# Phylogenetic barriers to horizontal transfer of antimicrobial peptide resistance genes in the human gut microbiota

Bálint Kintses [1,8]*, Orsolya Méhi[1,8], Eszter Ari [1,2,8], Mónika Számel[1,3], Ádám Györkei[1],
Pramod K. Jangir[1,3], István Nagy[4,5], Ferenc Pál[1], Gergely Fekete[1], Roland Tengölics[1], Ákos Nyerges [1,3],
István Likó[6], Anita Bálint[7], Tamás Molnár[7], Balázs Bálint [4], Bálint Márk Vásárhelyi[4],
Misshelle Bustamante[2], Balázs Papp [1]* and Csaba Pál[1]*

The human gut microbiota has adapted to the presence of antimicrobial peptides (AMPs), which are ancient components of immune defence. Despite its medical importance, it has remained unclear whether AMP resistance genes in the gut microbiome are available for genetic exchange between bacterial species. Here, we show that AMP resistance and antibiotic resistance genes differ in their mobilization patterns and functional compatibilities with new bacterial hosts. First, whereas AMP resistance genes are widespread in the gut microbiome, their rate of horizontal transfer is lower than that of antibiotic resistance genes. Second, gut microbiota culturing and functional metagenomics have revealed that AMP resistance genes originating from phylogenetically distant bacteria have only a limited potential to confer resistance in *Escherichia coli*, an intrinsically susceptible species. Taken together, functional compatibility with the new bacterial host emerges as a key factor limiting the genetic exchange of AMP resistance genes. Finally, our results suggest that AMPs induce highly specific changes in the composition of the human microbiota, with implications for disease risks.

The maintenance of homeostasis between the gut microbiota and human host tissues entails a complex co-evolutionary relationship[1,2]. Mucosal barriers covering the intestinal epithelium restrict microbes to the lumen, control the composition of commensal inhabitants and ensure the removal of pathogens[3,4]. Cationic host antimicrobial peptides (AMPs) have crucial roles in this process[5]. They are among the most ancient and efficient components of the innate immune defence in multicellular organisms and have retained their efficacy for millions of years[5,6]. As AMPs have a broad spectrum of activity, much effort has been put into finding potential antibacterial drugs among AMPs[7,8].

However, therapeutic use of AMPs may drive bacterial evolution of resistance to our own immunity peptides[9,10]. Therefore, it is of central importance to establish whether genes that influence AMP resistance (AMP resistance genes) in the gut microbiome are available for genetic exchange with other bacterial species. Several lines of observation support the plausibility of this scenario. The gut bacterial community is a rich source of mobile antibiotic resistance genes[11], and certain abundant gut bacterial species exhibit high levels of intrinsic resistance to AMPs[12]. Moreover, even single genes can confer high AMP resistance in Bacteroidetes[12]. However, beyond the recent discovery of a horizontally spreading resistance gene family[13,14], the mobility of AMP-resistance-encoding genes across bacterial species has remained unknown.

Here, we have applied an integrated approach to systematically characterize the mobilization potential of the AMP resistance gene reservoir in the human gut microbiome. First, we examined the patterns of horizontal gene transfer events involving AMP resistance genes by analysing bacterial genome sequences from the human gut and naturally occurring plasmid sequences from the human microbiome. Next, we experimentally compared the functional compatibility of AMP resistance versus antibiotic resistance genes from the gut microbiome with a susceptible host, *Escherichia coli*, by performing functional metagenomic selections and by culturing the gut microbiome in the presence of diverse AMPs and small-molecule antibiotics. Together, these analyses revealed that AMP resistance genes are less frequently mobilized owing to lack of functional compatibility with new bacterial hosts.

## Results

**Infrequent horizontal transfer of AMP resistance genes in the gut microbiota.** We begin by asking whether the genetic determinants of resistance to AMPs and antibiotics, respectively, differ in their rate of horizontal transfer in the human gut microbiota. To systematically address this issue, we first collected previously characterized AMP and antibiotic resistance genes from the literature and databases, yielding a comprehensive catalogue of 114 and 199 AMP resistance and antibiotic resistance gene families, respectively (see Methods and Supplementary Table 1). By definition, AMP resistance

[1]Synthetic and Systems Biology Unit, Institute of Biochemistry, Biological Research Centre of the Hungarian Academy of Sciences, Szeged, Hungary. [2]Department of Genetics, Eötvös Loránd University, Budapest, Hungary. [3]Doctoral School in Biology, Faculty of Science and Informatics, University of Szeged, Szeged, Hungary. [4]SeqOmics Biotechnology Ltd, Mórahalom, Hungary. [5]Sequencing Platform, Institute of Biochemistry, Biological Research Centre of the Hungarian Academy of Sciences, Szeged, Hungary. [6]Hereditary Endocrine Tumors Research Group, Hungarian Academy of Sciences and Semmelweis University, Budapest, Hungary. [7]1st Department of Internal Medicine, Albert Szent-Györgyi Health Centre, University of Szeged, Szeged, Hungary. [8]These authors contributed equally: Bálint Kintses, Orsolya Méhi, Eszter Ari. *e-mail: kintses.balint@brc.mta.hu; pappb@brc.hu; cpal@brc.hu

genes influence bacterial susceptibility to at least one AMP when mutated (see Methods). Next, we compared the frequencies of these previously identified AMP and antibiotic resistance genes in a catalogue of 37,853 horizontally transferred genes from 567 genome sequences of phylogenetically diverse bacterial species in the human gut microbiota[15]. This mobile gene catalogue relies on the identification of nearly identical genes that are shared by distantly related bacterial genomes and thereby provides a snapshot of the gene set subjected to recent horizontal gene transfer events in a representative sample of the human gut microbiome[15]. We identified homologues of the literature-curated resistance genes for which at least one transfer event was reported (that is, those present in the mobile gene pool; see Methods and Supplementary Table 2).

We found that the relative frequency of AMP resistance genes within the pool of mobile genes was 4.8-fold lower than that of antibiotic- resistance genes, in spite of their similar frequencies in the genomes of the gut microbiota (Fig. 1a, Supplementary Fig. 1a and Supplementary Table 2). Notably, the relative under-representation of AMP resistance genes in the mobile gene pool cannot be simply attributed to the large physiological differences between Gram-negative and Gram-positive bacteria. When these two bacterial groups were considered individually, AMP resistance genes remained under-represented in the mobile gene pool compared to antibiotic resistance genes (Fig. 1a). Moreover, the unique transferred AMP resistance genes were shared between fewer bacterial species, indicating fewer transfer events per gene (Fig. 1b). Notably, 65% of these transfer events mobilized AMP-transporting efflux pumps between bacteria within the Firmicutes phylum (Fig. 1c, Supplementary Fig. 1b and Supplementary Table 3).

To further support the low mobility of AMP resistance genes, we next explored whether AMP resistance genes from our literature-curated list are associated with naturally occurring plasmid sequences or integrative conjugative elements (ICEs) in the human gut microbiome (see Methods). Strikingly, while 86% of the antibiotic resistance genes had close homologues on plasmids, only 33% of the AMP resistance genes did (Fig. 1d and Supplementary Table 4). Moreover, the plasmids carrying individual antibiotic resistance genes were more widespread across bacterial species than those carrying AMP resistance genes (Fig. 1e). Notably, many of these plasmid-encoded AMP resistance genes were proteases and efflux pumps carried by virulence or multi-drug resistance plasmids in the human microbiome (Supplementary Table 4). Finally, as with plasmid sequences, we found that disproportionately fewer AMP resistance genes had homologues in ICEs than antibiotic resistance genes (Supplementary Fig. 2 and Supplementary Table 4).

Overall, these results suggest that AMP resistance genes are less frequently transferred across bacterial species in the human microbiota.

**Short genomic fragments from the gut microbiota rarely confer AMP resistance.** One possible reason for the low mobilization of AMP resistance genes could be that AMP resistance is an intrinsic property of certain bacteria shaped by multi-gene networks[10]. Genes involved in AMP resistance may display strong epistatic interactions, and therefore they may have little or no impact on resistance individually. If this was so, horizontal gene transfer of single genes or transcriptional units encoded by short genomic fragments would not provide resistance in the recipient bacterial species. Indeed, AMPs interact with the cell membrane, a highly interconnected cellular structure, and membership in complex cellular subsystems has been shown to limit horizontal gene transfer[16,17].

To investigate this scenario, we experimentally compared the ability of short genomic fragments to transfer resistance phenotypes towards AMPs versus antibiotics. To this end, we applied an established functional metagenomic protocol[11,18] to identify random 1.5–5 kb DNA fragments in the gut microbiome that confer

resistance in an intrinsically susceptible *E. coli* strain. Importantly, the length distributions of the known AMP resistance and antibiotic resistance genes are well within this fragment size range (Supplementary Fig. 3), indicating that our protocol is suitable to capture single resistance genes for both AMPs and antibiotics. Metagenomic DNA from human gut faecal samples was isolated from two unrelated, healthy individuals who had not taken any antibiotics for at least one year. The resulting DNA samples were cut, and fragments between 1.5 and 5 kb were shotgun cloned into a plasmid to express the genetic information in *E. coli* K-12. About 2 million members from each library, corresponding to a total coverage of 8 Gb (the size of ~2,000 bacterial genomes), were then selected on solid culture medium in the presence of one of 12 diverse AMPs and 11 antibiotics (Supplementary Table 5) at concentrations where the wild-type host strain is susceptible. Finally, using a third-generation long-read sequencing pipeline[19], the number of unique DNA fragments conferring resistance (that is, resistance contigs) was determined.

In agreement with earlier studies[11,20], multiple resistant clones emerged against all tested antibiotics (Fig. 2a and Supplementary Table 6). In sharp contrast, no resistance was conferred against half of the AMPs tested, and, in general, the number of unique AMP resistance contigs ($N = 34$) was substantially lower than the number of unique antibiotic resistance contigs ($N = 119$) (Fig. 2a and Supplementary Table 6). Polymyxin B—an antimicrobial peptide used as a last-resort drug in the treatment of multidrug-resistant Gram-negative bacterial infections[21]—is a notable exception to this trend, with a relatively high number of unique resistance contigs (Fig. 2a). Indeed, a resistance gene (*mcr-1*) against polymyxin B is rapidly spreading horizontally worldwide, representing an alarming global healthcare issue[22]. In contrast to polymyxin B, we detected only one unique contig conferring resistance to LL37, a human AMP abundantly secreted in the intestinal epithelium[23] (Supplementary Table 6). These specific AMP resistance genes are involved in cell surface modification, peptide proteolysis and regulation of the outer membrane stress response (Table 1, Supplementary Table 6 and Supplementary Fig. 4).

If lack of functional compatibility with the host cell prevents AMP resistance genes from exerting their phenotypic effects, then DNA fragments identified in our screen should more often come from phylogenetically closely related bacteria. A total of 53% of the contigs showed over 95% sequence identity to bacterial genome sequences from the HMP database[24] (see Methods and Supplementary Fig. 5), allowing us to infer the source taxa with high accuracy (Supplementary Table 6). Indeed, while antibiotic resistance contigs were over-represented from Firmicutes, the most abundant phylum in the gut, AMP resistance contigs originated excessively from Proteobacteria, which are phylogenetically close relatives of the host *E. coli* (Fig. 2b and Supplementary Fig. 6). Notably, this trend was not driven by polymyxin B only, but was valid for the rest of the AMPs as well (Supplementary Fig. 6).

Whereas these patterns are consistent with the hypothesis that the genetic determinants of AMP resistance are difficult to transfer via short genomic fragments owing to a lack of functional compatibility with the new host, another explanation is also plausible. In particular, AMP-resistant bacteria might be relatively rare in the human gut microbiota, so AMP resistance genes from these bacteria might simply remain undetected. However, as explained below, we can rule out this alternative hypothesis.

