic devices were fabricated following the protocol described in Wedd et al[51]. On
830 the day before the experiment, the bacterial strains (*metG*$_{WT}$, *metG*$_{PP}$, and background) were
831 inoculated into TSB and incubated overnight with 200 rpm shaking at 37˚C for 18 hours. After
832 incubation, the cultures were pooled in equal proportions and loaded into the device. Individual
833 strains were identified based on their chromosomally integrated fluorescence protein markers,
834 which produced bright signals in dedicated fluorescence channels. Just before the image
835 acquisition, the two syringes with flow medium (with and without inducer) were prepared, as
836 well as with and without CRO (6x MIC, 0.4µg/ml). These media were connected to designated
837 flow lanes in the microfluidic device and switched during the experiment. Time-lapse multi-
838 colour imaging was conducted to monitor the exit dynamics and post-exit growth and division
839 of the cells. Images were captured using a Nikon Eclipse Ti2 inverted microscope with a
840 Hamamatsu C14440-20UP camera. All images were saved in Nikon's ND2 file format and
841 processed to extract single-cell growth traces using the pipeline described by Wedd et al[51],
842 which uses a combination of synthetic training images and machine-learning models for
843 single-cell segmentation[100]. Cells were tracked across frames using a custom script
844 (https://github.com/erezli/MMLineageTracking). The extracted time-series data were analysed
845 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.

Data availability statement

Nucleotide sequences of pWPMR2 and pWPMR2var obtained using Nanopore sequencing are deposited with accession numbers PQ180110 and PQ521219 (respectively) in GenBank. All metadata for the strains and their individual genome accession numbers are provided in Supplementary Table 1.

Unique Biological material availability statement

Requests to access unique biological materials used in this manuscript will be considered and should be directed to the corresponding author.

Code availability statement

Unique code used to track cells across frames in the microfluidic single cell experiments is available in GitHub (https://github.com/erezli/MMLineageTracking) and the custom python scripts used to extract time-series data such as lag time are available upon request from the corresponding author.

Author contributions

Conceptualization – KSB, JPJH, SB, RL
Data curation – KSB, LC, SCB, CRB, CJ, RL
Formal analysis – CRB, YLT, JPJH, LS, GBB, MDS, CEC, SCB, PR, SB, RL
Funding acquisition – KSB, JPJH, CJ, MDS, LC, SB
Investigation – KSB, LC, YLT, JPJH, LS, SN, MDS, CEC, SCB, PR, SB, RL
Methodology – KSB, LC, YLT, PR, SB. RL
Project administration – KSB, MDS

Resources – KSB, CJ
Supervision – KSB, LC, JPJH, SCB, CJ, SB
Visualisation – KSB, CRB, YLT, GBB, MDS, CEC, LS, JPJH, SB, RL
Writing – original draft – KSB, YLT, JPJH, GBB, MDS, CEC, SB, RL
Writing – review & editing – KSB, LCEM, LC, CRB, YLT, JPJH, SN, MDS, CEC, SCB, PR, CJ, SB, RL, GBB, LS

