Supplementary Information

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

Bálint Kintses[1*§], Orsolya Méhi[1§], Eszter Ari[1,2§], 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*] & Csaba Pál[1*]

[1]Synthetic and Systems Biology Unit, Institute of Biochemistry, Biological Research Centre of the Hungarian Academy of Sciences, 6726 Szeged, Hungary.

[2]Department of Genetics, Eötvös Loránd University, 1117 Budapest, Hungary.

[3]Doctoral School in Biology, Faculty of Science and Informatics, University of Szeged, Szeged, Hungary

[4]SeqOmics Biotechnology Ltd., 6782 Mórahalom, Hungary.

[5]Sequencing Platform, Institute of Biochemistry, Biological Research Centre of the Hungarian Academy of Sciences, 6726 Szeged, Hungary.

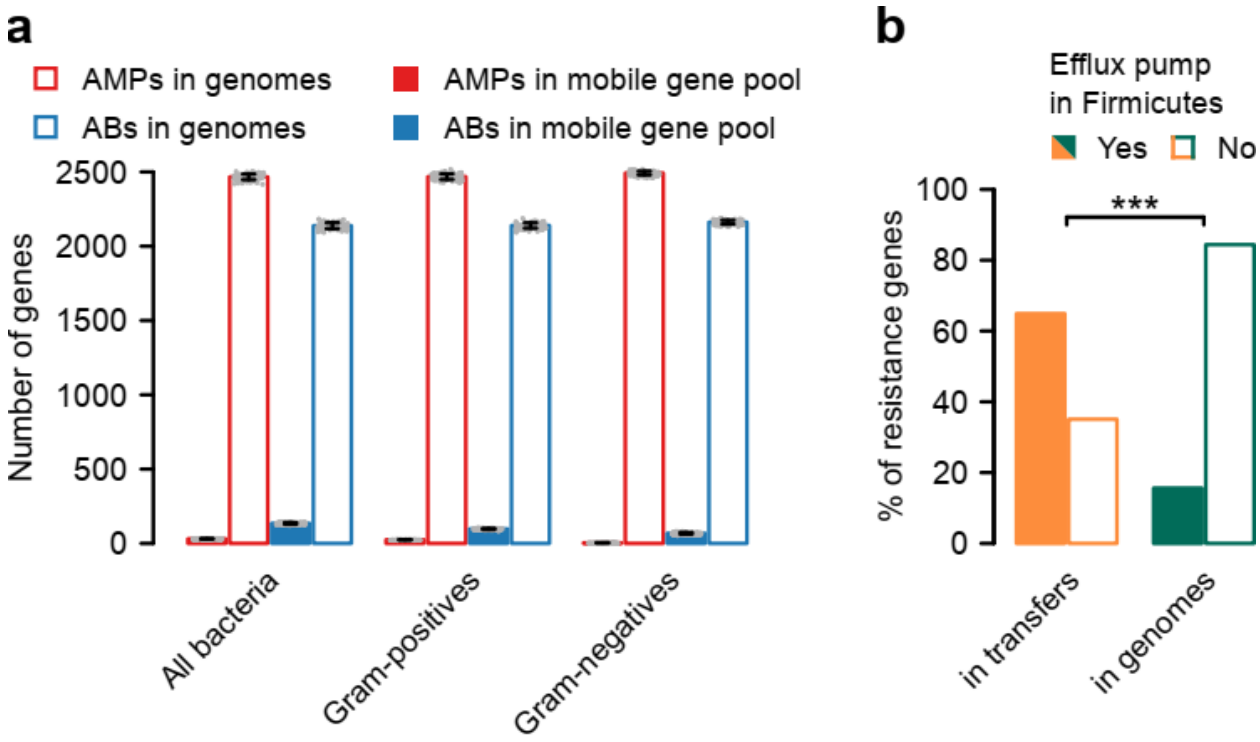
[6]Hereditary Endocrine Tumors Research Group, Hungarian Academy of Sciences and Semmelweis University, 1088 Budapest, Hungary.

[7]1st Department of Internal Medicine, Albert Szent-Györgyi Health Centre, University of Szeged, 6720 Szeged, Hungary

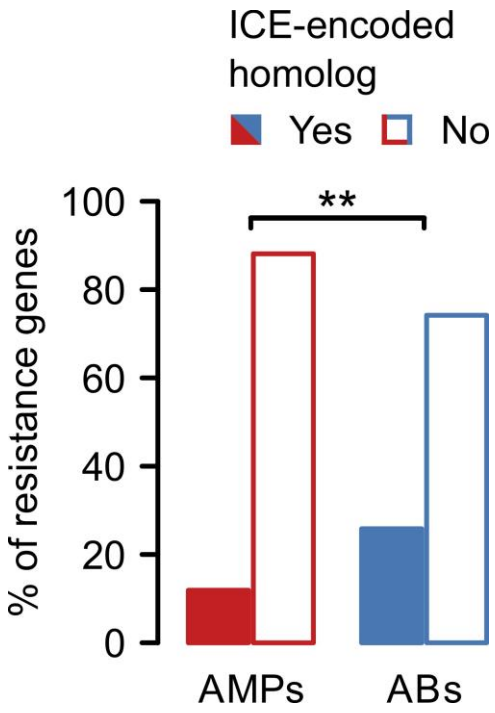
*Correspondence to cpal@brc.hu, pappb@brc.hu or kintses.balint@brc.mta.hu

[§]These authors contributed equally to this work.

