

# Clinically relevant antibiotic resistance genes are linked to a limited set of taxa within gut microbiome worldwide

---

Received: 10 August 2023

---

Accepted: 27 October 2023

---

Published online: 14 November 2023

---

 Check for updates

---

A list of authors and their affiliations appears at the end of the paper

The acquisition of antimicrobial resistance (AR) genes has rendered important pathogens nearly or fully unresponsive to antibiotics. It has been suggested that pathogens acquire AR traits from the gut microbiota, which collectively serve as a global reservoir for AR genes conferring resistance to all classes of antibiotics. However, only a subset of AR genes confers resistance to clinically relevant antibiotics, and, although these AR gene profiles are well-characterized for common pathogens, less is known about their taxonomic associations and transfer potential within diverse members of the gut microbiota. We examined a collection of 14,850 human metagenomes and 1666 environmental metagenomes from 33 countries, in addition to nearly 600,000 isolate genomes, to gain insight into the global prevalence and taxonomic range of clinically relevant AR genes. We find that several of the most concerning AR genes, such as those encoding the cephalosporinase *CTX-M* and carbapenemases *KPC*, *IMP*, *NDM*, and *VIM*, remain taxonomically restricted to Proteobacteria. Even *cfiA*, the most common carbapenemase gene within the human gut microbiome, remains tightly restricted to *Bacteroides*, despite being found on a mobilizable plasmid. We confirmed these findings in gut microbiome samples from India, Honduras, Pakistan, and Vietnam, using a high-sensitivity single-cell fusion PCR approach. Focusing on a set of genes encoding carbapenemases and cephalosporinases, thus far restricted to *Bacteroides* species, we find that few mutations are required for efficacy in a different phylum, raising the question of why these genes have not spread more widely. Overall, these data suggest that globally prevalent, clinically relevant AR genes have not yet established themselves across diverse commensal gut microbiota.

Preventing the spread of multidrug- and pandrug-resistant pathogenic bacteria remains a primary focus of global health efforts<sup>1</sup>. Recognition that the commensal gut microbiota harbor extensive numbers of diverse AR genes<sup>2–4</sup>, engage in horizontal gene transfer (HGT) at higher rates than microbiota in other environments<sup>2,5,6</sup>, and may serve as a stable reservoir for pathogenic acquisition<sup>5,7,8</sup> has prompted broader AR surveillance beyond clinical isolates<sup>9–13</sup>. The potential for the spread

of AR genes between taxa through HGT<sup>14–17</sup> has raised further concern, especially for *mcr* genes, which confer resistance against colistin, and carbapenemases. However, many of the annotated AR genes observed to transfer within microbiomes, including many common Class A beta-lactamase variants<sup>18</sup>, do not confer phenotypic resistance to clinically relevant antibiotics. Understanding the global spread and HGT potential of clinically relevant AR genes—particularly those associated

---

 e-mail: [ibrto@cornell.edu](mailto:ibrto@cornell.edu)

with ‘last resort’ antibiotics—would be helpful, yet surprisingly little is known about the gene-taxa associations within microbiomes at large.

Here, using a combined analysis of 14,229 human gut metagenomes and nearly 600,000 isolate genomes, we find that certain AR genes are prevalent throughout global populations’ microbiomes; however, thus far associated with few taxa. To determine whether there may be any additional bacterial hosts for twelve such clinically relevant genes, we applied a single-cell fusion PCR method to associate the genes with the 16S rRNA sequence of the host. We find that these genes are harbored by the set of taxa in which they were previously observed, despite being mobilizable, and, in some instances, functional in taxonomically distant organisms. This study challenges the notion that AR genes have spread more broadly within the gut microbiome than currently realized.

## Results

### Many clinically relevant AR genes are found in a few isolated taxa

To define the established taxonomic breadth of clinically relevant AR genes and their distribution within human-associated microbiota, we first examined the global prevalence of 46 known AR gene families in 14,229 human gut microbiome samples, as well as an additional 248 oral and 261 skin microbiome samples spanning 33 countries compiled from the curated MetagenomicData package<sup>19</sup> (Supplementary Data 1–3). We then coupled this with a taxonomic analysis of AR genes within nearly 600,000 sequenced genome isolates from GenBank (Supplementary Data 4, 5). The 46 clinically relevant gene families selected confer resistance to 18 classes of antibiotics, defined by the Comprehensive Antibiotic Resistance Database (CARD)<sup>20</sup> (Supplementary Data 6), and largely overlap with the World Health Organization’s List of Essential Medicines<sup>21</sup> (Fig. 1, all gene families shown in Supplementary Fig. 1). We examined the prevalence of all identifiable AR markers in CARD, excluding efflux pumps, as their role in antibiotic resistance versus other biological functions may be difficult to annotate and confirm<sup>22</sup> and they are less frequently mobilized<sup>23</sup>. We further excluded metabolic genes recently shown to confer resistance<sup>24</sup>, due to the difficulties of distinguishing their physiologic role from their role in AR.

Notwithstanding small differences in read depth, AR gene detection method, or individuals’ health status (Supplementary Figs. 2–4), AR gene families exhibited a relatively even geographic distribution across continents but differed between the gut microbiomes of Western and non-Western populations, as defined by Pasolli et al.<sup>25</sup> (Supplementary Fig. 5). This split has been observed in other potential AR gene reservoirs, such as sewage, and bacterial isolate sources and may be explained by socio-economic factors<sup>26–29</sup>. AR gene families also exhibited a bimodal distribution, such that they were either detected in many countries or very few (Supplementary Fig. 6). Gene-specific prevalence in the oral and skin microbiomes mirrored, albeit with reduced prevalence, those found within the gut (Fig. 1). Together, these findings concretely illustrate the reality that AR is a “global commons” problem.

We next assessed the host ranges of different AR gene families in isolate genomes. Here, certain taxa tend to harbor a disproportionately higher number of genes (Fig. 1). The Gammaproteobacteria, Betaproteobacteria, and Bacilli are among those taxa, though these organisms tend to be over-represented, as they comprise pathogenic strains, several representing prominent AR threats, and have thus been surveyed to a greater extent. Actinomycetia, which includes *Streptomyces*—known to be a source of AR genes—also harbor many AR gene families.

It may be expected that gene families with greater taxonomic range are generally more prevalent globally. That is the case (Supplementary Fig. 7, two-sided Pearson’s  $\rho = 0.558$ ,  $p$ -value = 0.000896), however, unexpectedly, a subset of AR gene families was found to be

restricted to a single taxonomic class (e.g., NDM and *CMY*) (Fig. 1, Supplementary Fig. 1). This restricted distribution was particularly surprising given their association with plasmids<sup>30,31</sup>, which we confirmed with our analysis of AR gene-plasmid associations (Fig. 1). Two cephalosporinase genes *cepA* and *cblA*, as well as the carbapenemase gene *cfiA*, were similarly found to be confined to the genus *Bacteroides*, despite this being the most abundant and prevalent genus in Western countries<sup>32</sup>, theoretically providing ample opportunity to transfer.

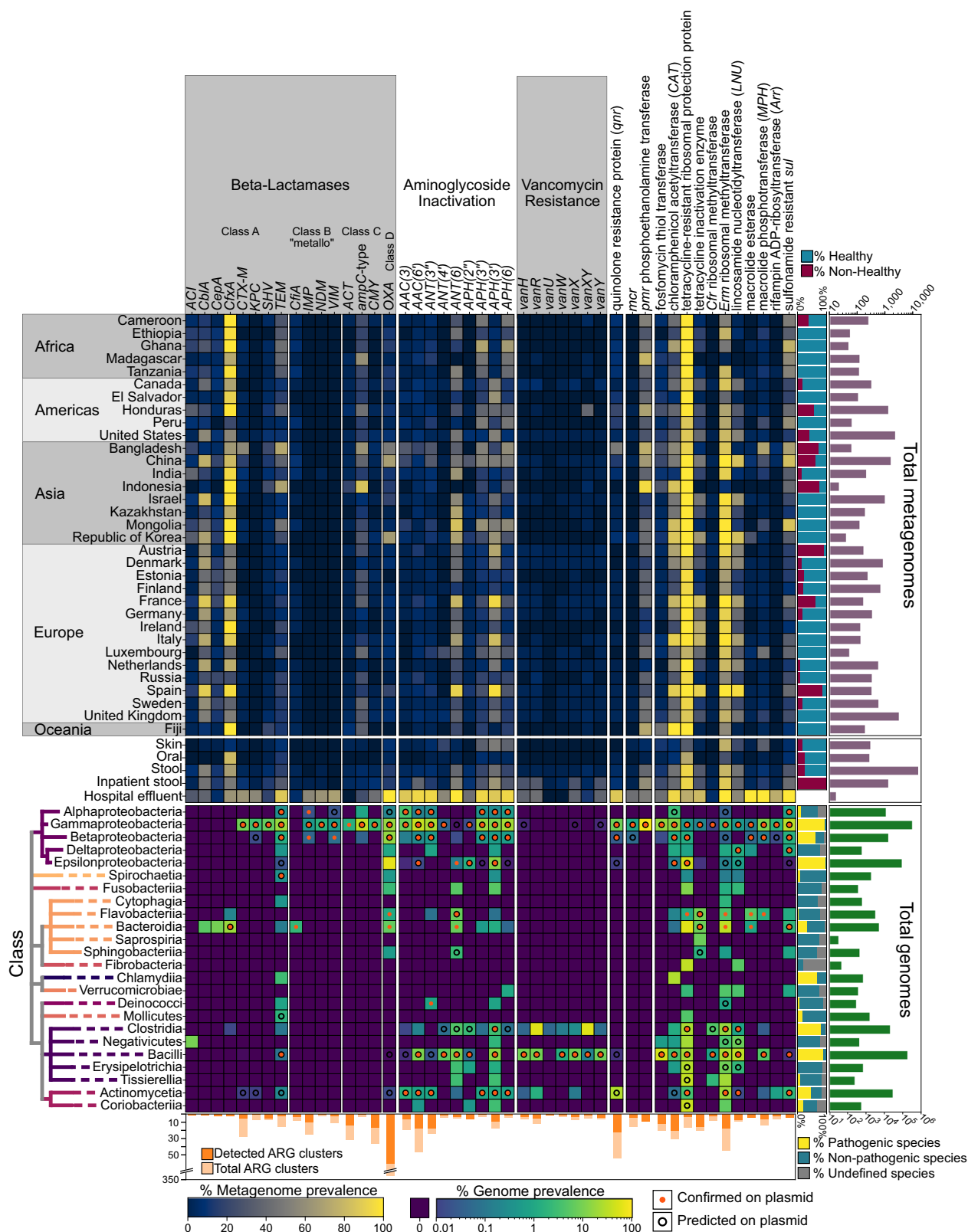
### Factors related to AR gene prevalence and taxonomic host range