898 **References**

899
900

901 1. Pulford, C. V. *et al.* Stepwise evolution of *Salmonella Typhimurium* ST313 causing
902 bloodstream infection in Africa. *Nat. Microbiol.* **6**, 327–338 (2021).
903 2. Johnson, J. R., Johnston, B., Clabots, C., Kuskowski, M. A. & Castanheira, M. *Escherichia*
904 *coli* sequence type ST131 as the major cause of serious multidrug-resistant *E. coli*
905 infections in the United States. *Clinical infectious diseases: an official publication of the*
906 *Infectious Diseases Society of America* vol. 51 286–294 (2010).
907 3. Banerjee, R. & Johnson, J. R. A new clone sweeps clean: The enigmatic emergence of
908 *Escherichia coli* sequence type 131. *Antimicrob. Agents Chemother.* **58**, 4997–5004
909 (2014).
910 4. Edquist, P. J., Makitalo, B., Olsson-Liljequist, B., Soderblom, T. & Wisell, K. T.
911 Epidemiology extended-spectrum beta-lactamase-producing *Escherichia coli* Sweden
912 2007–2011. *Clin Microbiol Infect* **20**, O344–O352 (2007).
913 5. Whitmer, G. R., Moorthy, G. & Arshad, M. The pandemic *Escherichia coli* sequence type
914 131 strain is acquired even in the absence of antibiotic exposure. *PLoS Pathog.* **15**,
915 e1008162 (2019).
916 6. García-Fernández, A. *et al. Klebsiella pneumoniae* ST258 producing KPC-3 identified in
917 italy carries novel plasmids and OmpK36/OmpK35 porin variants. *Antimicrob. Agents*
918 *Chemother.* **56**, 2143–2145 (2012).
919 7. Thompson, C. N., Duy, P. T. & Baker, S. The rising dominance of *Shigella sonnei:* An
920 intercontinental shift in the etiology of bacillary dysentery. *PLoS Negl. Trop. Dis.* **9**,
921 e0003708 (2015).
922 8. Bennett, R J ,De Silva, P. M. , Bengtsson, R. J. , Horsburgh, M. J. , Blower, T. R. and
923 Baker, K. S. Temporal GWAS identifies widely distributed putative adhesin contributing
924 pathogen success *Shigella* spp. *bioRxivorg* (2022) doi:10.1101/2022.08.23.504947.
925 9. Mason, L. C. E. *et al.* The evolution and international spread of extensively drug resistant
926 *Shigella sonnei. Nat. Commun.* **14**, 1983 (2023).

927 10. Baker, K. S. *et al.* Intercontinental dissemination of azithromycin-resistant shigellosis
928 through sexual transmission: a cross-sectional study. *Lancet Infect. Dis.* **15**, 913–921
929 (2015).
930 11. Mitchell, H. D. *et al.* Linkage of whole genome sequencing, epidemiological, and clinical
931 data to understand the genetic diversity and clinical outcomes of *Shigella flexneri* among
932 men who have sex with men in England. *Microbiol. Spectr.* **9**, e0121321 (2021).
933 12. Mitchell, H. D. *et al.* Use of whole-genome sequencing to identify clusters of *Shigella*
934 *flexneri* associated with sexual transmission in men who have sex with men in England:
935 a validation study using linked behavioural data. *Microb. Genom.* **5**, e000311 (2019).
936 13. Lefèvre, S. *et al.* Rapid emergence of extensively drug-resistant *Shigella sonnei* in
937 France. *Nat. Commun.* **14**, 462 (2023).
938 14. Baker, K. S. *et al.* Genomic epidemiology of *Shigella* in the United Kingdom shows
939 transmission of pathogen sublineages and determinants of antimicrobial resistance. *Sci.*
940 *Rep.* **8**, 7389 (2018).
941 15. Hoffmann, C. *et al.* High rates of quinolone-resistant strains of *Shigella sonnei* in HIV-
942 infected MSM. *Infection* **41**, 999–1003 (2013).
943 16. Gilbart, V. L. *et al.* Sex, drugs and smart phone applications: findings from semistructured
944 interviews with men who have sex with men diagnosed with *Shigella flexneri* 3a in
945 England and Wales. *Sex. Transm. Infect.* **91**, 598–602 (2015).
946 17. Baker, K. S. *et al.* Horizontal antimicrobial resistance transfer drives epidemics of multiple
947 *Shigella* species. *Nat. Commun.* **9**, 1462 (2018).
948 18. Ingle, D. J. *et al.* Co-circulation of multidrug-resistant *Shigella* among men who have sex
949 with men in Australia. *Clin. Infect. Dis.* **69**, 1535–1544 (2019).
950 19. Olesen, S. W., Lipsitch, M. & Grad, Y. H. The role of ‘spillover’ in antibiotic resistance.
951 *Proc. Natl. Acad. Sci. U. S. A.* **117**, 29063–29068 (2020).
952 20. Mason, L. C. E. *et al.* Using demographics of patients to inform treatment of shigellosis in
953 England. *Lancet Microbe* 101026 (2024).