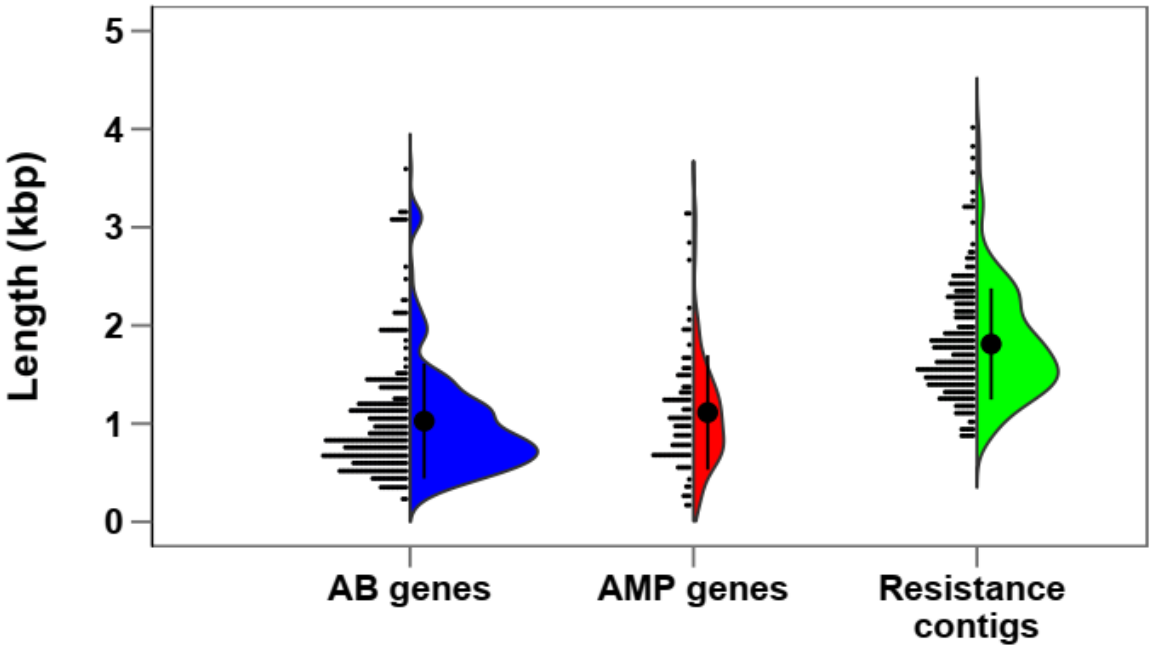
Supplementary Figures



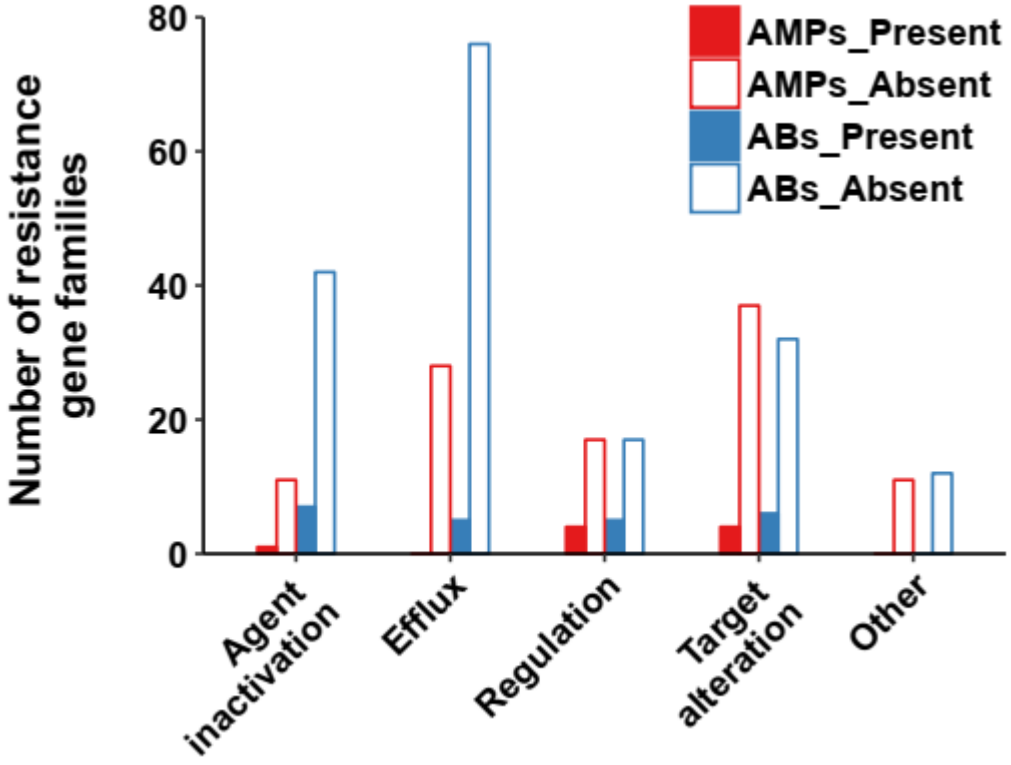
Supplementary Figure 1. The representation of the AMP- and antibiotic-resistance genes in the gut microbiome and in the mobile gene pool. **a)** The number of AMP- (red bars) and antibiotic-resistance genes (ABs, blue bars) that are detected in the genome sequences of the gut microbiota (empty bars) and in the mobile gene pool (filled bars). AMP- and antibiotic-resistance genes are similarly represented even if Gram-positive and Gram-negative species are considered separately. Resistance genes were identified using *blast* sequence similarity searches in the genome sequences of the gut microbiota (see Methods). In each case n=100 both for AMPs and ABs. Centre and error bars represent mean and SD calculated by randomly sampling 100 times from each of the 225 operational taxonomic units (OTUs), respectively (see Methods). OTUs were generated by collapsing genomes with lower than 2% 16S rRNA gene dissimilarities. The annotated resistance genes are provided in Supplementary Table 2. **b)** Efflux pumps in Firmicutes were overrepresented in the transfer events in the mobile gene pool (37 out of 57, n=57, orange in the figure, for data see Supplementary Table 3), compared to their