**AMP resistant gut bacteria are abundant and phylogenetically diverse.** To assess the diversity and taxonomic composition of gut bacteria displaying resistance to AMPs and antibiotics, we carried out anaerobic cultivations and selections of gut microbiota using a state-of-the-art protocol[25]. To this end, faecal samples were collected from seven healthy individuals (Faecal 7 mix, see Methods).
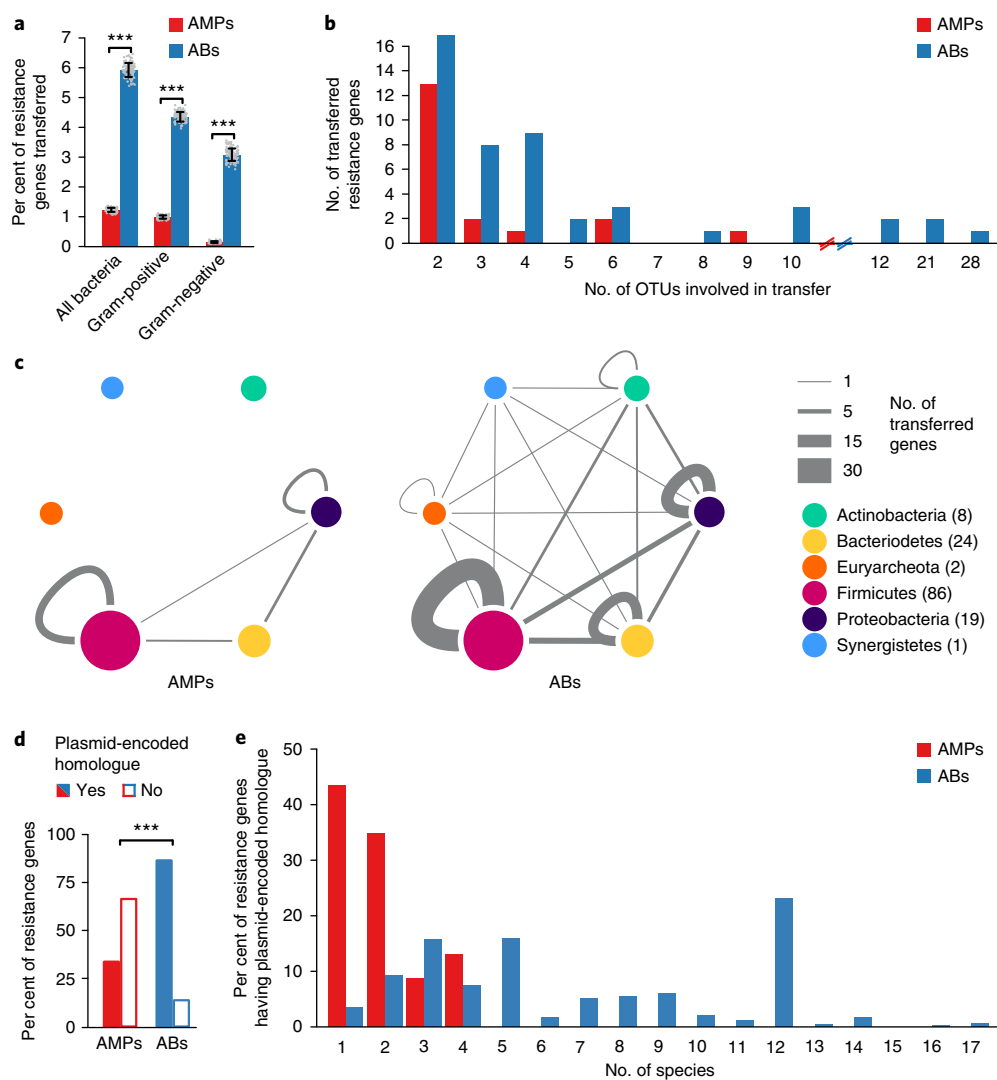
**Fig. 1 | AMP resistance genes are less frequently transferred in the human gut microbiome than antibiotic resistance genes. a**, Percentage of AMP resistance genes (AMP, red bars) and antibiotic resistance genes (AB, blue bars) detected as horizontally transferred (that is, present in the mobile gene pool). Resistance genes were identified using BLAST sequence similarity searches in the genome sequences of the gut microbiota (see Methods). ***Significant difference (left to right): $P = 2 \times 10^{-16}$, $4 \times 10^{-16}$, $2 \times 10^{-13}$ from two-sided binomial test, in each case $n = 100$ both for AMPs and ABs). Centre and error bars represent mean and s.d. values calculated by randomly sampling 100 times from each of the 225 operational taxonomic units (OTUs), respectively (see Methods and Supplementary Table 2). **b**, Unique mobile AMP resistance genes were involved in half as many between-species transfer events as antibiotic resistance genes ($P = 0.01$, two-sided negative binomial regression, $n = 19$ and 50 for AMP and antibiotic resistance genes, respectively). The result remains when a non-parametric test is used, which is less sensitive to outliers ($P = 0.02$, two-sided Mann–Whitney $U$ test). OTUs were generated as for **a** (Supplementary Table 3). On the x axes, the continuity of the scale breaks between 10 and 12. Above 12, only values with at least one transferred resistance gene are shown. **c**, Network representation of the mobile gene pool in the case of AMP and antibiotic resistance genes. Straight and curved lines represent genes that were shared between OTUs of different phyla and the same phylum, respectively. Line thickness represents the number of resistance genes shared between OTUs of six major phyla. Node size and numbers in parentheses indicate the number of OTUs in each phylum that shared at least one transferred AMP or antibiotic resistance gene. OTUs were generated as for **a**. **d**, Significantly fewer AMP resistance genes (46 out of 137) have a close homologue in naturally occurring plasmid sequences in the human microbiome as compared to antibiotic resistance genes (1,867 out of 2,163) ($P = 2.2 \times 10^{-6}$, two-sided Fisher's exact test). $n = 137$ and 2,163 for AMPs and ABs, respectively (see Methods and Supplementary Table 4). **e**, Plasmid-encoded homologues of individual AMP resistance genes were found in significantly fewer species in the human microbiome as compared to antibiotic resistance genes ($P = 5.6 \times 10^{-15}$, two-sided negative binomial regression, $n = 23$ for AMPs and $n = 1,772$ for antibiotics, see Methods).

As expected[25], the cultivation protocol allowed representative sampling of the gut microbiota: we could cultivate 65–74% of the gut microbial community at the family level in the absence of any drug treatment (Supplementary Fig. 7 and Supplementary Table 7). Next, the same faecal samples were cultivated in the presence of one of 5–5 representative AMPs and antibiotics (Supplementary Table 8). We applied drug dosages that retained 0.01–0.1% of the total cell

populations from untreated cultivations (Supplementary Table 8) and assessed the taxonomic composition of these cultures by 16S rRNA sequencing (see Methods).

Remarkably, the diversities of the AMP-treated and untreated bacterial cultures did not differ significantly from each other (Fig. 3a), despite marked differences in their taxonomic compositions (Fig. 3b). AMP-treated samples contained several bacterial
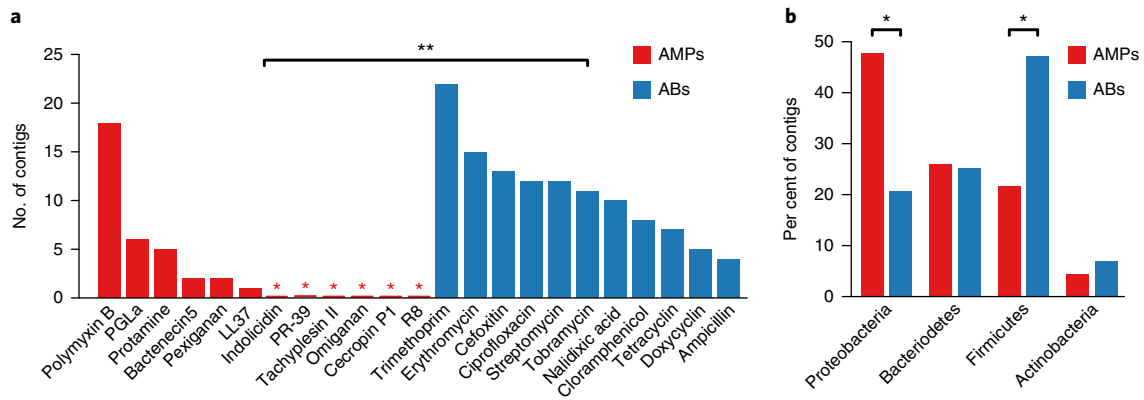
**Fig. 2 | In *E. coli*, short genomic fragments from the human gut microbiota confer AMP resistance less frequently than antibiotic resistance. a,** Functional selection of metagenomic libraries with 12 AMPs (red bars) resulted in fewer distinct resistance-conferring DNA contigs than with 11 conventional small-molecule antibiotics (ABs, blue bars). $P=0.002$ from two-sided negative binomial regression, $n=34$ AMP resistance contigs, $n=119$ antibiotic resistance contigs. Red asterisks indicate zero values. **Significant difference between AMPs and antibiotics. **b,** Phylum-level distribution (%) of the AMP resistance (red bars) and antibiotic resistance contigs (blue bars). In the case of AMPs, significantly more resistance contigs originate from the Proteobacteria phylum ($P=0.015$, two-tailed Fisher's exact test, $n=110$), while contigs originating from the Firmicutes phylum are under-represented ($P=0.033$, two-tailed Fisher's exact test, $n=110$). *Significant difference between AMPs and antibiotics for a given phylum.

**Table 1 | List of putative AMP resistance genes identified from our functional metagenomic screens**

| Resistance gene | Resistance gene function | Classification of resistance function | Estimated source organism | UniProt ID[d] | AMP |
|---|---|---|---|---|---|
| *lpxF*[a] | Lipid A 4'-phosphatase | Target alteration | *Parabacteroides merdae* | A7AEQ0 | Polymyxin B, pexiganan |
| *lpxF*[b] | Lipid A 4'-phosphatase | | *Parabacteroides* sp. | A0A355WPL4 | Polymyxin B, pexiganan |
| *lpxF*[c] | Lipid A 4'-phosphatase | | NA | A0A0F5J1Y0 | Polymyxin B |
| *arnTEF* | 4-amino-4-deoxy-ʟ-arabinose modification of lipid A | | *E. coli* MS 107-1 | P76473 Q47377 P76474 | Polymyxin B |
| *pmrAB (qseBC)* | Response regulator | Regulation | *Sutterella wadsworthensis* 2_1_59BFAA | A0A139KIR1 A0A139KIX7 | Polymyxin B |
| *pmrD* | Signal transduction protein | | *E. coli* MS 107-1 | P37590 | Polymyxin B |
| *ompT* | Protease 7 precursor | Agent inactivation | *E. coli* MS 196-1 | P09169 | Bactenecin 5, LL37, pexiganan |

These resistance genes could be functionally annotated based on a literature-curated catalogue of resistance genes (Supplementary Table 1). NA (not available) indicates that the donor organism could not be identified based on the Human Microbiome Project Database (see Methods). ᵃCharacterized in this work (Fig. 4b,c and Supplementary Fig. 11); annotated as undecaprenyl pyrophosphate phosphatase in Supplementary Table 6. ᵇAnnotated as *bcrC* in Supplementary Table 6. ᶜAnnotated as *ybjG* in Supplementary Table 10. ᵈClosest homologues of the metagenomic hits in UniProt database.

families from the Firmicutes and Actinobacteria phyla, which are phylogenetically distant from *E. coli* (Fig. 3c). Notably, exposure to AMP stress provided a competitive growth advantage to bacterial families that remained undetected in the untreated samples (Fig. 3c). The examples include Desulfovibrionaceae, Clostridiaceae and Eubacteriaceae. Desulfovibrionaceae is a clinically relevant bacterial family that is linked to ulcerative colitis[26]—an inflammatory condition with elevated AMP levels[27], while Clostridiaceae and Eubacteriaceae have key roles in maintaining gut homeostasis[28] (Fig. 3c). In sharp contrast, the diversity of antibiotic-treated cultures dropped significantly compared to both the untreated and AMP-treated cultures (Fig. 3a and Supplementary Fig. 8). Several bacterial families had a significantly lower abundance in the antibiotic-treated cultures than in the untreated ones (Fig. 3c). These results indicate that the human gut is inhabited by a large number of bacterial families across all major phyla in the gut that exhibit intrinsic resistance to AMPs.