954 21. Worley, J. N. *et al.* Genomic drivers of multidrug-resistant *Shigella* affecting vulnerable
955 patient populations in the United States and abroad. *MBio* **12**, e03188-20 (2021).
956 22. Bardsley, M. *et al.* Persistent transmission of shigellosis in England is associated with a
957 recently emerged multidrug-resistant strain of *Shigella sonnei*. *J. Clin. Microbiol.* **58**,
958 e01692-19 (2020).
959 23. Hawkey, J. *et al.* Global population structure and genotyping framework for genomic
960 surveillance of the major dysentery pathogen, *Shigella sonnei*. *Nat. Commun.* **12**, 2684
961 (2021).
962 24. Collins, C. & Didelot, X. A phylogenetic method to perform genome-wide association
963 studies in microbes that accounts for population structure and recombination. *PLoS*
964 *Comput. Biol.* **14**, e1005958 (2018).
965 25. Fridman, O., Goldberg, A., Ronin, I., Shoresh, N. & Balaban, N. Q. Optimization of lag
966 time underlies antibiotic tolerance in evolved bacterial populations. *Nature* **513**, 418–421
967 (2014).
968 26. Blattman, S. B. *et al.* Identification and genetic dissection of convergent persister cell
969 states. *Nature* **636**, 438–446 (2024).
970 27. Khare, A. & Tavazoie, S. Extreme antibiotic persistence via heterogeneity-generating
971 mutations targeting translation. *mSystems* **5**, 00847–00819 (2020).
972 28. Girgis, H. S., Harris, K. & Tavazoie, S. Large mutational target size for rapid emergence
973 of bacterial persistence. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 12740–12745 (2012).
974 29. Levin-Reisman, I. *et al.* Antibiotic tolerance facilitates the evolution of resistance. *Science*
975 **355**, 826–830 (2017).
976 30. Azam, A.H., Chihara, K., Kondo, K., Nakamura, T., Ojima, S., Tamura, A., Yamashita,
977 W., Cui, L., Takahashi, Y., Watashi, K., & Kiga, K. Viruses encode tRNA and anti-retron
978 to evade bacterial immunity. *bioRxivorg* (2023) doi:10.1101/2023.03.15.532788.
979 31. Wein, T. *et al.* Essential gene acquisition destabilizes plasmid inheritance. *PLoS Genet.*
980 **17**, e1009656 (2021).

981 32. Lozano-Terol, G. *et al.* Relative impact of three growth conditions on the *Escherichia coli*
982 protein acetylome. *iScience* **27**, 109017 (2024).
983 33. Kremer, M. *et al.* Bacteria employ lysine acetylation of transcriptional regulators to adapt
984 gene expression to cellular metabolism. *Nat. Commun.* **15**, 1674 (2024).
985 34. Baker, K. S., Dallman, T. J., Thomson, N. R. & Jenkins, C. An outbreak of a rare Shiga-
986 toxin-producing *Escherichia coli* serotype (O117:H7) among men who have sex with men.
987 *Microb. Genom.* **4**, e000181 (2018).
988 35. Ahmed, S. A. *et al.* Genomic comparison of *Escherichia coli* O104:H4 isolates from 2009
989 and 2011 reveals plasmid, and prophage heterogeneity, including shiga toxin encoding
990 phage stx2. *PLoS One* **7**, e48228 (2012).
991 36. Stephens, C. *et al.* F plasmids are the major carriers of antibiotic resistance genes in
992 human-associated commensal *Escherichia coli. mSphere* **5**, e00709-20 (2020).
993 37. Pfeifer, E., Moura de Sousa, J. A., Touchon, M. & Rocha, E. P. C. Bacteria have
994 numerous distinctive groups of phage-plasmids with conserved phage and variable
995 plasmid gene repertoires. *Nucleic Acids Res.* **49**, 2655–2673 (2021).
996 38. Thorley, K. *et al.* Emergence of extensively drug-resistant and multidrug-resistant *Shigella*
997 *flexneri* serotype 2a associated with sexual transmission among gay, bisexual, and other
998 men who have sex with men, in England: a descriptive epidemiological study. *Lancet*
999 *Infect. Dis.* **23**, 732–739 (2023).
1000 39. Schmartz, G. P. *et al.* PLSDB: advancing a comprehensive database of bacterial
1001 plasmids. *Nucleic Acids Res.* **50**, D273–D278 (2022).
1002 40. diCenzo, G. C., Mengoni, A. & Perrin, E. Chromids aid genome expansion and functional
1003 diversification in the family Burkholderiaceae. *Mol. Biol. Evol.* **36**, 562–574 (2019).
1004 41. Clough, S. E., Elphinstone, J. G. & Friman, V.-P. Plant pathogenic bacterium *Ralstonia*
1005 *solanacearum* can rapidly evolve tolerance to antimicrobials produced by *Pseudomonas*
1006 biocontrol bacteria. *J. Evol. Biol.* **37**, 225–237 (2024).
1007 42. Nair, S. *et al.* Presence of phage-plasmids in multiple serovars of *Salmonella enterica*.
1008 *Microb. Genom.* **10**, 001247 (2024).