We next sought to determine the selective pressure imposed by antibiotic use on global AR geographic distribution and taxonomic spread. Antibiotic use on an individual level is generally associated with an increase in AR gene abundance in gut microbiomes<sup>11,33,34</sup>. Historical data on global antibiotic use however are sparse, due to the lack of centralized reporting that accounts for human and animal consumption, and environmental use. Nevertheless, recent data from the World Health Organization (WHO) representing 65 countries<sup>35</sup> afford the opportunity to examine global use of oral and parenteral use of antibiotics with AR gene prevalence in human gut microbiomes.

We expected higher global antibiotic use to correlate with greater AR gene prevalence in microbiomes. Yet, this was not the case. Beta-lactams, including penicillins, categorized as Access drugs, under the WHO’s suggested use categories of Access, Watch and Reserve, were the single most consumed category of antibiotics worldwide ( $36.7\% \pm 12.25$  of total consumption per country)<sup>35</sup>. Despite the copious use of broad-spectrum beta-lactams, only two beta-lactamase genes, *cfxA* and *cblA*, were found to be highly prevalent in human gut microbiomes (Fig. 1). This was unexpected, as beta-lactamases are diverse, and many, even those that are most concerning, confer resistance to multiple beta-lactams, including amoxicillin, one of the most widely consumed drugs<sup>35</sup>.

The most striking observation was a general lack of carbapenemases in gut microbiomes. Watch-group beta-lactams, including carbapenems, penems, monobactams, and third-, fourth-, and fifth-generation cephalosporins, comprised the second most consumed category of antibiotics ( $13.99\% \pm 8.05$  of total use per country). Aside from *cfiA*, which has surprisingly high prevalence, we identified only 8 individuals out of 14,229 harboring any other carbapenemase in their gut microbiomes (Supplementary Data 2, 3). This finding contrasts with reports of nosocomial spread of these genes<sup>36,37</sup>. For example, another carbapenemase, *NDM-1*, was first identified in a clinical *Klebsiella pneumoniae* isolate from an inpatient who was simultaneously colonized with commensal *Escherichia coli* carrying the gene<sup>38</sup>. Despite this potential reservoir, the gene family *NDM* was observed in only 3 samples worldwide. Similarly, *OXA-48*, the only gene in the *OXA* family with significant carbapenemase activity<sup>39,40</sup> (which, unlike other phylogeny-based AR gene families, is based on phenotypic resistance to oxacillin alone), was identified in only 5 metagenomes. By contrast, chloramphenicol-resistance genes were found to be pervasive across global metagenomes, even though chloramphenicol has long been considered a last-resort antibiotic due to toxicity and has relatively low usage ( $0.29\% \pm 0.68$  of total use per country)<sup>35</sup>.

Since the majority of the metagenomic samples we analyzed were from outpatient cohorts (Supplementary Fig. 4), we wanted to directly compare the AR gene profiles in these samples with those in other suspected reservoirs, including gut metagenomes obtained from inpatient cohorts, hospital effluent, wastewater treatment plants, and livestock<sup>41–55</sup> (Supplementary Data 7–9). Hospital effluent is the only suspected reservoir that was strongly enriched for clinically relevant AR genes, including *mcr* and *CTX-M*, and the carbapenemase-encoding genes *KPC*, *NDM*, *VIM*, and *IMP* (Fig. 1, Supplementary Fig. 1). We hypothesize that the enrichment of carbapenemases in hospital effluent is



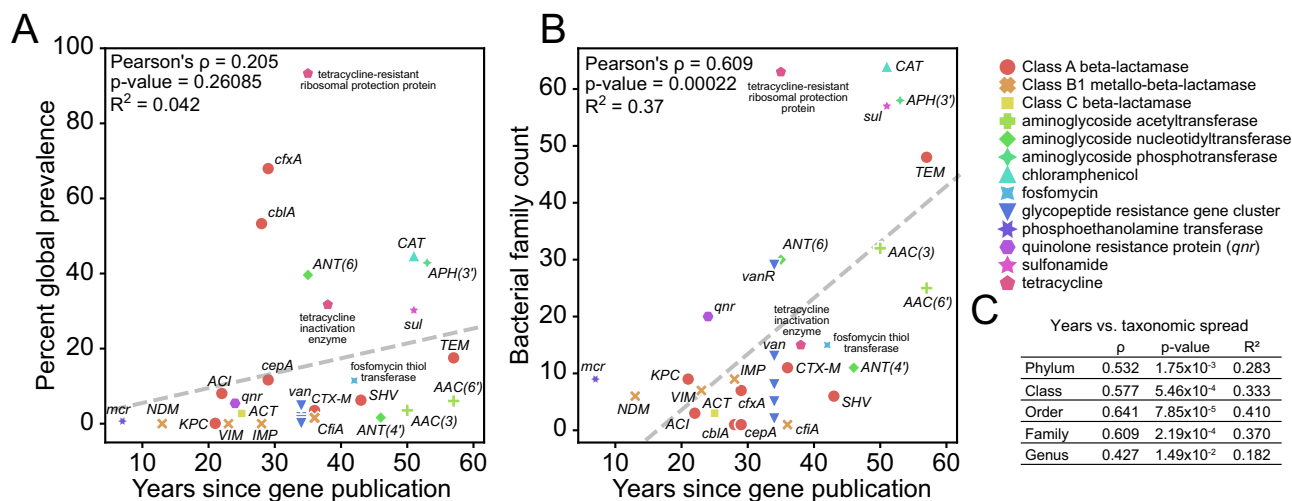
due to aerobic conditions selecting for Proteobacteria, which commonly harbor diverse ARG genes.

To address the possibility that ARG gene spread, both geographically and between taxa, may simply be the result of time, we used the first report of an identified ARG gene family in the literature, a previously established metric<sup>56</sup>, as a proxy for the year at which it emerged (Supplementary Data 6). Whereas time since emergence has

had little bearing on the global prevalence of ARG gene families (Fig. 2A), time has allowed for greater taxonomic distribution (two-sided Pearson's  $\rho = 0.609$ ,  $p\text{-value} = 2.2 \times 10^{-4}$ , at the family-level) (Fig. 2B, C). Genes within the *mcr* family serve as a counterexample, having emerged relatively recently<sup>57</sup> and, despite their low prevalence in the human gut microbiome, have rapidly spread to many taxa (21 genera across 2 classes). Although polymyxins, to which *mcr* genes confer

**Fig. 1 | Distribution of clinically relevant AR genes across global human-associated microbiomes globally and cultured isolates.** The top heatmap shows the prevalence of AR gene families in human gut-associated microbiomes sequenced from each country. Immediately below is a heatmap showing the total prevalence in gut, skin and oral microbiomes. The bottom heatmap shows the prevalence of AR gene families in sequenced isolates by taxonomic class. Color scales represent the percent of metagenomes (top) or isolates (bottom) for which the gene family was detected. At right of the top heatmap are two barplots, one showing the proportion of samples from healthy (blue) and not-healthy (red)

individuals, as classified by curatedMetagenomicDataCuration<sup>19</sup>, and a second showing the total number of samples from each country. At right of the bottom heatmap are two barplots, the first showing the proportion of isolates from pathogenic (yellow), non-pathogenic (red), or unannotated species (gray), as determined by CARD, and the second showing the total number of isolates surveyed per class. At the bottom, a barplot shows the total number of AR gene variants in each gene family (light orange) and the number detected in this study (dark orange) in the metagenomic datasets. Nearly all genes in each family shown were identified in the isolate genomes.



**Fig. 2 | Time allows for genes to become more prevalent and spread more broadly.** **A, B** For each gene family, the time since initially reported in the scientific literature, a proxy for AR gene emergence, is plotted against the prevalence in human gut microbiomes (left) and number of genera in which the gene is found

(right). Marker color/shape indicate the AR gene class. Two-sided Pearson's  $\rho$ , p-value and  $R^2$  are reported. **C** Statistics (two-sided Pearson's  $\rho$ , p-value and  $R^2$ ) are reported for the years since AR gene emergence and taxonomic spread computed at different taxonomic levels.

resistance, are infrequently prescribed today, they have historically been used for growth promotion in livestock. Their prevalence may have been limited by the disproportionate fitness defects caused by these genes compared to other AR genes<sup>58</sup>. Rather, their persistence within communities is thought to be stabilized by being situated within transposons on conjugative plasmids that may be maintained by other selective means<sup>59,60</sup>.

### Lack of hidden taxonomic reservoir of clinically relevant AR genes in gut microbiomes

Because the taxonomic diversity of the AR gene reservoir within human gut microbiomes is unknown, we tested whether clinically relevant AR gene family-taxa associations could be explained solely by using a collection of isolates sequenced thus far. To this end, we examined all 14,229 gut metagenomes from curatedMetagenomicData for the presence of at least one known host for each AR gene family. Save for a few AR gene families in a small number of people, isolate-based associations alone were sufficient to explain the distribution of AR gene families (Supplemental Fig. 8), hinting at the lack of hidden AR gene reservoirs. Only *mcr* genes in 6 total individuals lacked previously identified hosts at the phylum level, indicating that either there is an undefined reservoir for *mcr* genes, or known hosts have too low abundances to detect.