**Human microbiota harbours a large reservoir of AMP resistance genes.** Next, we assessed if the high taxonomic diversity in the AMP-resistant microbiota corresponds to a diverse reservoir of AMP resistance genes. To this end, we annotated previously identified AMP- and antibiotic-resistance genes in a set of gut bacterial genomes[16] representing bacterial families that were detected in our culturing experiments following AMP and antibiotic selection, respectively (Fig. 3c; for details, see Methods). Remarkably, 65% of our literature-curated AMP resistance gene families (Supplementary Table 1) were represented in at least one of these genomes (Supplementary Table 2), which is similar to the value for antibiotic resistance gene families (58%). Finally, AMP resistance gene families, on average, were 39% more widespread in these species than the same figure for antibiotic resistance genes (Supplementary Fig. 9). Thus, the human gut harbours diverse AMP-resistant bacteria and a large reservoir of AMP resistance genes.
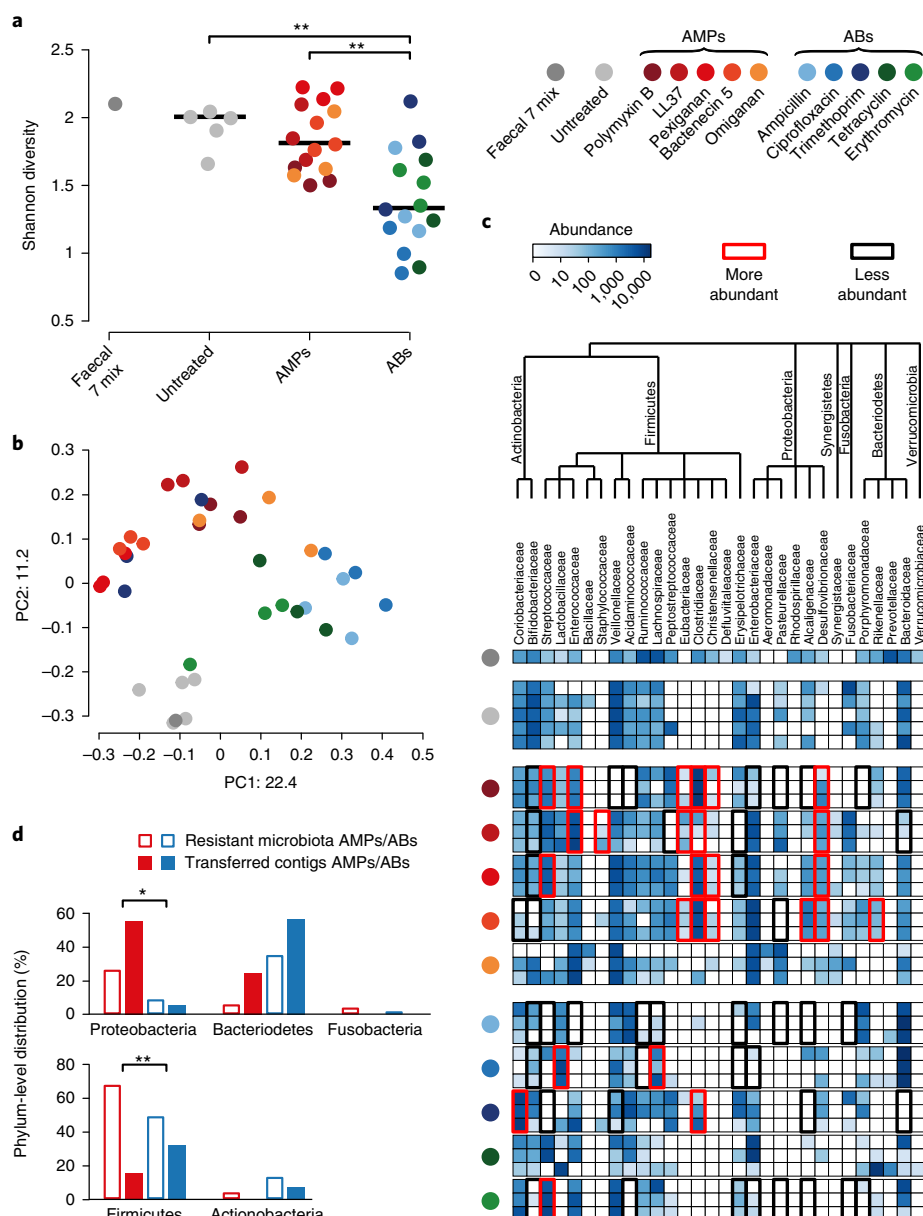
**Fig. 3 | Culturing reveals that short genomic DNA fragments from the gut microbiota have a limited potential to transfer AMP resistance to *E. coli*.**
**a**, Diversities of the cultured microbiota and the original faecal sample (Faecal 7 mix). Data represent Shannon alpha diversity indices at the family level based on 16S rRNA profiling of the V4 region. AMP/antibiotic (AB) treatments are colour-coded. Untreated samples were grown in the absence of any AMP or antibiotic. **Significant difference from two-sided Mann–Whitney $U$ test, $P = 0.005$ for untreated versus ABs, $P = 3 \times 10^{-4}$ for AMPs versus ABs. Sample sizes were 5, 15 and 15 for untreated, AMPs and ABs, respectively. Central horizontal bars represent median values. **b**, Principal coordinate analysis plot based on unweighted UniFrac distances[65], separating the AMP- and antibiotic-resistant and untreated microbial communities ($P = 0.001$, permutational multivariate analysis of variance test, $n = 36$). **c**, Differential abundance analyses between the untreated and AMP- and antibiotic-resistant microbiota at the family level (see Methods). Brackets depict a significant increase (red) or decrease (black) in abundance of a given family as a consequence of AMP or antibiotic (AB) treatment based on pairwise two-sided negative binomial tests (see Methods). Sample sizes were 5 and 3–3 for untreated and AMPs and ABs, respectively. The phylogenetic tree is based on the assignment of the NCBI Taxonomy database and created using the ete3 toolkit[34]. **d**, Phylum-level distributions of resistant gut bacteria and resistance DNA contigs originating from them. Compared to their relative frequencies in the drug-treated cultured microbiota (based on colony numbers), the phylogenetically close Proteobacteria contributed disproportionately more AMP resistance (red bars) than antibiotic resistance (AB, blue bars) contigs, whereas the opposite pattern was seen for the distantly related Firmicutes. Asterisks indicate significant interaction terms in logistic regression models: *$P = 0.018$ and **$P = 0.003$ for Proteobacteria and Firmicutes, respectively (for more details see Methods and Supplementary Table 11). Sample sizes were 22,651 and 24,336 for the total colony number of AMP-resistant and AB-resistant microbiota, respectively, and 33 and 54 for the number of transferred AMP-resistant and AB-resistant contigs, respectively.

**Phylogenetic constraints on the functional compatibility of AMP resistance genes.** We next directly tested whether the shortage of AMP resistance DNA fragments from distantly related bacteria can be explained by the low potential of genomic fragments to transfer AMP resistance phenotypes to *E. coli*. To this end, we constructed metagenomic libraries from the AMP- and antibiotic-resistant microbiota
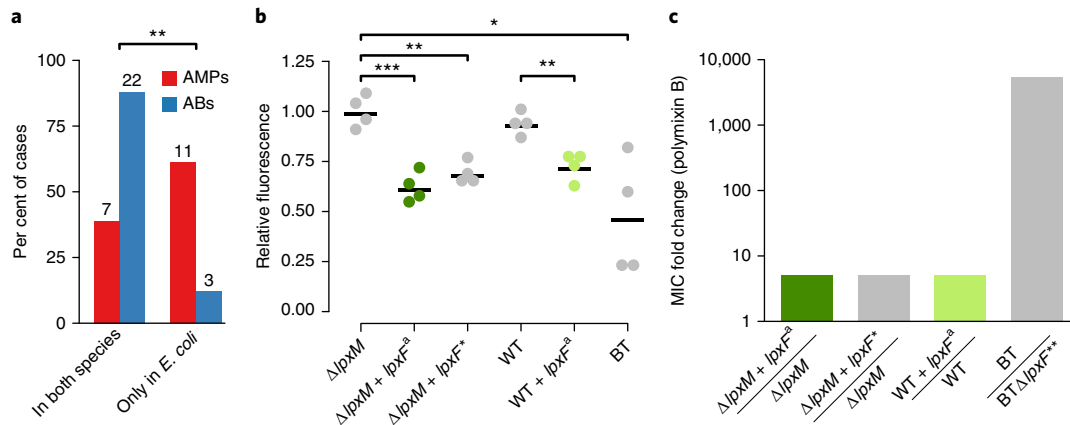
**Fig. 4 | AMP resistance DNA fragments provide host-dependent phenotypic effects. a**, A significantly lower proportion of AMP resistance DNA fragments (AMPs, red bars) conferred resistance in both *E. coli* and *S. enterica* compared to antibiotic resistance DNA fragments (ABs, blue bars), suggesting weaker between-species conservation of the AMP resistance phenotypes. Asterisks indicate significant difference, $P = 0.0011$, two-sided Fisher's exact test, $n = 16$ for AMPs, $n = 25$ for ABs. **b**, The *lpxF* orthologue from *P. merdae*, isolated in our screen (marked as lpxF[a] here and in Table 1; represented in green) and a previously characterized fully functional *lpxF* from *Francisella novicida*[61] (marked as *lpxF**) decrease the net negative surface charge of Δ*lpxM E. coli* to a similar extent, and close to the level of wild-type *Bacteroides thetaiotaomicron* (BT) expressing its native *lpxF*. LpxF[a] has a similar effect both in Δ*lpxM* and wild-type *E. coli* (dark and light green dots). The fluorescence signal is proportional to the binding of the fluorescein isothiocyanate (FITC)-labelled poly-ʟ-lysine polycationic molecules. Less poly-ʟ-lysine binding reflects a less negative net cell surface charge[62]. Significant differences: *$P = 0.03$, **$P = 0.001$ and ***$P = 0.0004$, respectively, Welch's two-sample *t*-test, $n = 4$ biological replicates; central horizontal bars represent mean values. Corresponding microscopic pictures are shown in Supplementary Fig. 11. **c**, LpxF[a] increases polymyxin B resistance of both Δ*lpxM* and wild-type *E. coli* only fivefold (dark and light green bars) ($n = 3$), to the same extent as LpxF from *F. novicida* (marked as *lpxF**) ($n = 3$). In contrast, lpxF in its original host, *B. thetaiotaomicron* (marked as *lpxF***) provides a 5,000-fold increment in polymyxin B resistance[13].

cultures. From each AMP and antibiotic treatment, two biological replicates were generated (see Methods), resulting in 10–10 libraries, covering 25.6 Gb and 14 Gb DNA, respectively (Supplementary Table 9). These metagenomic libraries were next screened on the corresponding AMP- or antibiotic-containing solid medium. Finally, the phylogenetic sources of the resulting AMP resistance and antibiotic resistance contigs were inferred with high confidence (Supplementary Fig. 10 and Supplementary Table 10). Compared to their relative frequencies in the drug-treated cultured microbiota, the phylogenetically close Proteobacteria contributed disproportionately more AMP resistance than antibiotic resistance DNA fragments, whereas the opposite pattern was seen for the distantly related Firmicutes (Fig. 3d and Supplementary Table 11). Taken together, phylogenetically diverse gut bacterial species show AMP resistance, but there is a shortage of transferable AMP resistance DNA fragments from phylogenetically distant relatives of *E. coli*.