1009 43. Pfeifer, E. & Rocha, E. P. C. Phage-plasmids promote recombination and emergence of
1010 phages and plasmids. *Nat. Commun.* **15**, 1545 (2024).
1011 44. Silva, C., Calva, E., Fernández-Mora, M., Puente, J. L. & Vinuesa, P. Population analysis
1012 of D6-like plasmid prophage variants associated with specific IncC plasmid types in the
1013 emerging *Salmonella Typhimurium* ST213 genotype. *PLoS One* **14**, e0223975 (2019).
1014 45. Levin-Reisman, I., Brauner, A., Ronin, I. & Balaban, N. Q. Epistasis between antibiotic
1015 tolerance, persistence, and resistance mutations. *Proc. Natl. Acad. Sci. U. S. A.* **116**,
1016 14734–14739 (2019).
1017 46. Sandman, Z. & Iqbal., O. A. *Azithromycin*. (Treasure Island (FL): StatPearls Publishing,
1018 2025).
1019 47. Pollock, A. A. *et al.* Pharmacokinetic characteristics of intravenous ceftriaxone in normal
1020 adults. *Antimicrob. Agents Chemother.* **22**, 816–823 (1982).
1021 48. Bignell, C., Fitzgerald, M., Guideline Development Group & British Association for Sexual
1022 Health and HIV UK. UK national guideline for the management of gonorrhoea in adults,
1023 2011. *Int. J. STD AIDS* **22**, 541–547 (2011).
1024 49. Nwokolo, N. C. *et al.* 2015 UK national guideline for the management of infection with
1025 *Chlamydia trachomatis*. *Int. J. STD AIDS* **27**, 251–267 (2016).
1026 50. Zaslaver, A. *et al.* A comprehensive library of fluorescent transcriptional reporters for
1027 *Escherichia coli*. *Nat. Methods* **3**, 623–628 (2006).
1028 51. Wedd, C. *et al.* Single-cell imaging of the lytic phage life cycle in bacteria. *bioRxiv* (2024)
1029 doi:10.1101/2024.04.11.588870.
1030 52. Stefanovic, A. *et al.* Increased severity of multidrug-resistant *Shigella sonnei* infections in
1031 people experiencing homelessness. *Clin. Infect. Dis.* **80**, 339–346 (2025).
1032 53. Vecilla, D. F. *et al.* First report of *Shigella sonnei* carrying a blaCTX-M-15 sexually
1033 transmitted among men who have sex with men. *Infection* **53**, 443–448 (2025).
1034 54. Charles H., Greig D.R., Swift C., Olonade I., Simms I., Sinka K., Baker K.S., Godbole G.,
1035 Jenkins C. Outbreak of sexually transmitted *S. sonnei* $bla_{\text{CTX-M-15}}$ in England: an