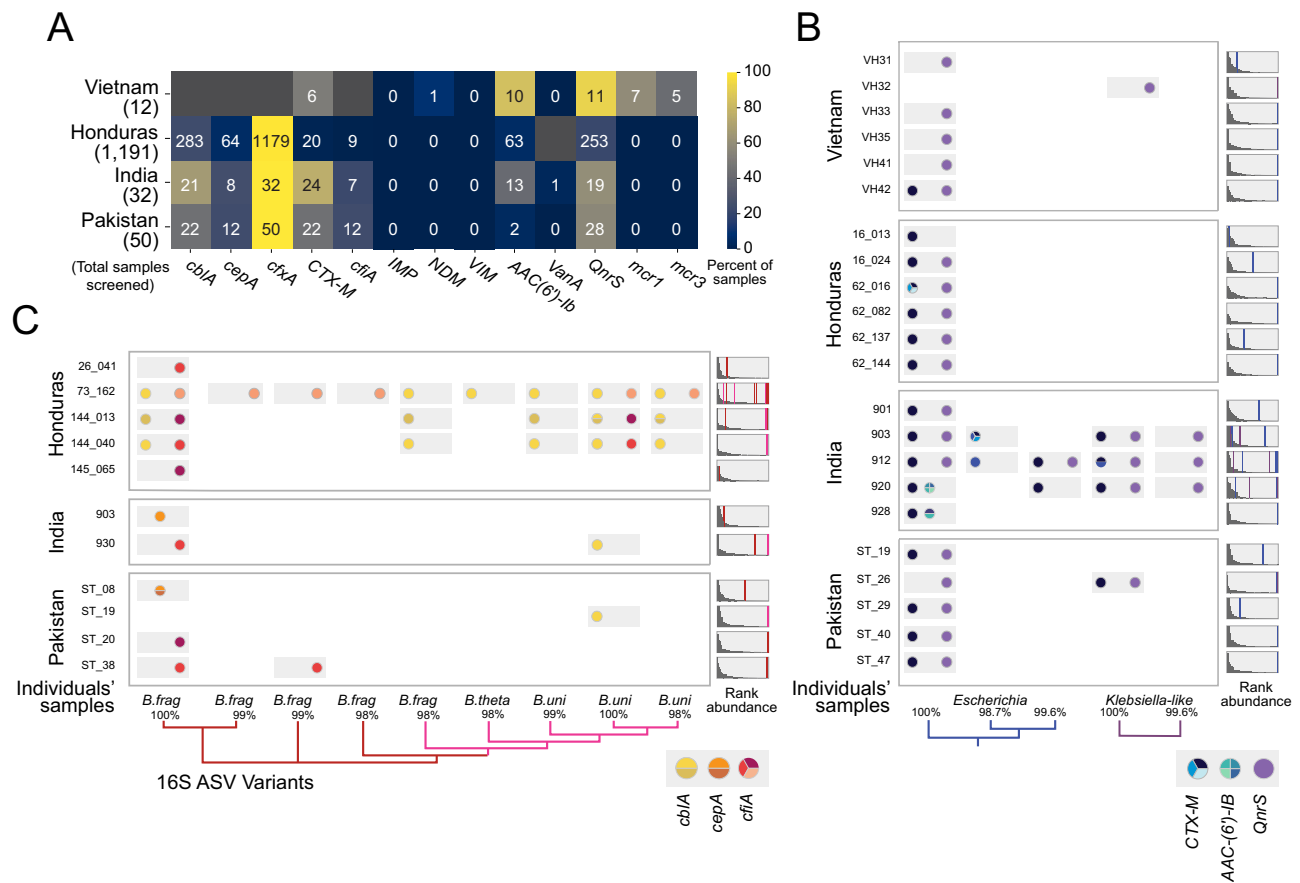
To experimentally elucidate AR gene-taxa associations within gut microbiomes, we applied OIL-PCR (one-step isolation and lysis PCR), a sensitive, culture-independent single-cell fusion PCR approach to associate extrachromosomal DNA with their associated genomes<sup>61</sup>. OIL-PCR overcomes technical limitations which have, until now, obscured our ability to assay the full taxonomic range of these genes. Metagenomic genome assemblies are neither able to robustly associate plasmids nor integrated mobile elements with specific bacterial

hosts. And while culture-based screening can provide a window into AR host range, surveying the entire gut microbiota across many samples may be impractical due to diverse culturing conditions<sup>62-64</sup>.

To identify candidate samples for OIL-PCR, we tested 100 gut microbiome samples from India, Pakistan, and Vietnam for the presence of twelve clinically relevant AR genes of interest (*cbIA*, *cepA*, *cfxA*, *CTX-M*, *cfIA*, *IMP*, *NDM*, *VIM*, *AAC(6)-Ib*, *VanA*, *QnrS*, *mcr1*, and *mcr3*) using a probe-based quantitative PCR approach (Fig. 3A, Supplementary Data 10). We also computationally screened 1187 additional samples from Honduras, for which we had metagenomes, by aligning reads to AR genes (same as in Fig. 1). These results validated the prevalence trends observed in our analysis of global gut microbiomes. We selected thirty-one samples for OIL-PCR and focused on six genes: *CTX-M*, *aac(6)-Ib-cr* and *qnrS*, which have been shown to be mobilizable, and *Bacteroides*-specific *cbIA*, *cepA* and *cfIA* (Fig. 3B, C, Supplementary Data 10, 11). We found that all six genes were restricted to the genera expected based on the isolate genomes and included the presence of multiple amplicon sequence variants (ASVs) within the same sample, each carrying a distinct set of AR gene variants, with relative abundances as low as 0.0044%. We identified several novel associations: *cfIA* with *B. uniformis* and *cbIA* with *Bacteroides thetaotaomicron* and *B. uniformis*, though we were unable to exclude the possibility that these associations resulted from chimeras between highly similar 16S rRNA sequences formed during library preparation. These results thus provide orthogonal experimental support for the absence of a prevalent unculturable commensal reservoir for these genes.

### *Bacteroides*-specific beta-lactamases can be transferred and are functional in Enterobacteriaceae

Commensal beta-lactamases conferring resistance against clinically relevant antibiotics, including *cepA*, *cbIA*, *cfxA* and *cfIA*, have the



**Fig. 3 | OIL-PCR confirms that select AR genes are taxonomically restricted. A** A heatmap showing the number and percentage of gut microbiome samples that were positively screened for each AR gene, by quantitative PCR (for Vietnam, India and Pakistan) or metagenomic read alignment (Honduras). Bolded AR genes were further screened for taxonomic associations by OIL-PCR. Dark gray boxes indicate qPCR reactions that were not performed. **B** OIL-PCR was performed to detect taxonomic associations for *CTX-M*, *AAC(6)-IB*, and *qnrS* in positive gut microbiome samples from the international cohorts. Each row represents an individual's gut microbiome sample. Colors represent AR gene variants. Phylogenies show the

relationships between detected ASVs. The percent similarity is listed. To the right of each sample is a rank-ordered distribution of abundances of the individual ASVs in the individual's gut microbiome, as determined by 16S rRNA sequencing. Lines, colored according to the species designated in the phylogenies, are placed at the rank of the ASVs detected by OIL-PCR in each sample. **C** OIL-PCR results performed in *cfiA*, *cblA*, and *cepA*, displayed in the same manner as in **(B)**. 31 individuals' samples are shown in **(B)** and **(C)**, with two individuals having been tested for the genes in both **(B)** and **(C)**.

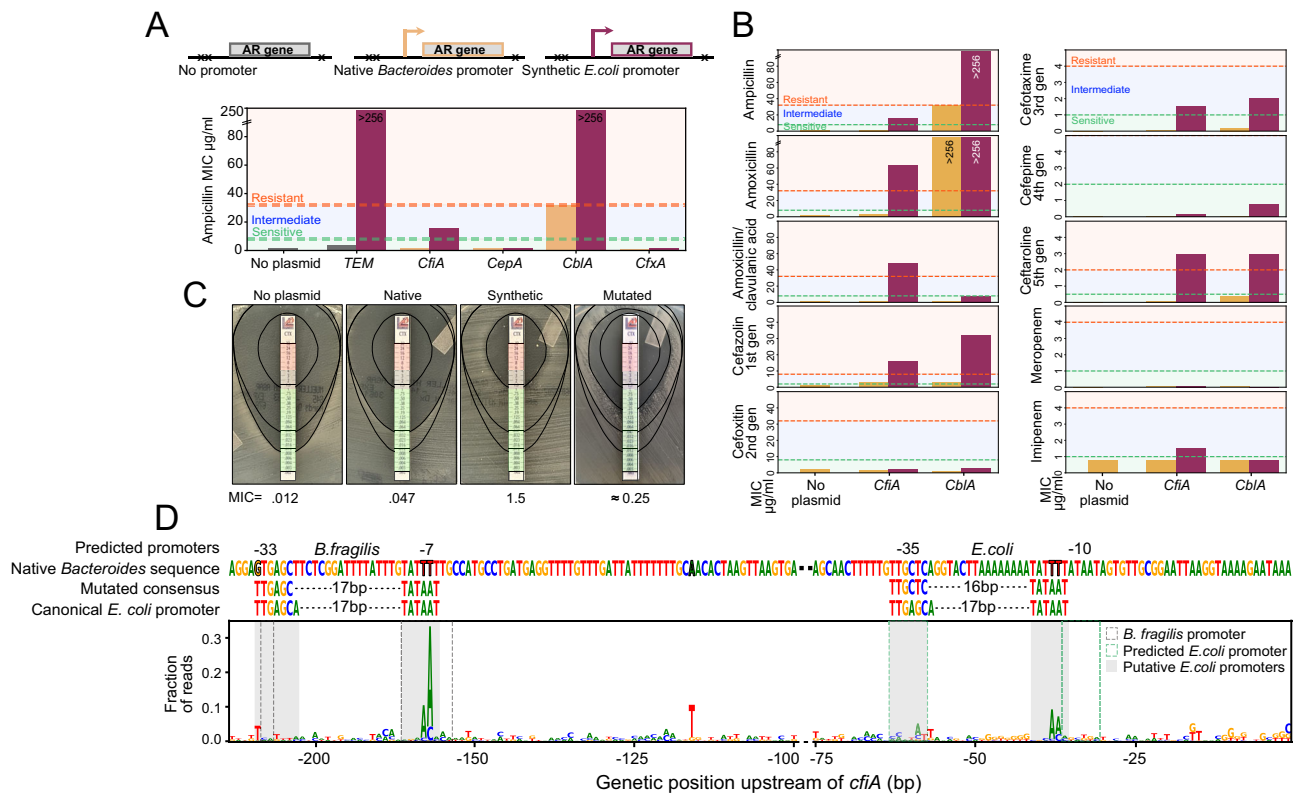
potential to exacerbate AR infections if transferred to pathogens within the Gammaproteobacteria. Aside from *cfxA*, which has been sporadically found in other phyla, these genes are currently restricted to the *Bacteroides*. While inter-phylum HGT is considered a rare event—with few transfers recorded between isolates within the Bacteroidetes and Gammaproteobacteria<sup>5,65</sup>—the high relative abundance of Bacteroidetes within gut microbiomes might be expected to result in higher contact rates with enteric pathogens, and therefore an increased probability of HGT. We therefore sought to explore potential barriers of beta-lactamase HGT beyond *Bacteroides*.