**Pervasive genetic background dependence of AMP resistance genes.** Finally, we present evidence that DNA fragments that confer resistance to AMPs and were isolated from our screens show stronger genetic background dependence than those conferring resistance to antibiotics.

To systematically test the genetic background-dependency of AMP resistance genes, we examined how DNA fragments that provide AMP or antibiotic resistance in *E. coli* influence drug susceptibility in a related *Enterobacter* species, *Salmonella enterica*. We analysed a representative set of 41 resistance-conferring DNA fragments derived from our screens (Supplementary Table 12) by measuring the levels of resistance provided by them in both *E. coli* and *S. enterica*. Strikingly, while 88% of the antibiotic resistance DNA fragments provided resistance in both host species, only 38.9% of AMP resistance DNA fragments did so (Fig. 4a and Supplementary Table 12). Thus, the phenotypic effect of AMP resistance genes frequently depends on the genetic background, even when closely related hosts are compared.

As an example, we finally focused on a putative orthologue of a previously characterized AMP resistance gene, *lpxF* (ref. [12]). LpxF is a key determinant of AMP resistance in Bacteroidetes, a member of the human gut microbiota. By decreasing the net negative surface charge of the bacterial cell, it provides a 5,000-fold increment in polymyxin B resistance in these species[13,29]. To test the impact of one of the *lpxF* orthologues identified in our screen (denoted *lpxF[a]* in Table 1) on AMP resistance in a new bacterial species, we expressed it in wild-type and a mutant *E. coli* strain (Δ*lpxM*) as well. Similarly to Bacteroidetes, the Δ*lpxM* strain uniquely synthesizes penta-acylated lipid A molecules, the natural substrate of *lpxF*, and is therefore especially suitable for testing the impact of the isolated *lpxF* orthologue[30]. Reassuringly, surface charge measurements confirmed that the *lpxF[a]* is fully functional in *E. coli* (Fig. 4b and Supplementary Fig. 11). However, it provided a mere fivefold increase in polymyxin B resistance in *E. coli*, even in the Δ*lpxM* background (Fig. 4c). Notably, unlike the previously characterized *lpxF* from *F. novicida*[30], the *lpxF* orthologue isolated from our screen was active both on the penta- and hexa-acylated lipid A molecules (Fig. 4b,c). Therefore, substrate specificity alone is unlikely to limit transfer of *lpxF[a]* into *E. coli* or other Enterobacteriaceae species. The compromised resistance phenotype conferred by *lpxF[a]* in the new host shows that the function of other genes is also required to achieve the high AMP resistance level seen in the donor bacterium.

## Discussion

This work systematically investigated the mobility of AMP resistance versus antibiotic resistance genes in the gut microbiome. We report that AMP resistance genes are less frequently transferred between members of the gut microbiota than antibiotic resistance genes (Fig. 1). In principle, this pattern could be explained by at least two independent factors: shortage of relevant selection regimes during the recent evolutionary history of the gut microbiota and lack of functional compatibility of AMP resistance genes on transfer to a new host. We focused on testing the second possibility due to

its experimental tractability and relevance to forecast the mobility of resistance genes upon AMP treatment. In a series of experiments, we showed that phylogenetically diverse gut bacteria display high levels of AMP resistance (Fig. 3), yet the underlying resistance genes individually often fail to confer resistance upon transfer to *E. coli* (Figs. 2 and 3). Furthermore, we demonstrated that the AMP resistance conferred by 1.5–5 kb genomic fragments often depends on the genetic background of the recipient bacterium (Fig. 4). Together, these results support the notion that horizontal acquisition of AMP resistance is constrained by phylogenetic barriers owing to functional incompatibility with the new host cell[31].

An important issue is whether the simultaneous transfer of multiple AMP resistance genes carried by longer DNA segments is feasible and can provide resistance to recipient bacteria. While this is certainly a realistic possibility, bioinformatic analyses support the conclusions derived from the metagenomic screens: AMP resistance genes are relatively rare in the mobile gene pool and on plasmids in nature (Fig. 1).

We speculate that the large differences in functional compatibility between antibiotic resistance and AMP resistance genes might be caused by the latter being more often part of highly interconnected cellular subsystems, such as cell envelope biosynthesis pathways. We note that the compromised benefit may not be the only manifestation of functional incompatibility and the exclusive reason for the limited presence of AMP resistance genes in the mobile gene pool and on plasmids (Fig. 1). It is also plausible that some AMP resistance genes have severe deleterious side effects in the new host in addition to conferring a compromised resistance. For example, the introduction of *lpxF* into bacterial pathogens reduces virulence in mice, probably because it perturbs the stability of the bacterial outer membrane in enterobacterial species[32]. Future works should elucidate whether AMP resistance genes are especially prone to induce deleterious side effects compared to antibiotic resistance ones. Clearly, deciphering the biochemical underpinnings of functional incompatibility of AMP resistance genes remains an area for future research.

Our results also provide mechanistic insights into the functional capacity of AMPs to control the composition and stability of the gut microbiome over evolutionary timescales. Specifically, as phylogenetic barriers limit the horizontal transfer of AMP resistance mechanisms, the exact dosages and combinations of AMPs could prove to be critical for the long-term advantage of gut bacterial species involved in human health. Indeed, our work indicates that specific AMP stresses can lead to an increase in the amount of bacteria linked to ulcerative colitis (Desulfovibrionaceae).

Another important and unresolved issue is why natural AMPs that are part of the human innate immune system have remained effective for millions of years without detectable resistance in several bacterial species. One possibility, supported by our work, is that the acquisition of resistance through horizontal gene transfer from human gut bacteria is limited, most probably due to compromised functional compatibility in the recipient bacteria. We do not wish to claim, however, that AMPs in clinical use would generally be resistance-free. In agreement with the prevalence of polymyxin B resistance DNA fragments (Fig. 2a), a plasmid conferring colistin resistance is spreading globally[33]. Rather, our work highlights major differences in the frequencies and mechanisms of resistance across AMPs, with the ultimate aim to identify antimicrobial agents less prone to resistance.

## Methods

**Establishing a comprehensive AMP resistance gene data set.** Even though several databases have been created for antibiotic resistance genes, a comprehensive list of AMP resistance genes has not been compiled so far. We therefore carried out a systematic literature mining in PubMed NCBI and Google Scholar with the keywords 'antimicrobial peptide' + 'resistance'. From the identified publications, genes with experimentally confirmed influence on AMP susceptibility were

included in our manually curated AMP resistance gene data set (Supplementary Table 1). Altogether, 138 AMP resistance genes were identified. As a next step, the compiled AMP resistance genes were classified into resistance gene families (orthologous gene groups or orthogroups) using *EggNOG-mapper* software (version 0.12.7) on the bacterial EggNOG 4.5.1 database[34]. Then, AMP resistance genes were classified into broad functional categories analogous to the classification of antibiotic resistance genes in The Comprehensive Antibiotic Resistance Database (CARD)[35]. To obtain a comparable data set for known antibiotic resistance genes we downloaded the CARD database[35]. Genes associated AMP resistance in CARD were filtered out and the remaining antibiotic resistance genes from CARD were grouped into resistance gene families in the same way as AMP resistance genes using the EggNOG database (Supplementary Table 1).

**Analysis of the mobile gene pool of the gut microbiota.** A previously published mobile gene catalogue of the human gut microbiota[15] was analysed to compare the patterns of horizontal gene transfer events involving AMP and antibiotic resistance genes across a wide range of bacteria. This mobile gene catalogue relies on the identification of nearly identical genes in distantly related bacterial genomes and thereby provides a snapshot of the gene set subjected to recent horizontal gene transfer events in a representative sample of the gut microbiota. The goal in our analysis was to determine the presence/absence pattern of the AMP and antibiotic resistance genes not only in the mobile gene pool but also in the 567 genomes from which the mobile gene pool was derived. In this way, not only the horizontally transferred resistance genes were identified but also those that have not been detected in such transfer events, but were present in the gut microbiome. To this end, the genomes and proteomes used by Brito and colleagues[15] were downloaded from the Human Microbiome Project (HMP) database (https://www.hmpdacc.org/HMRGD/) and from the Fijicomp website (http://www.fijicomp.org). DNA sequences derived from the latter database were used for open reading frame (ORF) prediction with Prodigal software (version 2.6.3[36]). Then, a sequence similarity search was applied to the compiled list of proteins encoded in the analysed genomes to identify those that were present in the mobile gene pool as well. The sequence similarity search between the mobile gene pool and the proteins from the genomes was carried out with the BLASTX option of Diamond software (version 0.9.10[37]) with 50% sequence coverage and 100% sequence identity (parameters were chosen to reproduce the original publication of the mobile gene pool[15]). Of the 37,870 unique mobile genes in the mobile gene pool, we identified 37,184 in the genomes (98.28%).

Next, both the antibiotic and AMP resistance genes were identified among the mobile genes and among those that have not been detected in the mobile gene pool but were present in the genomes. For this functional annotation, a BLAST search was carried out against the antibiotic resistance genes from the CARD database with the BLASTP option of the Diamond software with strict parameters (*e*-value < 10[−5], >40% identity at the protein level and 80% query sequence coverage) (Supplementary Table 2). In a similar vein, AMP resistance genes were identified by performing a BLASTP sequence similarity search against the manually curated list of AMP resistance genes (Supplementary Table 1). Antibiotic and AMP resistance genes in our databases were classified into resistance gene families by the EggNOG-mapper software on the bacterial EggNOG database (Supplementary Table 2). For the annotated resistance gene list in the mobile gene pool, see Supplementary Table 3.

To compare the relative frequency of the AMP and antibiotic resistance genes in the mobile gene pool (Fig. 1a), we restricted the analysis to one genome per species. This was necessary to avoid sampling bias because different species were represented by unequal numbers of genomes in the data set. To this end, 16S rRNA sequences were determined for each genome (for HMP genomes they were downloaded from the Silva database[38], and for Fijicomp genomes they were identified directly in the genomes using RNAmmer software (version 1.2[39]). Then, genomes with fewer than 2% 16S rRNA gene dissimilarities were collapsed into genome groups ('species' or OTUs) using average linkage clustering as described in the publication of the mobile gene pool[15]. Each such genome group was represented by one randomly chosen genome for the statistical analysis presented in Fig. 1a (note that the random sampling was repeated 100 times, yielding an estimate of standard error). Resistance genes in the mobile gene pool that resulted in a BLAST hit both from the AMP and the antibiotic resistance databases were excluded from the analysis. The remaining resistance genes annotated in the mobile gene pool were counted and plotted as the percentage of the total number of annotated resistance genes in the genome sequences ('All bacteria' from Fig. 1a). Additionally, the analysis was separately performed for the Gram-negative and Gram-positive bacterial genomes as well ('Gram-positive' and 'Gram-negative' in Fig. 1a). For the network representation of the mobile gene pool, we used Cytoscape[40].

For each unique mobile AMP or antibiotic resistance gene we also estimated the minimum number of independent transfer events (Fig. 1b) by counting the number of genome groups (that is, OTUs) in which the gene is present in the mobile gene pool[16] (Supplementary Table 3).