1036 epidemiological and genomic investigation. *medRxiv* (2024)
1037 doi:10.1101/2024.10.14.24314996.
1038 55. Ingle, D. J. *et al.* Informal genomic surveillance of regional distribution of *Salmonella Typhi*
1039 genotypes and antimicrobial resistance via returning travellers. *PLoS Negl. Trop. Dis.* **13**,
1040 e0007620 (2019).
1041 56. Pfeifer, E., Bonnin, R. A. & Rocha, E. P. C. Phage-plasmids spread antibiotic resistance
1042 genes through infection and lysogenic conversion. *MBio* **13**, e0185122 (2022).
1043 57. Fifer, H. *et al.* British Association of Sexual Health and HIV UK National Guideline for the
1044 Management of infection with *Neisseria gonorrhoeae*, 2025. *Int. J. STD AIDS*
1045 9564624251345196 (2025).
1046 58. Allen, H. *et al.* Evidence for re-infection and persistent carriage of *Shigella* species in adult
1047 males reporting domestically acquired infection in England. *Clin. Microbiol. Infect.* **27**,
1048 126.e7-126.e13 (2021).
1049 59. Bengtsson, R. J. *et al.* Accessory genome dynamics and structural variation of *Shigella*
1050 from persistent infections. *MBio* **12**, e00254-21 (2021).
1051 60. van den Berg, D. F., van der Steen, B. A., Costa, A. R. & Brouns, S. J. J. Phage tRNAs
1052 evade tRNA-targeting host defenses through anticodon loop mutations. *Elife* **12**, e85183
1053 (2023).
1054 61. Charles, H. *et al.* Trends in shigellosis notifications in England, January 2016 to March
1055 2023. *Epidemiol. Infect.* **152**, e115 (2024).
1056 62. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina
1057 sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
1058 63. Liu, Y., Schröder, J. & Schmidt, B. Musket: a multistage k-mer spectrum-based error
1059 corrector for Illumina sequence data. *Bioinformatics* **29**, 308–315 (2013).
1060 64. Croucher, N. J. *et al.* Rapid phylogenetic analysis of large samples of recombinant
1061 bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* **43**, e15 (2015).
1062 65. Minh, B. Q. *et al.* IQ-TREE 2: New models and efficient methods for phylogenetic
1063 inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).

1064 66. Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model
1065 choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
1066 67. Heritage, S. MBASR: Workflow-simplified ancestral state reconstruction of discrete traits
1067 with MrBayes in the R environment. *bioRxiv* (2021) doi:10.1101/2021.01.10.426107.
1068 68. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
1069 *arXiv [q-bio.GN]* (2013).
1070 69. Chen, S. Ultrafast one-pass FASTQ data preprocessing, quality control, and
1071 deduplication using fastp. *Imeta* **2**, e107 (2023).
1072 70. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods*
1073 **9**, 357–359 (2012).
1074 71. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008
1075 (2021).
1076 72. Marco-Sola, S., Sammeth, M., Guigó, R. & Ribeca, P. The GEM mapper: fast, accurate
1077 and versatile alignment by filtration. *Nat. Methods* **9**, 1185–1188 (2012).
1078 73. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads
1079 using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
1080 74. Wick, R. R., Schultz, M. B., Zobel, J. & Holt, K. E. Bandage: interactive visualization of de
1081 novo genome assemblies. *Bioinformatics* **31**, 3350–3352 (2015).
1082 75. De Coster, W. & Rademakers, R. NanoPack2: population-scale evaluation of long-read
1083 sequencing data. *Bioinformatics* **39**, btad311 (2023).
1084 76. De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M. & Van Broeckhoven, C. NanoPack:
1085 visualizing and processing long-read sequencing data. *Bioinformatics* **34**, 2666–2669
1086 (2018).
1087 77. Wishart, D. S. *et al.* PHASTEST: faster than PHASTER, better than PHAST. *Nucleic*
1088 *Acids Res.* **51**, W443–W450 (2023).
1089 78. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421
1090 (2009).