We first investigated if these AR genes could confer a resistance phenotype in *E. coli*. We placed *cepA*, *cblA*, *cfxA*, and *cfiA* under control of either their native promoter or a synthetically designed promoter active in *E. coli*<sup>66,67</sup>. We transformed each construct into the auxotrophic strain DH5 $\alpha$ , for the purpose of biocontainment, and measured the minimum inhibitory concentration (MIC) of each AR gene to ampicillin (Fig. 4A). Both *cblA* and *cfiA* are shown to be functional when expressed by an *E. coli* promoter, yet only *cblA* was also functional under control of its native *Bacteroides* promoter, suggesting that *cblA* could provide immediate selective benefit in the presence of ampicillin. We further found that when adequately expressed, *cblA* and *cfiA* were capable of conferring *E. coli* with resistance to other next-generation cephalosporins (Fig. 4B). *cfiA* also provides resistance to

the beta-lactamase inhibitor clavulanic acid and the carbapenem imipenem, largely matching the literature<sup>68–70</sup>. Furthermore, the expression of these genes in *E. coli* did not result in a significant growth defect, when compared to a wildtype strain or a strain expressing a *TEM* beta-lactamase commonly found in *E. coli* (Supplemental Fig. 9).

We were especially intrigued by *cfiA* because of its clinical relevance and activity in *E. coli*. We investigated the evidence for mobilization of *cfiA*. The mobilizable plasmid on which *cfiA* is situated, pBFUK1, has been shown to transfer from *B. uniformis* to *E. coli*, and vice versa<sup>71–73</sup>. Furthermore, *cfiA* is carried on a composite transposon composed of two IS1380-like elements. While the transposase and inverted repeat sequences could be found in Gammaproteobacteria, a complete transposable element containing both repeats was only found in *Bacteroides*.

We hypothesized that some minor mutations in the native *B. fragilis* *cfiA* promoter would allow its expression in *E. coli*, especially considering the presence of a predicted *E. coli* promoter downstream of the native *B. fragilis* transcriptional start site. We screened a library of randomly mutated *B. fragilis* *cfiA* promoters for increased MIC to cefotaxime, a third-generation cephalosporin, in *E. coli* (Fig. 4C), finding that most mutations occurred either within the *B. fragilis* promoter binding site or within the predicted *E. coli* promoter, thus creating a TATAAT box within proper range of the –35 region with only



**Fig. 4 | *Bacteroides*-specific beta-lactamases CfiA and CblA are functional in *E. coli*.** **A** AR genes TEM, *cfiA*, *cepA*, *cblA*, and *cfxA* were cloned onto plasmids either with no promoter (TEM only), their native *Bacteroides* promoter, or a synthetic *E. coli* promoter, transformed into *E. coli* and tested on ampicillin to determine their minimum inhibitory concentrations (MIC). **B** MICs to various antibiotics were determined using antibiotic strips on agar for *E. coli* carrying *cfiA* and *cblA*, placed under its native *Bacteroides* promoter (yellow) and an *E. coli*-specific promoter (purple). *E. coli* without plasmid were used as a control. **C** *E. coli* tested for growth

on agar plates treated with cefotaxime strips. *E. coli* harboring no plasmid, or a plasmid harboring *cfiA* under control of its native *B. fragilis* promoter, the synthetic *E. coli* promoter, or a mutant library of *B. fragilis* *cfiA* promoters (Mutated). **D** Histogram of mutations within the *B. fragilis* native promoter identified in the mutated *cfiA* promoter library after 18 h of growth in 0.5 μg/ml of cefotaxime. The native *B. fragilis* promoter sequence is shown, as are the predicted *B. fragilis* and *E. coli* promoters in the native sequence, the mutated consensus sequence and a depiction of the canonical *E. coli* promoter.

two mutations (Fig. 4D). This suggests that neither expression, function, nor transfer are the factors limiting the taxonomic spread of the carbapenemase *cfiA*.

## Discussion

Surveillance of AR genes within the human microbiome is grounded on the perceived threat that these genes may be transferred to pathogens. Given the prevalence of AR genes in the gut, the selective pressures imposed by antibiotics, and the observation that the gut environment is conducive to HGT<sup>5,17,62,74–78</sup>, it is expected that AR genes would have been transferred to a wide range of organisms, thereby creating a diverse reservoir of AR genes. While this is the case for some AR genes, others have surprisingly restricted host associations. This is most striking for commensal-associated beta-lactamases, despite their high global prevalence.

While antibiotic stewardship has emerged as a critical strategy to reduce the circulation of multi-drug resistant pathogens, it may be less influential in shaping the global reservoir of AR strains within the human gut. For example, *Bacteroides* often carry aminoglycoside-resistance genes from Proteobacteria, despite their intrinsic resistance to this drug<sup>79</sup>, essentially decoupling antibiotic-mediated selection from use, as also shown by others<sup>80</sup>. This dissociation may be due to the functions of co-localizing genes on AR gene-encoding plasmids<sup>81</sup>, alternative physiologic functions of AR genes<sup>24</sup>, multi-drug efflux pumps that may provide sufficient resistance, and differences in antibiotic concentrations felt across the intestinal microenvironment<sup>82</sup>, among other reasons<sup>83</sup>. Microbial community

effects may also play a major role. For instance, beta-lactamases are secreted<sup>84</sup>, thereby mitigating the effects of antibiotics on susceptible bacteria within the community. Bacteria are similarly protected while situated within a biofilm<sup>85,86</sup>. Changes to community structure also have the potential to affect the landscape of HGT by modulating contact rates between organisms<sup>87</sup>. These examples illustrate some of the ways in which selective pressures imposed by antibiotics may not directly affect sensitive bacteria or bacteria carrying resistance markers.

Although there appears to be limited societal transmission of certain AR strains, nosocomial transmission of these strains may be markedly higher. This, along with increased antibiotic use, may promote the emergence of novel AR strains, especially within the hospital environment. The strong association of time with the taxonomic distribution of AR genes suggests that their emergence and spread within and across gut microbiomes may reflect varying rates of engraftment and persistence<sup>34,88,89</sup>. Curtailing the expanding reservoir of AR strains may therefore require strategies tailored to specific AR gene families. Indeed, fecal microbiota transplantation has been shown to effectively reduce the carriage of vancomycin-resistant *Enterococci*<sup>90</sup>. Phage and gene editing machinery have similarly been shown to reduce the carriage of AR genes<sup>91–93</sup>. Challenges notwithstanding, the geographic and taxonomic spread of AR genes within human microbiomes has grown with time, highlighting the need for continued surveillance of isolates and microbiomes and more explicit studies on the factors that allow, or restrict, certain AR genes from establishing a global reservoir within commensal gut microbiomes.

## Methods

### Public metagenome acquisition

The curatedMetagenomicData datasets package<sup>25</sup> was used to compile a metadata table for all metagenomes included in the Waldron dataset. (Supplementary Data 1). Samples were filtered for at least 5 million reads. Fastq files were downloaded from EBI using NCBI accession numbers acquired from the metadata to query FTP links (wget -qO- "https://www.ebi.ac.uk/ena/portal/api/filereport?accession = <NCBIaccession > &result=read\_run&fields=fastq ftp").

Raw reads were then downloaded using the compiled FTP links (wget -c --retry-connrefused -P <download\_location > <ftp\_link >). Due to the large variability in Illumina sequencing methods across datasets, all metagenomes were processed as single-end reads by concatenating the fastq read files into a single file. For analysis of inpatient, wastewater, air, and animal metagenomes, metadata and accession numbers were manually curated from multiple studies and downloaded with the same method (Supplementary Data 7).