**Identification of homologues of the literature-curated AMP resistance and antibiotic resistance genes in naturally occurring plasmid sequences and in ICEs.** For the identification of plasmid-encoded AMP resistance genes, we

used two independent databases: a plasmid-encoded protein database from the NCBI Reference Sequence Database (Refseq)[41] (ftp://ftp.ncbi.nih.gov/refseq/release/plasmid/) and a curated plasmid database containing 2,097 entire plasmid sequences from the Enterobacteriaceae bacteria[42] (https://figshare.com/s/18de8bdcbba47dbaba41). From the Refseq database, protein sequences were downloaded and both the antibiotic and AMP resistance proteins were identified with the BLASTP algorithm using our literature-curated lists of resistance genes. Hits were accepted only if they showed >40% sequence similarity over 80% of the length of the subject protein, with an $e$-value less than $10^{-5}$. Because this Refseq data set contains plasmid-encoded proteins from various sources in addition to the human microbiota, we filtered our hits using a previously compiled list of species from the human microbiota[43]. Thus, only those resistance gene hits were retained that are present on plasmids from human-associated bacteria. From the Enterobacteriaceae-specific plasmid database, we downloaded the translated nucleotide sequences for all six reading frames and carried out the similarity search as above using BLASTP with 80% query coverage, and >40% sequence identity. For this data set, we manually checked the presence of the plasmid-encoded resistance genes in the human microbiome using the NCBI database (Supplementary Table 4). Finally, we took the union of these two data sets to calculate the percentage of resistance genes residing on plasmids (Fig. 1d).

To estimate the species-level distribution of the plasmid-encoded resistance proteins we used the Refseq data set only as it gives information on the identity of species from which the plasmids were isolated. Specifically, for each AMP and antibiotic resistance gene that resulted in plasmid-encoded homologues, we counted the number of species that bore at least one plasmid-encoded homologue of the given AMP and antibiotic resistance gene (Fig. 1e).

For the identification of AMP resistance genes that are associated with ICEs, we used a database that contains 16,820 cargo genes associated with ICE sequences from human gut bacterial genome sequences (Intestinal Microbiome Mobile Element Database, https://immedb.org/)[44]. Nucleotide sequences were downloaded and, following translation of the sequences with the tBLASTn algorithm, both the antibiotic and AMP resistance proteins were identified using our literature-curated lists of resistance genes with the same BLAST parameters as before (>40% sequence similarity over at least 80% of the length of the subject protein). Because the literature-curated lists of resistance genes contain proteins from various sources in addition to the human gut microbiota, only those homologues were considered in the set of non-ICE-associated resistance genes that were detected in the genome sequences of the HMP database (Supplementary Table 4).

**Construction of gut metagenomic libraries.** To sample the gut resistome, we applied a previously established small-insert shotgun metagenomic protocol[11] with small modifications. This method identifies small genomic fragments that decrease drug susceptibility when random genomic fragments are expressed from a multicopy plasmid with an inducible promoter. For the construction of the metagenomic libraries, human stool samples were obtained from two healthy unrelated individuals who had not taken any antibiotics for at least one year before sample donation. Throughout the entire study we complied with all relevant ethical regulations. The protocol related to human faecal sample collection was approved by the Ethical Review Board of the Albert Szent-Györgyi Health Centre, University of Szeged (approval ID: 42/2017-SZTE). Written informed consent from each participant was obtained before faecal sample collection. Protocols related to human faecal sample handling were approved by the Ministry of Agriculture (Hungary) (approval ID: TMF/146-9/2017). Gut community DNA was isolated immediately after sample donation using the ZR Fecal DNA MiniPrep kit (Zymo Research) according to the manufacturer's instructions (http://www.zymoresearch.com/downloads/dl/file/id/91/d6010i.pdf). Subsequently, 10 μg of metagenomic DNA from each sample was partially digested with 0.25 U MluCI restriction enzyme (New England BioLabs) in 10× CutSmart Buffer (New England BioLabs) at 37 °C for 20 min, followed by heat inactivation at 85 °C for 20 min. MluCI is a four-base cutter restriction enzyme that produces overhangs complementary to the ones that EcoRI produces. By varying the incubation time or the enzyme concentration, the size range of the resulting DNA fragments can be set. The fragmented DNA was size selected by electrophoresis on a 1% (mass/volume) agarose gel in 1× Tris-acetate-EDTA (TAE) buffer. A gel slice corresponding to 1,500–5,000 bp was excised from the gel and DNA was isolated using a GeneJET Gel Extraction and DNA Cleanup Micro Kit (Thermo Scientific). pZErO-2 plasmid DNA (5 μg; ThermoFisher) was digested with 25 U EcoRI restriction enzyme (Fermentas) in 10× EcoRI Buffer (Fermentas) for 2 h, followed by 20 min heat inactivation at 65 °C. After purification with a DNA Clean & Concentrator-5 kit (Zymo Research), digested pZErO-2 plasmid was dephosphorylated with FastAP alkaline phosphatase (Thermo Scientific) as follows: 4 μg plasmid DNA was incubated with 4 U enzyme in 10× FastDigest buffer at 37 °C for 1 h, followed by 5 min heat inactivation at 74 °C and purification with DNA Clean & Concentrator-5 kit (Zymo Research). DNA was ligated into pZErO-2 at the EcoRI site using the Rapid DNA ligation kit (Thermo Scientific). The ligation reaction was performed in 15 μl total volume using a 5:1 insert–vector ratio: 4.5 μl (310 ng) digested and gel purified DNA insert, 0.65 μl (62 ng) EcoRI-cut pZErO-2 vector, 3 μl 5× ligation buffer, 0.75 μl 10 mM ATP, 4.1 μl dH₂O, 2 μl T4 DNA ligase (5 U μl⁻¹). The ligation mixture was incubated at 16 °C overnight, followed by heat inactivation at 65 °C for 10 min.

Before transformation, the ligation mixture was purified with DNA Clean & Concentrator-5 kit (Zymo Research). A 3.5 μl volume of the resulting ligation mixture was transformed by electroporation into 50 μl of electrocompetent E. coli DH10B cells (Invitrogen). Electroporation was carried out with a standard protocol for a 1 mm electroporation cuvette. Cells were recovered in 1 ml SOC medium, followed by 1 h incubation at 37 °C. A 500 μl volume of the recovered cells was plated onto square Petri dishes containing Luria Bertani (LB) agar supplemented with 50 μg ml⁻¹ kanamycin. To assess the library size (number of colony-forming units, c.f.u.), 1 μl of the electroporated cells was saved for plating on a separate Petri dish containing Luria Bertani (LB) agar supplemented with 50 μg ml⁻¹ kanamycin. From each plate, 10 clones were randomly picked for colony PCR to confirm the presence and the size distribution of the inserts. PCR was performed using the Sp6-T7 primer-pair (Supplementary Table 13) flanking the EcoRI site of the multiple cloning site of the pZErO-2 vector. The sizes of the PCR products were determined by gel electrophoresis and the average insert size was calculated to be 2–3 kb. The size of each library was determined by multiplying the average insert size by the number of total c.f.u.s. The size distributions of the libraries varied between 4.4 and 16 Gb coverage with this protocol, which is in line with a previously published state-of-the-art protocol[11,18]. The resulting colonies from the Petri dishes were collected and the plasmid library was isolated using an InnuPREP Plasmid Mini Kit (Analytic Jena). Between 30 and 60 ng of isolated plasmid library was transformed by electroporation into 40 μl electrocompetent E. coli BW25113 (prepared as described in ref. [44]). This E. coli strain was used for the functional selections (see next section). After electroporation, cells were recovered in 1 ml of LB medium for 1 h at 37 °C. Special care was taken to achieve high electroporation efficiency to cover 10–100 times the original library size. In this way, we ensured that most library members were electroporated from the plasmid library. The 1 ml recovered cell culture was added to 9 ml of LB medium supplemented with 50 μg ml⁻¹ kanamycin, and grown at 37 °C for 2–3 h until it reached the $7.5 \times 10^8$ to $10 \times 10^8$ cell density (optical density at 600 nm (OD$_{600}$ of 1.5–2). Cell aliquots were frozen in 20% glycerol and kept at −80 °C for subsequent functional selection experiments.

Metagenomic libraries were generated from the uncultured microbiota (the total DNA extracted from the stool samples) and also from the cultured microbiota (the genomic DNA extracted from the cultured pooled microbiota). For details see the section 'Cultivation of the gut microbiota under anaerobic conditions and DNA extraction'.

**Functional metagenomic selections for AMP resistance and antibiotic resistance genes.** Functional selections for resistance were carried out on solid plates containing one of the 12 antimicrobial peptides or 11 antibiotics (Supplementary Table 5). Instead of the plating assay commonly used in the field[11], we applied a modified gradient plate assay[45] where bacteria are exposed to a concentration gradient of the antimicrobial instead of a single concentration. We found that this strategy improves the reproducibility of AMP selections, where changes in the resistance levels are relatively small compared to that in the case of antibiotics. The growth medium in these plates was a modified minimal salt medium (MS) with reduced salt concentration (1 g (NH₄)₂SO₄, 3 g KH₂PO₄, 7 g K₂HPO₄, 100 μl MgSO₄ (1 M), 540 μl FeCl₃ (1 mg ml⁻¹), 20 μl thiamine (50 mg ml⁻¹), 20 ml casamino acids (BD) (10% (mass/volume), 5 ml glucose (40% (mass/volume)) in a final volume of 1 l), because most AMPs are not effective in vitro in the presence of high salt concentrations. In the case of the AMP-containing plates, the solidifying agent was changed to 1.5% (mass/volume) low melting point agarose (UltraPure LMP Agarose, Invitrogen) from 1.5% (mass/volume) agar to prevent any heat-induced structural damage of the peptides during plate pouring. Antibiotics and AMPs were purchased from Sigma and ProteoGenix, respectively. Onto each of the gradient plates (Tray plates, SPL Life Sciences) $2 \times 10^8$ cells were plated out from the thawed stocks of E. coli BW25113 bearing the metagenomic plasmid libraries. In this way, each metagenomic library member was represented about 10–100 times on each plate. We found this necessary for good reproducibility of our experiments. Subsequently, plates were incubated at 30 °C for 24 h. For each functional selection, a control plate was prepared with the same number of E. coli BW25113 plated out. These cells contained the pZErO-2 plasmid with a random metagenomic DNA insert that has no effect on AMP and antibiotic resistance. This control plate showed the minimum inhibitory concentration (MIC) of the antimicrobial without the effect of a resistance plasmid. The empty plasmid was not applicable as a control because in the absence of a DNA insert the CcdB toxic protein is expressed from the plasmid. To isolate the resistant clones from the library plates, sporadic colonies were identified above the MIC level (defined using the control plate) by visual inspection. These clones were collected by scraping them into 2 ml LB broth and then stored at −80 °C.

**Validation of the resistance-conferring metagenomic DNA fragments.** Following selection of the metagenomic libraries, the putative resistance phenotypes conferred by the plasmid selections were confirmed for a representative fraction of the colonies. From each selection at least 20 colonies were picked and the MIC increase was determined by a standard broth microdilution method[46], as described in the section 'Quantification of the resistance gains that metagenomic

contigs provide'. For these measurements, the same control plasmid was used as in the functional selections. MIC values were determined after 16–24 h of incubation at 30 °C with continuous shaking at 240 r.p.m. Plasmids from validated resistant clones were re-transformed into the BW25113 *E. coli* strain, followed by a second MIC determination to exclude the possibility that the increase in the MIC was induced by genomic mutations. Plasmids not showing a MIC increase in the validation protocol were excluded from further analysis. Predominantly, clones situated closer than 1 cm to the MIC level on the gradient plates did not confer resistance during validation. To avoid these false positive resistance plasmids, colonies at the borderline were not collected for further analysis without confirming their phenotype. The rest of the colonies were collected by scraping them into 2 ml of LB broth. Bacterial samples were stored at −80 °C in 20% (mass/ volume) glycerol. When only a few clones were on the plates, all were tested for resistance to make sure that we did not lose potential hits. We encountered such situations only in the case of AMPs. If the number of resistant clones on a plate was less than or equal to 20, plasmids were isolated individually from the MIC validated clones and sent for Sanger bidirectional sequencing with the Sp6-T7 primer pair (Supplementary Table 13). When the resistance-conferring insert was longer than covered by the initial sequencing, we applied a primer walking strategy to sequence the middle part of the insert.