1091 79. Carattoli, A. & Hasman, H. PlasmidFinder and in silico pMLST: Identification and typing
1092 of Plasmid replicons in whole-genome sequencing (WGS). *Methods Mol. Biol.* **2075**, 285–
1093 294 (2020).
1094 80. Feldgarden, M. *et al.* AMRFinderPlus and the Reference Gene Catalog facilitate
1095 examination of the genomic links among antimicrobial resistance, stress response, and
1096 virulence. *Sci. Rep.* **11**, 12728 (2021).
1097 81. Tesson, F. *et al.* Systematic and quantitative view of the antiviral arsenal of prokaryotes.
1098 *Nat. Commun.* **13**, 2561 (2022).
1099 82. Tesson, F. *et al.* Exploring the diversity of anti-defense systems across prokaryotes,
1100 phages and mobile genetic elements. *Nucleic Acids Res.* **53**, 13 (2024).
1101 83. Nishimura, Y. *et al.* ViPTree: the viral proteomic tree server. *Bioinformatics* **33**, 2379–
1102 2380 (2017).
1103 84. Mihara, T. *et al.* Linking virus genomes with host taxonomy. *Viruses* **8**, 66 (2016).
1104 85. Tonkin-Hill, G. *et al.* Producing polished prokaryotic pangenomes with the Panaroo
1105 pipeline. *Genome Biol.* **21**, 180 (2020).
1106 86. Schwengers, O. *et al.* Bakta: rapid and standardized annotation of bacterial genomes via
1107 alignment-free sequence identification. *Microb. Genom.* **7**, 000685 (2021).
1108 87. Sullivan, M. J., Petty, N. K. & Beatson, S. A. Easyfig: a genome comparison visualizer.
1109 *Bioinformatics* **27**, 1009–1010 (2011).
1110 88. Carver, T. J. *et al.* ACT: The Artemis Comparison Tool. *Bioinformatics* **21**, 3422–3423
1111 (2005).
1112 89. Hunt, M., Lima, L., Shen, W., Lees, J. & Iqbal, Z. AllTheBacteria - all bacterial genomes
1113 assembled, available and searchable. (2024) doi:10.1101/2024.03.08.584059.
1114 90. Shen, W. & Iqbal, Z. LexicMap: efficient sequence alignment against millions of
1115 prokaryotic genomes. *bioRxiv* (2024) doi:10.1101/2024.08.30.610459.
1116 91. Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q
1117 file manipulation. *PLoS One* **11**, e0163962 (2016).

1118 92. Lassmann, T. Kalign 3: multiple sequence alignment of large data sets. *Bioinformatics*
1119 **36**, 1928–1929 (2019).
1120 93. Page, A. J. *et al.* SNP-sites: rapid efficient extraction of SNPs from multi-FASTA
1121 alignments. *Microb. Genom.* **2**, e000056 (2016).
1122 94. Wickham, H. *et al.* Welcome to the tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
1123 95. Cook, R. *et al.* INfrastructure for a PHAge REference Database: Identification of large-
1124 scale biases in the current collection of cultured phage genomes. *Phage (New Rochelle)*
1125 **2**, 214–223 (2021).
1126 96. Bakshi, S. *et al.* Tracking bacterial lineages in complex and dynamic environments with
1127 applications for growth control and persistence. *Nat. Microbiol.* **6**, 783–791 (2021).
1128 97. Blazanin, M. gcplyr: an R package for microbial growth curve data analysis. *BMC*
1129 *Bioinformatics* **25**, 232 (2024).
1130 98. Smug, B. J., Opalek, M., Necki, M. & Wloch-Salamon, D. Microbial lag calculator: A shiny-
1131 based application and an R package for calculating the duration of microbial lag phase.
1132 *Methods Ecol. Evol.* **15**, 301–307 (2024).
1133 99. Wickham, H. Ggplot2. *Wiley Interdiscip. Rev. Comput. Stat.* **3**, 180–185 (2011).
1134 100. Hardo, G., Noka, M. & Bakshi, S. Synthetic Micrographs of Bacteria (SyMBac) allows
1135 accurate segmentation of bacterial cells using deep neural networks. *BMC Biol.* **20**, 263
1136 (2022).