### AR gene profiling of metagenomes

Concatenated fastq files were rimmed with bbTools version 38.96 (*bbduk.sh in=infile.fastq out=trimmed.fastq ref=adapters ktrim=r k=23 mink=11 hdist=1 tpe tpo*), and aligned to the Comprehensive Antibiotic Resistance Database<sup>20</sup> (CARD) protein homology model database using KMA<sup>94</sup> version 1.4.3 (*kma -i trimmed.fastq -o kma.txt -t db card\_db*). KMA results were analyzed using a custom python script (*kma\_result\_analysis\_V8.ipynb*). In short, positive hits were strictly filtered for template identity over 90%, template coverage over 80%, and p-values under 0.0001. Many hits within the KMA results mapped to single nucleotide gene variants within the CARD database, so CD-HIT<sup>95</sup> was used to cluster the database to 99% identity to reduce redundancy in the dataset. KMA results were then combined based on the CARD clustering. AR gene family profiles were assembled by counting the presence or absence of AR gene families within each metagenome. Individual sample profiles were then used to calculate the percent prevalence of each gene family by country and in all samples grouped by body site.

To confirm our results, we conducted an alternative analysis using HUMAnN3 data acquired from the curatedMetagenomiData package. In short, we ran the Resistance Gene Identifier (RGI)<sup>20</sup> version 5.2.0 on the UniRef90 database used by HUMAnN3 to identify resistance genes. We used the RGI results to build resistance gene profiles for each metagenome by filtering the HUMAnN3 output tables. These profiles were then processed similarly to the KMA analysis to build a heatmap for gene prevalence by country.

### Acquisition of isolate genome and plasmid sequences

Metadata for all genomes withing GenBank were acquired through NCBI ([https://ftp.ncbi.nlm.nih.gov/genomes/genbank/assembly\\_summary\\_genbank.txt](https://ftp.ncbi.nlm.nih.gov/genomes/genbank/assembly_summary_genbank.txt) 2022-2-11). Only Bacterial genomes with an assigned genus were used. For genomes marked as "excluded from RefSeq", only those assonated as "from large multi-isolate projects" and "fragmented assemblies" were used, excluding those derived from environmental sources, metagenomes, or single-cell sequencing. *Salmonella enterica* and *Escherichia coli* were both randomly subsampled to 100 thousand genomes each due to their overrepresentation in the dataset. FTP links were then compiled from the metadata and sequences downloaded with wget. Plasmid sequences were similarly acquired from RefSeq's plasmid database (<https://ftp.ncbi.nlm.nih.gov/genomes/refseq/plasmid/>).

### AR gene profiling of isolates

RGI<sup>20</sup> version 5.2.0 and the CARD database version 3.1.4 was used to detect AR genes in genomes and plasmids (*rgi main --input\_sequence inGenome.fasta --output\_file rigout.txt --input\_type contig --local --clean --alignment\_tool DIAMOND --num\_threads 5 --split\_prodigal\_jobs*

*--include\_loose*). Using a custom python script (*process\_rgi\_V14.ipynb*), RGI hits were filtered to remove "loose" hits and AR gene family profiles were constructed for each genome similar to the metagenomic profiles (Supplementary Data 5). AR gene family prevalence was then calculated for each Class with at least 10 representatives. Gene family/taxa associations with <2 detections were removed. Isolate contigs containing AR genes were classified as genomic or plasmid using PlasForest<sup>96</sup> version 1.4 (*python3 PlasForest.py f -r --threads 4 -i AR\_Contigs.fasta -o out\_plasforest.csv*)

### Human participant consent and sample acquisition

Human participant research was approved by the following committees: Cornell University Institutional Review Board (#1706007261, #1702006922, #1609006586), Aga Khan University Ethics Review Committee (#2019-0550-5166), Ethics Committee of the University of Oxford (OxTREC 38-15) and of Tien Giang Hospital Institutional Review Board (278/BVĐK), the Yale University Institutional Review Board (#2000020688), the Indian BJ Government Medical College Ethics Board, and Weill Cornell Medicine Institutional Review Board (#1503016041, #1504016114). Honduran study participants from 9 isolated villages in the western highlands of Honduras that were part of larger population-based cohort assembled for a different purpose<sup>97</sup> were asked to take part in this study. The goal of the microbiome sampling was to be as comprehensive as possible. Pakistani study participants comprised adults (over the age of 18) recruited via the existing community-based antimicrobial surveillance system established by the two union councils of the Matiari district. Participants were stratified based on ethnicity/caste and tribe and random representatives were chose across the communities. Vietnamese participants comprised adult farmers (over the age of 18) involved in studies conducted by the Oxford University Clinical Research Unit (OUCRU) in Vietnam. Indian participants comprised a subset of pregnant women enrolled in the PRACHITI study in Pune, India<sup>98</sup>. All women were over the age of 18 who presented to the antenatal clinic at BJ Government Medical College in Pune, India, with gestational age between 13–34 weeks. After obtaining consent, all human research participants from Honduras, India, Pakistan, and Vietnam were provided with stool acquisition kits and instructed on how to self-collect the samples, store and transport samples to the study coordinators. After voiding, participants kept samples cold using cold packs until they could be frozen at -80 °C.

### Sample processing, DNA extraction, and metagenomic sequencing

Aliquots of frozen stool samples from Honduras were shipped on dry ice to Cornell University. 200 mg of stool material was placed in 2 ml PowerBead Tubes with 2.8 mm ceramic beads (Qiagen, Cat. No.:13114-50) for DNA extraction using the Chemagic Stool gDNA Extraction Kit (Perkin Elmer, CMG-1076) on the Chemagic 360 Instrument (Perkin Elmer). Resulting genomic DNA quantity, quality, and purity were assessed via determination of the 260/280 nm for values of 1.7–2.0, and 260/230 absorbance ratios for values ≥ and 1% agarose gel electrophoresis to ensure that the gDNA was neither degraded nor displayed RNA contamination. Metagenomic libraries were prepared using KAPA Hyper Library Preparation Kit (KAPA Biosystems, Part#KK8504). Resulting libraries were sequenced on the Illumina NovaSeq (2×150).

### qPCR screening of samples

Gene variant sequences for each target gene were compiled from CARD and aligned in SnapGene. Primers targeting conserved regions were designed with 40–60% GC content and an annealing temperature of -58 °C. Multiple primer sets were designed for each gene and tested for optimal amplification using the OIL-PCR<sup>61</sup> master mix modified by omitting lysozyme and using a lower concentration of Phusion

polymerase (1U per 50 µl of reaction). Isolates acquired from the CDC & FDA Antimicrobial Resistance Isolate Bank were used as templates for testing primers. Amplification quality was evaluated based on clean bands on a gel as well as efficient nested PCR amplification using NEB Luna Universal qPCR Master Mix. Primer/probe sets were designed based on the validated OIL-PCR primers, using the round 1 and fusion primers for amplification, and the nested round 2 primer sequence for the probes (Supplementary Data 10). Primers were tested against purified genomic DNA as a positive control, and purified metagenomic DNA as a negative. Lastly, primers and probes were tested in triplex combinations to check for cross-reactivity between sets. Metagenomic DNA was purified using Qiagen DNeasy PowerSoil Kit (12888-100). Reactions were multiplexed as shown in Supplementary Data 10 and run in duplicate 15 µl reactions with 3 µl of genomic DNA on the Quant Studio 7 Pro for 50 cycles. Positive detections were defined by detected amplification across duplicates.

### 16S rRNA library preparations

16S rRNA libraries were prepared in duplicate 20 µl reactions as follows: 20 U/ml of Phusion Hot Start Flex DNA polymerase, 1x HF Buffer, 2 µM dNTPs, 200 nM of i519F and i786R universal 16S rRNA Illumina primers, and 25 ng of purified DNA template. Reactions were amplified as follows: 95 °C for 3 min, followed by 22 cycles of 95 °C for 5 s, 52 °C for 30 s, and 72 °C for 30 s, before final extension of 72 °C for 5 minutes. Duplicate reactions were pooled and treated with heat-labile Exonuclease I (NEB M0293L) to degrade the primers (10U ExoI per reaction, incubated at 37 °C for 25 min followed by 5 min at 80 °C). ExoI treated 16S amplicons were indexed as described previously<sup>61</sup>.

### OIL-PCR

Nycodenz purification of cells and OIL-PCR was performed with automation as described previously with some adjustments<sup>61</sup>. To remove clumped cells that can signal false gene associations, purified cells were passed through a 5 µm nylon syringe filter. To improve the sensitivity of OIL-PCR, we increased the input of cells 10-fold to 400 thousand cells per reaction based on previous experiments showing that OIL-PCR is more accurate with low abundance cells<sup>61</sup>. In addition, after performing the nested PCR step, multiplexed reactions were recombined and indexed in a single PCR reaction. Lastly, instead of performing Ampure bead purification between the nested PCR and indexing PCR, we used ExoI treatment to degrade nested primers as described for the 16S rRNA library preparations.

OIL-PCR sequences were analyzed with a modified pipeline to sort pooled libraries and allow an analysis of individual ASV variants. Using the script “s1.1\_run\_mrg\_trim\_ft.qsub.sh” reads were first merged using *usearch* mergepairs version 11.0.667 (*usearch -fastq\_mergepairs forward.fq -reverse reverse.fq -fastq\_pctid 80 -fastq\_minmergelen 350 -fastq\_maxmergelen 470 -fastq\_maxdiffs 10 -fastqout merged.fq*), and *cutadapt* version 3.4 was then used to trim primers (*cutadapt -a ^TARGET1\_primer...16S\_rev\$ -a ^TARGET2\_primer... 16S\_rev\$ -a ^TARGET3\_primer... 16S\_rev\$ --discard-untrimmed -o trimmed.fq merged.fq --cores = 2*), before quality filtering with *usearch* (*usearch -fastq\_filter trimmed.fq -fastq\_maxee 0.5 -fastaout filtered.fa*).