**Amplification of the resistance-conferring metagenomic DNA fragments.**
Plasmid pools from the scraped resistant clones were PCR amplified for subsequent PacBio sequencing[47]. To this end, the plasmid pools were first isolated from each metagenomic selection using an InnuPREP Plasmid Mini Kit (Analytic Jena). Then, these plasmid pools served as templates for subsequent PCR amplification reactions to amplify the inserts from the pooled plasmids. These PCR reactions were performed with barcoded Sp6 and T7 specific primers as forward and reverse primers, respectively. A 16-base PacBio barcode dual-end labelling scheme was used to label each plasmid pool for sample identification in subsequent PacBio sequencing. The primer sequences are presented in Supplementary Table 13. PCR reactions consisted of 30 ng of template DNA, Q5 Hot Start High-Fidelity 2× Master Mix (New England BioLabs) and 0.2 µM barcoded primers in a final reaction volume of 25 µl. Following an optimization process, the number of PCR cycles was reduced to 15 to minimize amplification bias. The following thermocycler conditions were used: 98 °C for 5 min, 15 cycles of 98 °C for 15 s + 69 °C for 30 s + 72 °C for 2 min, and 72 °C for 7 min. The amplified metagenomic inserts were then cleaned using the DNA Clean & Concentrator-5 kit (Zymo Research) and their concentration was measured with a Qubit fluorometer (Invitrogen). To get rid of the short amplicons (for example, primer dimers), which may interfere with the sequencing process, barcoded amplicons were mixed at an equimolar ratio and the sample was gel extracted following electrophoreses using 1% agarose gel. The sample was purified (Zymoclean Gel DNA Recovery Kit) before sequencing.

**PacBio sequencing and data analysis.** Sequences of the pooled PCR products were obtained from the Norwegian Sequencing Centre at the University of Oslo, Norway. The library was prepared using the Pacific Biosciences Amplicon library preparation protocol. Samples were sequenced with a Pacific Biosciences RS II instrument using P6-C4 chemistry and MagBead loading in one SMRT cell. A total of 61,641 reads were obtained with a mean length of 20,961 bp. Reads were filtered without demultiplexing using the *RS_subreads.1* pipeline of the SMRT Portal (version 2.3.0) using default settings (no. of passes, 1; minimum accuracy, 0.9). Following barcode detection and demultiplexing, reads were collapsed to consensus sequences using the long amplicon analysis pipeline of the SMRT Portal with default settings. We validated our sequencing effort on a mock sample containing nine previously sequenced DNA contigs originating from our metagenomic selections. Reassuringly, of the nine sequences, eight were present in the PacBio sequencing result with at least 99% sequence identity. The single non-detected contig was the longest (4,500 bp), which may indicate a bias of the pipeline towards shorter insert sizes.

**Functional annotation and resistance gene identification on the metagenomic contigs.** To functionally analyse the ORFs on the assembled contigs from the metagenomic selections, ORFs were predicted and annotated using the Prokka suite (version 1.11[48]) with bacterial prediction settings and an *e*-value threshold of 10[−5]. Within Prokka, the *Prodigal*[36] sub-script was modified to run without the *-c* parameter to identify highly probable ORFs, even if the ends were not closed. This was necessary because some contigs may have been shortened by a few residues during the cloning process or in the assembly due to low coverage, without losing their functionality. Other parameters were kept as default. Next, ORFs on the metagenomic contigs were functionally annotated with our resistance gene lists introduced in the section 'Analysis of the mobile gene pool of the gut microbiota'. Specifically, an ORF was classified as an antibiotic resistance gene when a sequence similarity search using BLASTP[49] against the CARD[8] database[35] resulted in an annotation with an *e*-value < 10[−5], identity > 30% and coverage > 80%. Here, the minimum sequence identity was lower than in the case of the analysis of the mobile gene pool, because the experimentally observed resistance phenotype provided extra confidence for the annotation. Similarly, AMP resistance genes

were identified by performing a BLASTP sequence similarity search against the manually curated list of AMP resistance genes (Supplementary Table 1). We note that 55.4% of the antibiotic resistance contigs carried a homologue of a known antibiotic resistance gene, and 23.5% of the 34 unique AMP resistant DNA contigs encoded a gene that has been associated with AMP resistance in the literature. This difference suggests that more AMP resistance genes await discovery. To estimate the identity of the donor organisms from which the assembled DNA contig sequences originated, a nucleotide sequence similarity search was carried out for the entire DNA contigs as query sequences against the genome sequences from the HMP database[24] using BLASTN, with an *e*-value threshold of <10[−10]. Taxonomy was assigned with the ete3 toolkit[34].

**Cultivation of the gut microbiota under anaerobic conditions and DNA extraction.** To compare the abundance and phylogenetic distribution of the AMP- and antibiotic-resistant gut bacteria, we applied a recently established anaerobic cultivation protocol of the human gut microbiota, with small modifications[25]. For this purpose, human faecal samples were obtained from seven healthy unrelated volunteers who had not taken any antibiotics for at least one year before sample donation. Ethical rules were observed throughout the whole study. Following defecation, stool samples were immediately placed into uncapped 50 ml plastic tubes (Sarstedt), deposited in anaerobic bags (Oxoid, Thermo Scientific); samples were transferred into the anaerobic chamber within 1 h after sample collection. All anaerobic experiments were performed in a Bactron IV anaerobic chamber (Shel Lab) filled with an atmosphere of 95% nitrogen and 5% hydrogen, with palladium catalysts. Two grams of the faecal samples were suspended in 20 ml of modified Gifu anaerobic medium (GAM) broth (HyServe). After 10 min of incubation, allowing the solid particles to settle down, the supernatants were supplemented with 20% glycerol, aliquoted and stored at −80 °C. Before cultivation of the microbiota, thawed aliquots from the samples of the seven individuals were combined in an equal ratio (we refer to this sample mix as 'Faecal 7 mix' sample) in the anaerobic chamber. This Faecal 7 mix sample was plated out anaerobically in the presence and absence of one of the five AMPs or one of the five antibiotics that were active in the culturing medium (Supplementary Table 5). The culturing medium was modified GAM, given that the best reconstruction of the composition and architecture of the human gut bacterial community could be obtained using this medium[25]. In the case of AMP-containing plates, the solidifying agent was low melting agarose instead of agar for the same reason as before (see section 'Functional metagenomic selections for AMP and antibiotic resistance genes'). Each AMP and antibiotic was applied in three different concentrations on separate plates with three replicates.

To mimic the thermal conditions encountered by the gut bacteria in the human intestinal tract, plates were incubated at 37 °C in the anaerobic chamber for 4 days. After a 4 day time interval, we observed a plateau in the number of colonies appearing. Following colony counts the plates had to fulfil two requirements to be selected for subsequent analysis. First, colony numbers needed to be in the range of 0.01–0.1% as compared to the colony numbers in the absence of any AMP and antibiotic treatment (we refer to these samples as 'untreated'). Second, colony numbers needed to be high, but still in the countable range (1,000–5,000 colonies). To be in this range, the untreated Faecal 7 mix sample was plated out in a concentration 1,000-fold more dilute than used for the AMP and antibiotic selections. For Trimethoprim the selection pressure could not be increased enough to select for the most resistant 0.01–0.1% of the population because 0.49% of the colonies were able to grow even at the solubility limit of this antibiotic. From the selected plates, resistant colonies were scraped and the pooled colonies from each plate were used to isolate genomic DNA with a ZR Fecal DNA MiniPrep kit (Zymo Research) according to the manufacturer's instructions. The isolated DNA samples were subsequently used for both small-insert shotgun library constructions (referred to as libraries from cultured microbiota, Supplementary Table 10) and for 16S rRNA-based community profiling. To estimate the relative resistance level of the gut microbiota compared to *E. coli* BW25113, we used the colony counts from the anaerobic cultivation experiments at each AMP and antibiotic concentration; that is, the resistance level of the gut microbiota against an AMP and antibiotic is defined as the AMP and antibiotic concentration at which only 0.1–0.01% of the untreated microbiota population could grow, analogous to MIC value determination for a single species. The susceptibility of *E. coli* was estimated in the same way as for the gut microbiota by plating out anaerobically the same number of cells as in the case of the Faecal 7 mix sample for each AMP and antibiotic treatment. Then we determined the AMP and antibiotic concentrations at which only 0.1–0.01% of the *E. coli* cells could grow. The relative resistance level of the gut microbiota was defined as the resistance level of the gut microbiota divided by the resistance level of *E. coli* (Supplementary Table 8).

**16S rRNA-based bacterial community profiling.** To determine the taxonomic composition of the AMP- and antibiotic-resistant gut bacterial communities, we sequenced and analysed the V4 region of the 16S rRNA genes from the Faecal 7 mix samples cultivated in the presence of individual AMPs and antibiotics. We also determined the phylogenetic composition of the uncultivated Faecal 7 mix sample. The V4 region of the 16S rRNA gene was PCR amplified with dual-indexed Illumina primer pairs, using different combinations of barcoded forward and reverse primers (v4.SA501-505 and v4.SA701-706, respectively,

Supplementary Table 13) as previously described[50]. The primers consist of the appropriate Illumina adapters, an 8 nt index sequence, a 10 nt pad sequence, a 2 nt linker and specific sites for the V4 region. The PCR reactions consisted of 1.5 μl (30 ng) of template DNA, 10 μl of Phusion HF buffer (Thermo Fisher Scientific), 4 μl of 2.5 mM deoxynucleotide triphosphates mix (dNTPs), 0.5 μl of Phusion DNA polymerase (2 U μl⁻¹) (Thermo Fisher Scientific), 1–1 μl of primers, 10 μM each, 3 μl DMSO (100%) and 29 μl of nuclease-free $H_2O$ in a final reaction volume of 50 μl. The following thermocycler conditions were used: 95 °C for 2 min, 25 cycles of 95 °C for 20 s + 56 °C for 15 s + 72 °C for 5 min and 72 °C for 10 min. Following gel electrophoreses of the PCR products, the 400 bp amplicons were extracted from the gel (Thermo Scientific GeneJET Gel Extraction Kit) and, following a second purification step (Zymo Research DNA Clean & Concentrator-5 Kit), were sequenced using the MiSeq Illumina platform. To prepare the samples for sequencing, the amplicons were quantified using a fluorometric method (Qubit dsDNA BR Assay Kit, Thermo Fischer Scientific) and libraries were mixed with Illumina PhiX in a ratio of 0.95:0.05. Sequencing on the Illumina MiSeq instrument was carried out with a v2 500 cycle sequencing kit (Illumina). 100 μM stock custom sequencing primers[49] were mixed with standard read1, index read and read2 sequencing primers included in the MiSeq cartridge.

After sequencing, 16S rRNA reads were demultiplexed and processed with the Mothur software (version 1.36.1[51]). There were 21,979 average counts per sample. To filter out the low read counts we followed the protocol of ref. [25]. The number of sequences per sample was equalized to 20,000 read counts using random re-sampling with a custom R script. Reassuringly, 20,000 read counts is well above the threshold where phylogenetic diversities show saturation in the samples (see the rarefaction curves of samples in Supplementary Fig. 12, which were calculated using the vegan (version 2.4–3[52]) R package). Sequences were merged at the level of 97% sequence identity and taxonomically assigned using the Silva ribosomal RNA database[38]. OTUs were classified at the family level because the V4 region allows accurate identification only down to this level[53] (Supplementary Table 7). After removal of reads that could not be classified, 362 OTUs remained. To evaluate the reproducibility of the cultivation and sequencing, we generated seven replicates from the untreated samples. Samples referred to as 'Untreated 1–5' originate from independent cultivation experiments started from different aliquots of the same five frozen samples (Supplementary Figs. 7 and 12). In the case of 'Untreated 5, 5i and 5ii' samples, cultivations were started from the same sample and cultures were grown in parallel (Supplementary Figs. 7 and 12).