Next the script “s2\_combine\_split\_sort\_V2.sh” was used to split target sequences from the 16S sequence using *cutadapt* to recognize the gene-specific fusion primer in each read (*cutadapt -a \$target\_fusion\_primer --discard-untrimmed -o output\_target.fa filtered.fa --suffix \_target --cores = 10*). *Cutadapt* was run separately for each fusion primer in multiplexed sequencing to simultaneously split and sort the target reads. *Cutadapt* was run a final time using the 519F 16S sequence to isolate the 16S portion of each read. Duplicate sequences were counted using *usearch -fastx\_uniques* (*usearch -fastx\_uniques output\_target.fa -fastaout unique.fa -relabel target\_uniq -sizeout*). Lastly, 16S sequences were clustered into OTUs using *usearch -cluster\_otus*

(*usearch -cluster\_otus uniques.fa -otus otus\_out.txt -relabel otu -uparse-out parsed\_file.txt -minsize 1*) and taxonomy assigned using MOTHUR RDP classifier with the SILVA version 132 using the script “s3\_cluster\_assign\_tax.sh”.

A custom python script (*s4\_Tax\_Blast\_Connect\_SNPS\_V7.ipynb*) was used to compile taxonomic-gene associations while keeping track of ASV variants. In short, fasta files were read into tables and sequences were used as identifiers to merge information about the identity of the target gene, assigned taxonomy for the 16S read, and count the number of reads for the association. A second python script was used to graph the output (*graphing\_results\_V4.ipynb*). In summary: 16S sequencing results were imported and merged with the OIL-PCR association to provide community abundance of each ASV and OUT. Target reads were clustered to 99% identity with CD-HIT. Detected associations were filtered for detection across replicates.

### Testing activity of *Bacteroides* genes in *E. coli*

We built a strongly insulated expression vector for testing commensal AR genes to reduce the possibility of leaky gene expression from stray RNA polymerase. We designed a gene block containing two strong terminators upstream of TEM-1—a common ampicillin resistance gene used routinely for cloning—expressed from a synthetically designed promoter and RBS<sup>66,67,99</sup>. Plasmid 353 (pARG-test-insulated) was assembled with a p15A origin and chloramphenicol resistance gene amplified from plasmid pdCas9-bacteria (Addgene 269588, primers BH01/BH02) using Gibson assembly (NEBuilder Hi-Fi Assembly Master Mix).

Sequences for *cfiA-2*, *cfxA-2*, *cepA*, and *cblA-1* were selected from CARD, favoring those that were found on plasmids or other mobile elements. *cfiA* was selected from the pBFUK1 plasmid. Each gene was synthesized along with 250–350 bp upstream to capture the native promoter region. For the native promoter, the expression plasmid 353 was linearized with primers BH01/BH03 and the *ach* gene block was inserted using Gibson assembly. For the synthetic promoter constructs, the plasmid was linearized with primers BH01/BH04 and each gene block was amplified with primer BH08 paired with BH09-BH12 to remove the promoter. The fragments were then assembled with Gibson. For the promoter-less construct, plasmid 353 was amplified with primers BH03/BH05, phosphorylated, and circularized with T4 ligase.

- *cfiA2* from plasmid pBFUK1 (accession AB646744)
- *cfxA2* from *B. fragilis* strain DCMOUH0085B (accession CP037440)
- *cepA* from *B. fragilis* strain 74985 CTn86 mobile element (accession MW169430)
- *cblA1* from Uncultured *Bacteroides* sp. Clone AMP (accession MH883562)

### Antibiotic resistance MIC determination

Plasmid glycerol stocks were inoculated into 5 ml LB + chloramphenicol and grown at 37 °C overnight. Cultures were then diluted 1:10 in fresh LB and evenly spread onto Muller-Hinton agar (Hardy Diagnostics G45) with a cotton swab and allowed to dry. MIC test strips (Liofilchem) were then placed onto the agar with tweezers and plates were placed at 37 °C for at least 24 h before imaging. MIC measurements were evaluated following the manufacturer’s instructions.

### Random mutagenesis library generation and screening

The backbone of plasmid 364 (pARG\_cfiA\_native) was amplified with primers BH03/BH97. The promoter region was amplified with primers BH98/BH99 using the GeneMorph II random mutagenesis kit (Agilent 200550) with 300 ng template and 30 cycles. Mutated promoters were assembled into the backbone using Gibson assembly, diluted, and plated for individual colonies (~24 thousand). Colonies were pooled and stocks were frozen in 20% glycerol. Mutant libraries were



grown overnight in 5 ml LB + chloramphenicol and diluted 1:100 in fresh LB + chloramphenicol with 0, 0.125, 0.5, 2, or 8 µg/ml of cefotaxime and allowed to grow overnight. Growth occurred at all antibiotic concentrations, likely due to compensatory genomic mutations. This was not observed for DH5a carrying *TEM*. Cultures were diluted 1:10 in TE and incubated 10 min at 99 °C to lyse. The raw lysate was used as a template for PCR with primers BI70/BI71 and processed following the 16 S rRNA library preparation protocol (ExoI treatment, index PCR, ampure cleanup, and pooling). Illumina reads were processed similar to 16 S rRNA reads, merging with usearch, trimming with cutadapt, quality filtering with usearch<sup>100</sup>, and counting uniques with usearch. Next, sequences were aligned to the original promoter with bwa mem version 0.7.17, converted to bam, sorted, and indexed with samtools<sup>101</sup> version 1.15.1. The program bam-readcount<sup>100,102</sup> was used to count the bases at each position of the reads. A custom python script (s5\_parse\_tsv\_and\_graph\_V05.ipynb) was then used to parse the bam-readcount output and the results were plotted using Logomaker<sup>103</sup>.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Sequence data was uploaded to the NCBI's Short Read Archive (Honduras metagenomes: [PRJNA999635](https://www.ncbi.nlm.nih.gov/sra/PRJNA999635); Neutropenic metagenomes: [PRJNA999651](https://www.ncbi.nlm.nih.gov/sra/PRJNA999651); OIL-PCR and 16 S sequencing: [PRJNA1001934](https://www.ncbi.nlm.nih.gov/sra/PRJNA1001934)).

### Code availability

Custom code is uploaded <https://github.com/britolab/Global-and-taxonomic-ARG-prevalence> <https://doi.org/10.5281/zenodo.10031964>

### References

- Murray, C. J. et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet* **399**, 629–655 (2022).
- Brito, I. L. et al. Mobile genes in the human microbiome are structured from global to individual scales. *Nature* **535**, 435–439 (2016).
- Pehrsson, E. C. et al. Interconnected microbiomes and resistomes in low-income human habitats. *Nature* **533**, 212–216 (2016).
- Zhang, Z. et al. Assessment of global health risk of antibiotic resistance genes. *Nat. Commun.* **13**, 1553 (2022).
- Zhou, H., Beltrán, J. F. & Brito, I. L. Functions predict horizontal gene transfer and the emergence of antibiotic resistance. *Sci. Adv.* **7**, eabj5056 (2021).
- Smillie, C. S. et al. Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* **480**, 241–244 (2011).
- Maier, L. et al. Unravelling the collateral damage of antibiotics on gut bacteria. *Nature* **599**, 120–124 (2021).
- Lamberte, L. E. & van Schaik, W. Antibiotic resistance in the commensal human gut microbiota. *Curr. Opin. Microbiol.* **68**, 102150 (2022).
- Ghosh, T. S., Gupta, S. S., Nair, G. B. & Mande, S. S. In silico analysis of antibiotic resistance genes in the gut microflora of individuals from diverse geographies and age-groups. *PLoS ONE* **8**, e83823 (2013).
- Rahman, S. F., Olm, M. R., Morowitz, M. J. & Banfield, J. F. Machine learning leveraging genomes from metagenomes identifies influential antibiotic resistance genes in the infant gut microbiome. *mSystems* **3**, e00123–17 (2018).
- Gibson, M. K. et al. Developmental dynamics of the preterm infant gut microbiota and antibiotic resistome. *Nat. Microbiol.* **1**, 1–10 (2016).
- Hu, Y. et al. Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota. *Nat. Commun.* **4**, 2151 (2013).
- Yassour, M. et al. Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. *Sci. Transl. Med.* **8**, 343ra81–343ra81 (2016).
- Forster, S. C. et al. Strain-level characterization of broad host range mobile genetic elements transferring antibiotic resistance from the human microbiome. *Nat. Commun.* **13**, 1445 (2022).
- Jiang, X. et al. Dissemination of antibiotic resistance genes from antibiotic producers to pathogens. *Nat. Commun.* **8**, 15784 (2017).
- Ronda, C., Chen, S. P., Cabral, V., Yaung, S. J. & Wang, H. H. Metagenomic engineering of the mammalian gut microbiome in situ. *Nat. Methods* **16**, 167–170 (2019).
- Munck, C., Sheth, R. U., Freedberg, D. E. & Wang, H. H. Recording mobile DNA in the gut microbiota using an Escherichia coli CRISPR-Cas spacer acquisition platform. *Nat. Commun.* **11**, 95 (2020).
- Sawa, T., Kooguchi, K. & Moriyama, K. Molecular diversity of extended-spectrum β-lactamases and carbapenemases, and antimicrobial resistance. *J. Intensive Care* **8**, 13 (2020).
- Pasolli E, Schiffer L, Manghi P, Renson A, Obenchain V, Truong DT, Beghini F, Malik F, Ramos M, Dowd JB, Huttenhower C, Morgan M, Segata N, Waldron L. Accessible, curated metagenomic data through ExperimentHub. *Nat Methods*. **14**, 1023–1024 (2017).
- Alcock, B. P. et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res.* **51**, D690–D699 (2023).
- WHO model list of essential medicines – 22nd list. <https://www.who.int/publications-detail-redirect/WHO-MHP-HPS-EML-2021.02> (2021)
- Du, D. et al. Multidrug efflux pumps: structure, function and regulation. *Nat. Rev. Microbiol.* **16**, 523–539 (2018).
- Nielsen, T. K., Browne, P. D. & Hansen, L. H. Antibiotic resistance genes are differentially mobilized according to resistance mechanism. *GigaScience* **11**, giac072 (2022).
- Lopatkin, A. J. et al. Clinically relevant mutations in core metabolic genes confer antibiotic resistance. *Science* **371**, eaba0862 (2021).
- Pasolli, E. et al. Accessible, curated metagenomic data through ExperimentHub. *Nat. Methods* **14**, 1023–1024 (2017).
- Hendriksen, R. S. et al. Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. *Nat. Commun.* **10**, 1124 (2019).
- Munk, P. et al. Genomic analysis of sewage from 101 countries reveals global landscape of antimicrobial resistance. *Nat. Commun.* **13**, 7251 (2022).
- Maugeri, A., Barchitta, M., Puglisi, F. & Agodi, A. Socio-economic, governance and health indicators shaping antimicrobial resistance: an ecological analysis of 30 European countries. *Glob. Health* **19**, 12 (2023).
- Collignon, P., Beggs, J. J., Walsh, T. R., Gandra, S. & Laxminarayan, R. Anthropological and socioeconomic factors contributing to global antimicrobial resistance: a univariate and multivariable analysis. *Lancet Planet. Health* **2**, e398–e405 (2018).
- Zhang, Z. et al. Genetic diversity and characteristics of bla<sub>NDM</sub>-positive plasmids in *Escherichia coli*. *Front. Microbiol.* **12**, 729952 (2021).
- Philippon, A., Arlet, G. & Jacoby, G. A. Plasmid-determined AmpC-type β-lactamases. *Antimicrob. Agents Chemother.* **46**, 1–11 (2002).
- Wexler, A. G. & Goodman, A. L. An insider's perspective: bacteroides as a window into the microbiome. *Nat. Microbiol.* **2**, 1–11 (2017).

33. Montassier, E. et al. Probiotics impact the antibiotic resistance gene reservoir along the human GI tract in a person-specific and antibiotic-dependent manner. *Nat. Microbiol.* **6**, 1043–1054 (2021).
34. Anthony, W. E. et al. Acute and persistent effects of commonly used antibiotics on the gut microbiome and resistome in healthy adults. *Cell Rep.* **39**, 110649 (2022).
35. WHO report on surveillance of antibiotic consumption. <https://www.who.int/publications-detail-redirect/who-report-on-surveillance-of-antibiotic-consumption>.
36. Feng, Y. et al. Rapid detection of new delhi metallo- $\beta$ -lactamase gene using recombinase-aided amplification directly on clinical samples from children. *Front. Microbiol.* **12**, 691289 (2021).
37. Day, K. M. et al. Prevalence and molecular characterization of Enterobacteriaceae producing NDM-1 carbapenemase at a military hospital in Pakistan and evaluation of two chromogenic media. *Diagn. Microbiol. Infect. Dis.* **75**, 187–191 (2013).
38. Yong, D. et al. Characterization of a new metallo-beta-lactamase gene, bla(NDM-1), and a novel erythromycin esterase gene carried on a unique genetic structure in Klebsiella pneumoniae sequence type 14 from India. *Antimicrob. Agents Chemother.* **53**, 5046–5054 (2009).
39. Moquet, O. et al. Class D OXA-48 Carbapenemase in Multidrug-Resistant Enterobacteria, Senegal. *Emerg. Infect. Dis.* **17**, 143–144 (2011).
40. Evans, B. A. & Amyes, S. G. B. OXA  $\beta$ -Lactamases. *Clin. Microbiol. Rev.* **27**, 241–263 (2014).
41. Zaheer, R. et al. Comparative diversity of microbiomes and resistomes in beef feedlots, downstream environments and urban sewage influent. *BMC Microbiol.* **19**, 197 (2019).
42. Zaheer, R. et al. Impact of sequencing depth on the characterization of the microbiome and resistome. *Sci. Rep.* **8**, 5890 (2018).
43. Yokoyama, M. et al. Microbial diversity and antimicrobial resistance in faecal samples from acute medical patients assessed through metagenomic sequencing. *PLoS ONE* **18**, e0282584 (2023).
44. Yan, J. et al. A compilation of fecal microbiome shotgun metagenomics from hematopoietic cell transplantation patients. *Sci. Data* **9**, 219 (2022).
45. Xiao, L. et al. A reference gene catalogue of the pig gut microbiome. *Nat. Microbiol.* **1**, 1–6 (2016).
46. Wu, D. et al. Inhalable antibiotic resistomes emitted from hospitals: metagenomic insights into bacterial hosts, clinical relevance, and environmental risks. *Microbiome* **10**, 19 (2022).
47. Talat, A., Blake, K. S., Dantas, G. & Khan, A. U. Metagenomic insight into microbiome and antibiotic resistance genes of high clinical concern in urban and rural hospital wastewater of northern India origin: a major reservoir of antimicrobial resistance. *Microbiol. Spectr.* **11**, e04102–e04122 (2023).
48. Takahashi, E. et al. Metagenomic analysis of diarrheal stools in Kolkata, India, indicates the possibility of subclinical infection of Vibrio cholerae O1. *Sci. Rep.* **12**, 19473 (2022).
49. Schneider, D., Aßmann, N., Wicke, D., Poehlein, A. & Daniel, R. Metagenomes of wastewater at different treatment stages in central Germany. *Microbiol. Resour. Announc.* **9**, e00201–e00220 (2020).
50. Majeed, H. J. et al. Evaluation of metagenomic-enabled antibiotic resistance surveillance at a conventional wastewater treatment plant. *Front. Microbiol.* **12**, 657954 (2021).
51. Lira, F., Vaz-Moreira, I., Tamames, J., Manaia, C. M. & Martínez, J. L. Metagenomic analysis of an urban resistome before and after wastewater treatment. *Sci. Rep.* **10**, 8174 (2020).
52. King, P. et al. Longitudinal metagenomic analysis of hospital air identifies clinically relevant microbes. *PLoS ONE* **11**, e0160124 (2016).
53. Kent, A. G., Vill, A. C., Shi, Q., Satlin, M. J. & Brito, I. L. Widespread transfer of mobile antibiotic resistance genes within individual gut microbiomes revealed through bacterial Hi-C. *Nat. Commun.* **11**, 4379 (2020).
54. Jankowski, P. et al. Metagenomic community composition and resistome analysis in a full-scale cold climate wastewater treatment plant. *Environ. Microbiome* **17**, 3 (2022).
55. Guo, X. et al. Metagenomic analysis of antibiotic resistance genes in untreated wastewater from three different hospitals. *Front. Microbiol.* **12**, 709051 (2021).
56. Jones, K. E. et al. Global trends in emerging infectious diseases. *Nature* **451**, 990–993 (2008).
57. Liu, Y.-Y. et al. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. *Lancet Infect. Dis.* **16**, 161–168 (2016).
58. Yang, Q. et al. Balancing mcr-1 expression and bacterial survival is a delicate equilibrium between essential cellular defence mechanisms. *Nat. Commun.* **8**, 2054 (2017).
59. Ogunlana, L. et al. Regulatory fine-tuning of mcr-1 increases bacterial fitness and stabilises antibiotic resistance in agricultural settings. *ISME J.* **17**, 2058–2069 (2023).
60. Lopatkin, A. J. et al. Persistence and reversal of plasmid-mediated antibiotic resistance. *Nat. Commun.* **8**, 1689 (2017).
61. Diebold, P. J., New, F. N., Hovan, M., Satlin, M. J. & Brito, I. L. Linking plasmid-based beta-lactamases to their bacterial hosts using single-cell fusion PCR. *eLife* **10**, e66834 (2021).
62. Groussin, M. et al. Elevated rates of horizontal gene transfer in the industrialized human microbiome. *Cell* **184**, 2053–2067.e18 (2021).
63. Forster, S. C. et al. A human gut bacterial genome and culture collection for improved metagenomic analyses. *Nat. Biotechnol.* **37**, 186–192 (2019).
64. Zou, Y. et al. 1520 reference genomes from cultivated human gut bacteria enable functional microbiome analyses. *Nat. Biotechnol.* **37**, 179–185 (2019).
65. Redondo-Salvo, S. et al. Pathways for horizontal gene transfer in bacteria revealed by a global map of their plasmids. *Nat. Commun.* **11**, 3602 (2020).
66. LaFleur, T. L., Hossain, A. & Salis, H. M. Automated model-predictive design of synthetic promoters to control transcriptional profiles in bacteria. *Nat. Commun.* **13**, 5159 (2022).
67. Reis, A. C. & Salis, H. M. An automated model test system for systematic development and improvement of gene expression models. *ACS Synth. Biol.* **9**, 3145–3156 (2020).
68. Smith, C. J., Bennett, T. K. & Parker, A. C. Molecular and genetic analysis of the Bacteroides uniformis cephalosporinase gene, cblA, encoding the species-specific beta-lactamase. *Antimicrob. Agents Chemother.* **38**, 1711–1715 (1994).
69. Rasmussen, B. A., Gluzman, Y. & Tally, F. P. Escherichia coli chromosomal mutations that permit direct cloning of the Bacteroides fragilis metallo-beta-lactamase gene, ccrA. *Mol. Microbiol.* **5**, 1211–1219 (1991).
70. Sóni, J. et al. Examination of cfiA-mediated carbapenem resistance in Bacteroides fragilis strains from a European antibiotic susceptibility survey. *Int. J. Antimicrob. Agents* **28**, 497–502 (2006).
71. Goto, T., Tanaka, K., Minh Tran, C. & Watanabe, K. Complete sequence of pBFUK1, a carbapenemase-harboring mobilizable plasmid from Bacteroides fragilis, and distribution of pBFUK1-like plasmids among carbapenem-resistant B. fragilis clinical isolates. *J. Antibiot.* **66**, 239–242 (2012).
72. Valentine, P. J., Shoemaker, N. B. & Salyers, A. A. Mobilization of Bacteroides plasmids by Bacteroides conjugal elements. *J. Bacteriol.* **170**, 1319–1324 (1988).