To quantify within-sample diversity from 16S rRNA data, we used the vegan R package to calculate the most commonly used alpha diversity indices[54,55] (Shannon index, Fig. 3a; Fisher's and inverse Simpson indices, Supplementary Fig. 8). Unweighted Unifrac distances (Fig. 3b) were computed with Phyloseq (version 1.22.3 R package[56]).

To identify differentially abundant bacterial families in the resistance microbiota, we applied *edgeR* (version 3.16.5 R package[57]) on the family-level 16S rRNA abundance data of the cultured microbiota samples, as suggested previously[58]. To this end, abundances were normalized using the TMM (trimmed mean of M-values) method[59] and then untreated and AMP- and antibiotic-treated samples were compared using negative binomial tests in a pairwise manner. We used the Benjamini–Hochberg false discover rate correction method to correct the *P* values for multiple testing[60].

**Comparing the prevalence of AMP and antibiotic resistance genes in gut microbial genomes.** To assess if the large taxonomic diversity in the AMP-resistant microbiota corresponds to a diverse reservoir of AMP resistance genes, we compared the prevalence of previously described AMP and antibiotic resistance gene families in a representative set of gut microbial genomes as follows. Genome sequences used for the analysis of the mobile gene pool (Supplementary Table 2) were filtered to retain only those sequences that belong to one of the AMP- and antibiotic-resistant bacterial families (Supplementary Table 7). In this analysis, we used only the genome sequences from the Human Microbiome Project, because the family-level taxonomic assignments of the genome sequences were sporadic in the Fijicomp cohort. In the remaining data set 24 AMP- and 26 antibiotic-resistant bacterial families were represented with 222 and 619 genome sequences, respectively. To avoid sampling bias due to the unequal representation of species among the genome sequences, one genome sequence was randomly picked from each genome group (genomes with lower than 2% 16S rRNA gene dissimilarities; for details see section 'Analysis of the mobile gene pool of the gut microbiota'). Then, for each known AMP and antibiotic resistance gene family, the genomes with at least one positive annotation were counted and plotted (Supplementary Fig. 9).

**Comparing phylum-level distributions of the resistant microbiota and the transferring resistance contigs.** To compare the phylum-level distributions of resistant gut bacteria and the resistance contigs originating from them (Fig. 3d), we carried out logistic regression analyses on count data (Supplementary Table 11) as follows. For each phylum, we fitted a logistic regression model to predict if a resistant gut microbiota colony or resistance-conferring contig belonged to that particular phylum or to the other phyla. Thus, each entry in the data set represented either a colony from the drug-treated cultivation experiment or a resistance contig detected in the functional metagenomics screen. The predictor

variables of the models were (1) whether the entry was a resistance contig or a cultivated colony, (2) the type of treatment (AMP or antibiotic) and (3) the interaction of these two variables. As we were interested in whether a given phylum contributed disproportionally more (or less) AMP than antibiotic resistance contigs compared to its relative frequency in the drug-treated cultured microbiota, we tested if the interaction term of the logistic regression model was significant (using the *glm* function of R).

**Comparison of resistance levels in *E. coli* and *S. enterica*.** To investigate whether the level of resistance provided by AMP resistance genes depends more on the genetic background of the recipient bacterium than in the case of antibiotic resistance genes, we measured how DNA fragments that provide AMP or antibiotic resistance to *E. coli* influence susceptibility in a related *Enterobacter* species, *S. enterica* serovar Typhimurium LT2. For this purpose, we used a representative set of plasmids carrying resistance DNA fragments that were isolated in our screens from AMP and antibiotic treatments (Supplementary Table 12). Special care was taken to avoid the inclusion of multiple plasmids carrying resistance genes with the same function or bias toward certain AMPs and antibiotics. To this end, from each AMP and antibiotic selection we chose 1–5 plasmids carrying resistance genes with different functions. This resulted in 16 plasmids and 25 AMPs and antibiotics. The provided resistance levels (MIC fold changes) were measured for both species with standard micro-dilution assays, in the same way as in the next section.

**Quantification of the resistance gains that metagenomic contigs provide.** To investigate the resistance gains that contigs provide for the recipient bacteria against AMP and antibiotic treatments, we carried out MIC measurements with selected contigs from the uncultured and cultured microbiota. The resistance gains were quantified by measuring MIC values with a standard broth microdilution method[61]. Briefly, 5,000 *E. coli* and *S. enterica* cells grown overnight in MS medium with 50 μg ml⁻¹ kanamycin were used to inoculate the wells of a 96-well plate. Three rows on the plate were inoculated with the strain in question and three rows with control cells. As a control, *E. coli* BW25113 or *S. enterica* serovar Typhimurium LT2 strains carrying the control plasmid were used as in the functional metagenomic selection experiments. Before inoculation, each well on the plate was pre-filled with 100 μl modified MS medium supplemented with the proper concentration of AMP or antibiotic. On the plate, each AMP and antibiotic was represented in 11 different concentrations (three wells per concentration per clone or control). Three wells per genotype contained only medium to check the growth in the absence of any antimicrobial. MIC values were determined by measuring $OD_{600}$ after 16–24 h of incubation at 30 °C with continuous shaking at 240 r.p.m. To minimize potential evaporation of the medium and consequent plate edge effects, 96-well plates were incubated at 30 °C following standard laboratory procedures for MIC measurements.

**Functional analysis of a putative LpxF from the metagenomic selections.** The aim of the functional characterization was twofold: (1) to support the functional prediction for the identified LpxF orthologues with a biochemical assay; (2) to estimate quantitatively the extent to which these LpxF orthologues reduce the net negative surface charge of the bacterial cell as compared to a previously characterized LpxF from *F. novicida* that removes >90% of the 4′-phosphate groups from the penta-acylated lipid A molecules and hence alters the charge of the outer membrane[62]. To estimate the surface charge of bacterial cells, we used a fluorescein isothiocyanate (FITC)-labelled poly-L-lysine (PLL) (FITC–PLL; Sigma-Aldrich) based assay. PLL is a polycationic molecule, widely applied to study the interaction between charged bilayer membranes and cationic peptides[63]. The following strains were used in this measurement: *E. coli* BW25113 (WT), *E. coli* BW25113 carrying the pZErO-2 plasmid with the LpxF orthologue from *P. merdae* (denoted *lpxF*[a] in Table 1), identified in our functional selection experiments (WT + *lpxF*), *E. coli* BW25113 Δ*lpxM* (Δ*lpxM*), *E. coli* BW25113 Δ*lpxM* carrying the pZErO-2 plasmid with the LpxF orthologue from *P. merdae* (Δ*lpxM* + *lpxF*), *E. coli* BW25113 Δ*lpxM* carrying the pWSK29 plasmid with LpxF from *F. novicida*[62] (Δ*lpxM* + *lpxF*\*) and *B. thetaiotaomicron*. We measured the phenotypic effect of both LpxF carrying plasmids on Δ*lpxM* genetic background, because LpxF from *F. novicida* cannot carry out its biological function in wild-type *E. coli*, only when the lipid A molecules are tetra- or penta-acylated as in the case of Δ*lpxM* *E. coli*[62]. *B. thetaiotaomicron* intrinsically expresses the BT1854 *lpxF* orthologue (lpxF\*\*), which is responsible for the high level of resistance of this strain against polymyxin B[12]. Before the surface charge measurements, cells were grown overnight in tryptone yeast extract glucose (TYG) medium[64] under anaerobic conditions. We grew all the strains in TYG medium to allow comparability with *B. thetaiotaomicron*. Cells were washed twice with phosphate-buffered saline (PBS) then resuspended to a cell density of $OD_{600} = 1$. Cells were incubated with 2 μl of 5 mg ml⁻¹ FITC–PLL and 100 μl of 1 μg ml⁻¹ 4,6-diamidino-2-phenylindole (DAPI) for 10 min, followed by centrifugation (4,500 r.p.m., 5 min). DAPI was used to identify the live cells. Cells were washed twice with PBS then diluted 100-fold in 100 μl of PBS and transferred into a black clear-bottomed 96-well microplate (Greiner Bio-One). Before fluorescent microscopy analysis, cells were collected to the bottom of the plate by centrifugation (4,500 rpm, 10 min). Pictures were taken with a PerkinElmer Operetta microscope using a ×60 high-NA objective to

visualize the cells. Images of two channels (DAPI and FITC–PLL) were collected from ten sites for each well. Mean fluorescent intensity for each well was calculated using Harmony High Content Imaging and Analysis Software. Experiments were carried out in four biological replicates.

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

## Data availability

The GenBank accession nos. for the PacBio sequencing data are MH883365–MH883616. 16S rRNA sequencing reads are available from the Sequence Read Archive (SRA) (BioProject PRJNA494380). All data generated or analysed during this study are included in this article and its Supplementary Information. For each figure, the availability of the analysed data is indicated in the legend. All accession numbers with information on the associated samples are provided as Supplementary Data.

## References

1. Dethlefsen, L., McFall-Ngai, M. & Relman, D. A. An ecological and evolutionary perspective on human–microbe mutualism and disease. *Nature* **449**, 811–818 (2007).
2. Nicholson, J. K. et al. Host–gut microbiota metabolic interactions. *Science* **336**, 1262–1267 (2012).
3. Littman, D. R. & Pamer, E. G. Role of the commensal microbiota in normal and pathogenic host immune responses. *Cell Host Microbe* **10**, 311–323 (2011).
4. Round, J. L. & Mazmanian, S. K. The gut microbiota shapes intestinal immune responses during health and disease. *Nat. Rev. Immunol.* **9**, 600–600 (2009).
5. Zasloff, M. Antimicrobial peptides of multicellular organisms. *Nature* **415**, 389–395 (2002).
6. Peschel, A. & Sahl, H. G. The co-evolution of host cationic antimicrobial peptides and microbial resistance. *Nat. Rev. Microbiol.* **4**, 529–536 (2006).
7. Hancock, R. & Patrzykat, A. Clinical development of cationic antimicrobial peptides: from natural to novel antibiotics. *Curr. Drug Target Infect. Disord.* **2**, 79–83 (2002).
8. Hancock, R. E. W. & Sahl, H.-G. Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. *Nat. Biotechnol.* **24**, 1551–1557 (2006).
9. Kubicek-Sutherland, J. Z. et al. Antimicrobial peptide exposure selects for *Staphylococcus aureus* resistance to human defence peptides. *J. Antimicrob. Chemother.* **72**, 115–127 (2017).
10. Andersson, D. I., Hughes, D. & Kubicek-Sutherland, J. Z. Mechanisms and consequences of bacterial resistance to antimicrobial peptides. *Drug Resist. Updat.* **26**, 43–57 (2016).
11. Sommer, M. O. A., Dantas, G. & Church, G. M. Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science* **325**, 1128–1131 (2009).
12. Cullen, T. W. et al. Gut microbiota. Antimicrobial peptide resistance mediates resilience of prominent gut commensals during inflammation. *Science* **347**, 170–175 (2015).
13. 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).
14. Chen, L. et al. Newly identified colistin resistance genes, *mcr-4* and *mcr-5*, from upper and lower alimentary tract of pigs and poultry in China. *PLoS ONE* **13**, e0193957 (2018).
15. Brito, I. L. et al. Mobile genes in the human microbiome are structured from global to individual scales. *Nature* **535**, 435–439 (2016).
16. Jain, R., Rivera, M. C. & Lake, J. A. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl Acad. Sci. USA* **96**, 3801–3806 (1999).
17. Cohen, O., Gophna, U. & Pupko, T. The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. *Mol. Biol. Evol.* **28**, 1481–1489 (2011).
18. Forsberg, K. J. et al. Bacterial phylogeny structures soil resistomes across habitats. *Nature* **509**, 612–616 (2014).
19. van der Helm, E. et al. Rapid resistome mapping using nanopore sequencing. *Nucleic Acids Res.* **45**, e61 (2017).
20. Pehrsson, E. C. et al. Interconnected microbiomes and resistomes in low-income human habitats. *Nature* **533**, 212–216 (2016).
21. Kaye, K. S., Pogue, J. M., Tran, T. B., Nation, R. L. & Li, J. Agents of last resort. *Infect. Dis. Clin. North Am.* **30**, 391–414 (2016).
22. Zhi, C., Lv, L., Yu, L.-F., Doi, Y. & Liu, J.-H. Dissemination of the *mcr-1* colistin resistance gene. *Lancet Infect. Dis.* **16**, 292–293 (2016).
23. Dürr, U. H. N., Sudheendra, U. S. & Ramamoorthy, A. LL-37, the only human member of the cathelicidin family of antimicrobial peptides. *Biochim. Biophys. Acta* **1758**, 1408–1425 (2006).
24. Methé, B. A. et al. A framework for human microbiome research. *Nature* **486**, 215–221 (2012).
25. Rettedal, E. A., Gumpert, H. & Sommer, M. O. A. Cultivation-based multiplex phenotyping of human gut microbiota allows targeted recovery of previously uncultured bacteria. *Nat. Commun.* **5**, 4714 (2014).
26. Rowan, F., Docherty, N. & Murphy, M. *Desulfovibrio* bacterial species are increased in ulcerative colitis. *Dis. Colon* **53**, 1530–1536 (2010).
27. Wehkamp, J. et al. Inducible and constitutive β-defensins are differentially expressed in Crohn's disease and ulcerative colitis. *Inflamm. Bowel Dis.* **9**, 215–223 (2003).
28. Lopetuso, L. R., Scaldaferri, F., Petito, V. & Gasbarrini, A. Commensal *Clostridia*: leading players in the maintenance of gut homeostasis. *Gut Pathog.* **5**, 23 (2013).
29. Coats, S. R., To, T. T., Jain, S., Braham, P. H. & Darveau, R. P. *Porphyromonas gingivalis* resistance to polymyxin B is determined by the lipid A 4′-phosphatase, PGN_0524. *Int. J. Oral Sci.* **1**, 126–135 (2009).
30. Ingram, B. O., Masoudi, A. & Raetz, C. R. H. *Escherichia coli* mutants that synthesize dephosphorylated lipid A molecules. *Biochemistry* **49**, 8325–8337 (2010).
31. Porse, A., Schou, T. S., Munck, C., Ellabaan, M. M. H. & Sommer, M. O. A. Biochemical mechanisms determine the functional compatibility of heterologous genes. *Nat. Commun.* **9**, 522 (2018).
32. Kong, Q. et al. Phosphate groups of lipid A are essential for *Salmonella enterica* serovar Typhimurium virulence and affect innate and adaptive immunity. *Infect. Immun.* **80**, 3215–3224 (2012).
33. Wang, R. et al. The global distribution and spread of the mobilized colistin resistance gene *mcr-1*. *Nat. Commun.* **9**, 1179 (2018).
34. Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: Reconstruction, analysis and visualization of phylogenomic data. *Mol. Biol. Evol.* **33**, 1635–1638 (2016).
35. McArthur, A. G. et al. The comprehensive antibiotic resistance database. *Antimicrob. Agents Chemother.* **57**, 3348–3357 (2013).
36. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
37. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
38. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).
39. Lagesen, K. et al. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**, 3100–3108 (2007).
40. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
41. O'Leary, N. A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).
42. Orlek, A. et al. A curated dataset of complete Enterobacteriaceae plasmids compiled from the NCBI nucleotide database. *Data Brief* **12**, 423–426 (2017).
43. Smillie, C. S. et al. Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* **480**, 241–244 (2011).
44. Jiang, X. et al. Comprehensive analysis of mobile genetic elements in the gut microbiome reveals a phylum-level niche-adaptive gene pool. Preprint at https://www.biorxiv.org/content/early/2017/11/06/214213 (2017).
45. Szybalski, W. Genetic studies on microbial cross resistance to toxic agents. IV. Cross resistance of *Bacillus megaterium* to forty-four antimicrobial drugs. *Appl. Microbiol.* **2**, 57–63 (1954).
46. Wiegand, I., Hilpert, K. & Hancock, R. E. W. Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nat. Protoc.* **3**, 163–175 (2008).
47. Rhoads, A. & Au, K. F. PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics* **13**, 278–289 (2015).
48. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
49. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
50. Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* **79**, 5112–5120 (2013).
51. Schloss, P. D. et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).
52. Oksanen, J. et al. vegan: community ecology package (2017); https://cran.r-project.org/web/packages/vegan/index.html
53. Chakravorty, S., Helb, D., Burday, M., Connell, N. & Alland, D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J. Microbiol. Methods* **69**, 330–339 (2007).

54. Fisher, R. A., Corbet, A. S. & Williams, C. B. The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* **12**, 42 (1943).
55. Simpson, E. H. Measurement of diversity. *Nature* **163**, 688 (1949).
56. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217 (2013).
57. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA–Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297 (2012).
58. Jonsson, V., Österlund, T., Nerman, O. & Kristiansson, E. Statistical evaluation of methods for identification of differentially abundant genes in comparative metagenomics. *BMC Genomics* **17**, 78 (2016).
59. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA–seq data. *Genome Biol.* **11**, R25 (2010).
60. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Source J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).
61. Sambrook, J. & Russell, D. W. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001).
62. Wang, X., McGrath, S. C., Cotter, R. J. & Raetz, C. R. H. Expression cloning and periplasmic orientation of the *Francisella novicida* lipid A 4′-phosphatase LpxF. *J. Biol. Chem.* **281**, 9321–9330 (2006).
63. Rossetti, F. F. et al. Interaction of poly(L-lysine)-g-poly(ethylene glycol) with supported phospholipid bilayers. *Biophys. J.* **87**, 1711–1721 (2004).
64. Bacic, M. K. & Smith, C. J. Laboratory maintenance and cultication of Bacteroides species. *Curr. Protoc. Microbiol.* **13**, 13C.1 (2008).
65. Lozupone, C. A., Hamady, M., Kelley, S. T. & Knight, R. Quantitative and qualitative diversity measures lead to different insights into factors that structure microbial communities. *Appl. Environ. Microbiol.* **73**, 1576–1585 (2007).

## Author contributions

B.K. and C.P. conceived of the project. B.K., O.M., E.A., A.N., B.P. and C.P. planned the experiments and data analyses. O.M., M.S., P.K.J. and R.T. performed the experiments. I.N. performed Illumina sequencing. A.B. and T.M. were responsible for faecal sample collection. B.K., O.M., E.A., A.G., F.P., G.F., I.L., B.B., B.M.V and M.B. analysed the experimental data and carried out bioinformatic analyses. B.K., O.M., B.P. and C.P. wrote the manuscript.

## Competing interests

B.M.V., B.B. and I.N. had consulting positions at SeqOmics Biotechnology Ltd at the time the study was conceived. SeqOmics Biotechnology Ltd was not directly involved in the design and execution of the experiments or in the writing of the manuscript. This does not alter the authors' adherence to all the Nature policies on sharing data and materials. The other authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41564-018-0313-5.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Correspondence and requests for materials** should be addressed to B.K. or B.P. or C.P.

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

# nature research

Corresponding author(s):   Csaba Pal, Balazs Papp and Balint Kintses

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see <u>Authors & Referees</u> and the <u>Editorial Policy Checklist</u>.

## Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

| n/a | Confirmed | |
|-----|-----------|---|
| ☐ | ☒ | The <u>exact sample size</u> (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☐ | ☒ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistics including <u>central tendency</u> (e.g. means) or other basic estimates (e.g. regression coefficient) AND <u>variation</u> (e.g. standard deviation) or associated <u>estimates of uncertainty</u> (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☐ | ☒ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☐ | ☒ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☐ | ☒ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |
| ☐ | ☒ | Clearly defined error bars<br>*State explicitly what error bars represent (e.g. SD, SE, CI)* |

*Our web collection on <u>statistics for biologists</u> may be useful.*

## Software and code

Policy information about <u>availability of computer code</u>

| Data collection | We did not use any code or software to collect data for this study |
|-----------------|---------------------------------------------------------------------|
| Data analysis | We did not use any custom code or mathematical algorithm that was not published before. Blast was carried out with blastx or blasp option of the Diamond software (version 0.9.10) or blast source code was downloaded from NCBI (downloaded on the 01.10.2018), 16S rRNA sequences were identified with RNAmmer software (version 1.2), network figure was drawn with Cytoscape, pacBio sequencing data was analysed with RS_subreads.1 pipeline on SMRT Portal (version 2.3.0), ORFs were predicted and annotated using the Prokka suite (version 1.11), 16S rRNA reads were demultiplexed and processed with the Mothur software (version 1.36.1), rarefaction curves of 16S rRNA samples were calculated using the vegan (version 2.4-3) R package, Unweighted Unifrac distances were computed by the Phyloseq (version 1.22.3 R package), differential abundance analysis was done with edgeR (version 3.16.5 R package). All references for these codes are available in the manuscript. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research <u>guidelines for submitting code & software</u> for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

> GenBank accession numbers for the PacBio sequencing data are MH883365-MH883616. All data generated or analysed during this study are included in this published article and its supplementary information files. For each figure the availability of the analysed data is provided in the figure legend.

# Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences　　☐ Behavioural & social sciences　　☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | Throughout the work we followed the standards of previously published protocols. In functional metagenomics the number of tested treatments (11 antibiotics and 12 antimicrobial peptides), library sizes (2-4 million clones per environments) and number of DNA sources (all together 7 individuals) met with the standards of previously published studies in the field of antibiotic resistance research (DOI:10.1126/science.1176950, DOI:10.1038/nature13377). The studied 11 antibiotics served as positive controls for the screens with antimicrobial peptides. The large number of resistant DNA fragments obtained with the antibiotic confirmed that our sample size was large enough for the presented study. For the micriobiome culturing we followed the same logic. We choose a sample size (7 individuals) that was similar to a previously published work (DOI:10.1038/ncomms5714) and than we carried out a thorough quality check with 16S rRNA analysis to test if bacterial diversity from the 7 individuals was in agreement with previous efforts to culture the microbiome. For the comparative genomic analysis the number of genomes was determined in a way to well represent the gut microbiome. |
| Data exclusions | No data were excluded. |
| Replication | All experimental screens were carried out at least in two replicates. Replicate numbers were increased to 3 in the case of microbiome culturing. |
| Randomization | Samples were not allocated to experimental groups. For one type of experiments all samples were handled by one person in charge. |
| Blinding | Investigators were not blinded during experiments or data process. |

# Reporting for specific materials, systems and methods

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☐ | ☒ Unique biological materials |
| ☒ | ☐ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology |
| ☒ | ☐ Animals and other organisms |
| ☐ | ☒ Human research participants |

## Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

## Unique biological materials

Policy information about availability of materials

| | |
|---|---|
| Obtaining unique materials | All materials were from standard commercial sources with the exception of the fecal samples that were unique materials. Antimicrobial peptides were custom synthesized at ProteoGenix (France). |

## Human research participants

| Population characteristics | Human stool samples were obtained from seven healthy unrelated individuals. The 3 male and 4 female participants were in the age range of 20-38 years. |
| --- | --- |
| Recruitment | Volunteer participants were selected based on the strict criteria that (1) they did not take any antibiotics at least for one year prior sample donation, (2) they are in a good health condition. These requirements are standard in the field and secure a bias free comparison of the antibiotic and antimicrobial peptide resistomes in the healthy human gut microbiome. |