73. Thomas, J. & Hecht, D. W. Interaction of *Bacteroides fragilis* pLV22a relaxase and transfer DNA with *Escherichia coli* RP4-TraG coupling protein. *Mol. Microbiol.* **66**, 948–960 (2007).
74. Coyne, M. J., Zitomersky, N. L., McGuire, A. M., Earl, A. M. & Comstock, L. E. Evidence of extensive DNA transfer between *Bacteroidales* Species within the human gut. *mBio* **5**, e01305–e01314 (2014).
75. Garud, N. R., Good, B. H., Hallatschek, O. & Pollard, K. S. Evolutionary dynamics of bacteria in the gut microbiome within and across hosts. *PLoS Biol.* **17**, e3000102 (2019).
76. Zlitni, S. et al. Strain-resolved microbiome sequencing reveals mobile elements that drive bacterial competition on a clinical timescale. *Genome Med.* **12**, 50 (2020).
77. Zhao, S. et al. Adaptive evolution within gut microbiomes of healthy people. *Cell Host Microbe* **25**, 656–667.e8 (2019).
78. Yaffe, E. & Relman, D. A. Tracking microbial evolution in the human gut using Hi-C reveals extensive horizontal gene transfer, persistence and adaptation. *Nat. Microbiol.* **5**, 343–353 (2020).
79. Bryan, L. E., Kowand, S. K. & Van Den Elzen, H. M. Mechanism of aminoglycoside antibiotic resistance in anaerobic bacteria: *Clostridium perfringens* and *Bacteroides fragilis*. *Antimicrob. Agents Chemother.* **15**, 7–13 (1979).
80. Pradier, L. & Bedhomme, S. Ecology, more than antibiotics consumption, is the major predictor for the global distribution of aminoglycoside-modifying enzymes. *eLife* **12**, e77015 (2023).
81. Gullberg, E. et al. Selection of a multidrug resistance plasmid by sublethal levels of antibiotics and heavy metals. *mBio* **5**, e01918–14 (2014).
82. Kim, S., Covington, A. & Pamer, E. G. The intestinal microbiota: antibiotics, colonization resistance, and enteric pathogens. *Immunol. Rev.* **279**, 90–105 (2017).
83. Vien, L. T. M. et al. The co-selection of fluoroquinolone resistance genes in the gut flora of vietnamese children. *PLoS ONE* **7**, e42919 (2012).
84. Kim, S. W. et al. Outer membrane vesicles from  $\beta$ -lactam-resistant *Escherichia coli* enable the survival of  $\beta$ -lactam-susceptible *E. coli* in the presence of  $\beta$ -lactam antibiotics. *Sci. Rep.* **8**, 5402 (2018).
85. Costerton, J. W., Stewart, P. S. & Greenberg, E. P. Bacterial biofilms: a common cause of persistent infections. *Science* **284**, 1318–1322 (1999).
86. Stewart, P. S. & William Costerton, J. Antibiotic resistance of bacteria in biofilms. *Lancet* **358**, 135–138 (2001).
87. Stecher, B. et al. Gut inflammation can boost horizontal gene transfer between pathogenic and commensal Enterobacteriaceae. *Proc. Natl Acad. Sci. USA* **109**, 1269–1274 (2012).
88. Palleja, A. et al. Recovery of gut microbiota of healthy adults following antibiotic exposure. *Nat. Microbiol.* **3**, 1255–1265 (2018).
89. MacPherson, C. W. et al. Gut bacterial microbiota and its resistome rapidly recover to basal state levels after short-term amoxicillin-clavulanic acid treatment in healthy adults. *Sci. Rep.* **8**, 11192 (2018).
90. Young, V. B. & Hayden, M. K. Environmental management in the gut: fecal transplantation to restore the intestinal ecosystem. *Infect. Dis. Lond. Engl.* **48**, 593–595 (2016).
91. Edgar, R., Friedman, N., Molshanski-Mor, S. & Qimron, U. Reversing bacterial resistance to antibiotics by phage-mediated delivery of dominant sensitive genes. *Appl. Environ. Microbiol.* **78**, 744–751 (2012).
92. Yosef, I., Manor, M., Kiro, R. & Qimron, U. Temperate and lytic bacteriophages programmed to sensitize and kill antibiotic-resistant bacteria. *Proc. Natl Acad. Sci. USA* **112**, 7267–7272 (2015).
93. Aslam, B. et al. CRISPR-Cas system: a potential alternative tool to cope antibiotic resistance. *Antimicrob. Resist. Infect. Control* **9**, 131 (2020).
94. Clausen, P. T. L. C., Aarestrup, F. M. & Lund, O. Rapid and precise alignment of raw reads against redundant databases with KMA. *BMC Bioinforma.* **19**, 307 (2018).
95. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinforma. Oxf. Engl.* **22**, 1658–1659 (2006).
96. Pradier, L., Tissot, T., Fiston-Lavier, A.-S. & Bedhomme, S. PlasForest: a homology-based random forest classifier for plasmid detection in genomic datasets. *BMC Bioinforma.* **22**, 349 (2021).
97. Shakya, H. B. et al. Exploiting social influence to magnify population-level behaviour change in maternal and child health: study protocol for a randomised controlled trial of network targeting algorithms in rural Honduras. *BMJ Open* **7**, e012996 (2017).
98. Bhosale, R. et al. Stages of pregnancy and HIV affect diagnosis of tuberculosis infection and *Mycobacterium tuberculosis* (MTB)-induced immune response: Findings from PRACHITI, a cohort study in Pune, India. *Int. J. Infect. Dis.* **112**, 205–211 (2021).
99. Cetnar, D. P. & Salis, H. M. Systematic quantification of sequence and structural determinants controlling mRNA stability in bacterial operons. *ACS Synth. Biol.* **10**, 318–332 (2021).
100. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
101. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Giga-Science* **10**, giab008 (2021).
102. Khanna, A. et al. Bam-readcount - rapid generation of basepair-resolution sequence metrics. Preprint at <https://arxiv.org/abs/2107.12817> (2021).
103. Tareen, A. & Kinney, J. B. Logomaker: beautiful sequence logos in Python. *Bioinformatics* **36**, 2272–2274 (2020).

## Acknowledgements

We thank Albert Vill and Adarsh Singh for their suggestions and assistance in carrying out this research. We thank Kyu Rhee and Craig Altier for their helpful comments on the manuscript. We thank the members of the Honduras sample collection team, and study members from the Zoonosis Group, Subdepartment of Animal Health in Tien Giang for their assistance in sample collection in Vietnam. We thank members of the Biotechnology Resource Center at Cornell University. This work was supported by funding from the U.S. National Institutes of Health (1R01AI151059, 1DP2HL141007), the National Sciences Foundation (1650122, 1661338), the Gates Foundation (OPP1161064), the NOMIS Foundation (GR108454-CON-80002144) and SCENERI project funded by AXA (to N.T.H.). During the course of this work, I.L.B. was a Pew Charitable Trusts Biomedical Research Scholar, Packard Foundation Fellow, and Sloan Foundation Fellow.

## Author contributions

Conceptualization: P.J.D. and I.L.B.; Methodology: P.J.D. and I.L.B.; Computational analysis: P.J.D.; Experimental work: P.J.D. and M.W.R.; Management of samples and acquisition: Q.S., N.V.T., F.U., S.A., V.K., P.D., M.A., N.T.H., N.A.C., N.T.I., S.A.A., and J.S.M.; Data curation: P.J.D.; Visualization: P.J.D. and I.L.B.; Writing—original draft: P.J.D. and I.L.B.; Writing—review and editing: P.J.D., M.W.R., Q.S., N.T.H., N.A.C., N.T.I., S.A.A., J.S.M. and I.L.B.; Project administration: I.L.B.; Funding acquisition: N.A.C. and I.L.B.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-42998-6>.

**Correspondence** and requests for materials should be addressed to Ilana L. Brito.

**Peer review information** *Nature Communications* thanks Frank Aarstrup, and the other, anonymous, reviewers for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

---

**Peter J. Diebold<sup>1</sup>, Matthew W. Rhee<sup>1</sup>, Qiaojuan Shi<sup>1</sup>, Nguyen Vinh Trung<sup>2</sup>, Fayaz Umrani<sup>3</sup>, Sheraz Ahmed<sup>3</sup>, Vandana Kulkarni<sup>4</sup>, Prasad Deshpande<sup>4</sup>, Mallika Alexander<sup>4</sup>, Ngo Thi Hoa<sup>2,5,6</sup>, Nicholas A. Christakis<sup>7</sup>, Najeeha Talat Iqbal<sup>3</sup>, Syed Asad Ali<sup>3</sup>, Jyoti S. Mathad<sup>8</sup> & Ilana L. Brito<sup>1</sup>** ✉

---

<sup>1</sup>Meinig School of Biomedical Engineering, Cornell University, Ithaca, NY, USA. <sup>2</sup>Oxford University Clinical Research Unit (OUCRU) in Ho Chi Minh City, Ho Chi Minh city, Viet Nam. <sup>3</sup>Aga Khan University, Karachi, Pakistan. <sup>4</sup>Johns Hopkins University Clinical Trials Unit, Byramjee Jeejeebhoy Government Medical College, Pune, Maharashtra, India. <sup>5</sup>Centre for Tropical Medicine, Nuffield Department of Medicine, University of Oxford, Oxford, UK. <sup>6</sup>Microbiology Department and Center for Tropical Medicine Research, Ngoc Thach University of Medicine, Ho Chi Minh city, Vietnam. <sup>7</sup>Yale University, New Haven, CT, USA. <sup>8</sup>Weill Cornell Medicine, New York, NY, USA. ✉e-mail: [ibrito@cornell.edu](mailto:ibrito@cornell.edu)