relative frequencies eamong the annotated AMP resistance genes in the genome sequences of the gut microbiota (975 out of 6247, n=6247, indicated with green in the figure, for data see Supplementary Table 2). Asterisks (***) indicate significant difference from two-sided Fisher's exact test, p=10^{-15} (for more details see Methods).



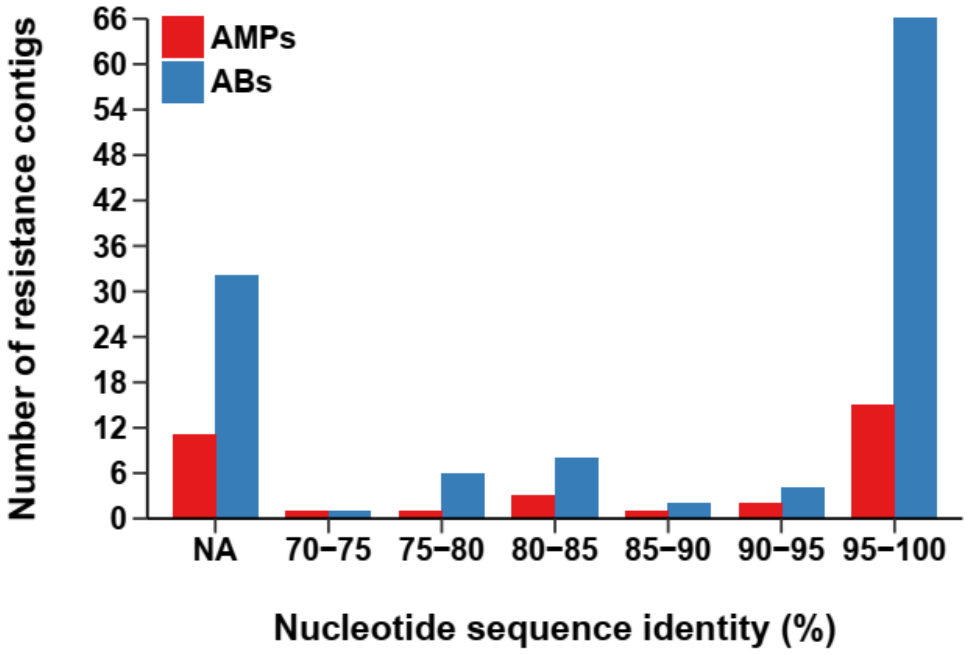
Supplementary Figure 2. Disproportionally fewer AMP resistance genes had homologs in integrative conjugative elements (ICEs) than antibiotic resistance genes. A significantly smaller fraction of the literature-curated AMP resistance genes (12 out of 101) has a close homolog in ICE sequences as a cargo protein than that of antibiotic resistance genes (79 out of 306). Sample sizes were 101 and 306 for AMPs and ABs, respectively. Asterisk indicates significant difference, p=0.036, two-sided Fisher's exact test. Since the literature-curated lists of resistance genes contain proteins from various sources in addition to the human gut microbiota, only those homologs were considered in the set of non-ICE-associated resistance genes that were detected in the genome sequences of the Human Microbiome Project database (Supplementary Table 2, for more details see Methods). Data is provided in Supplementary Table 4.



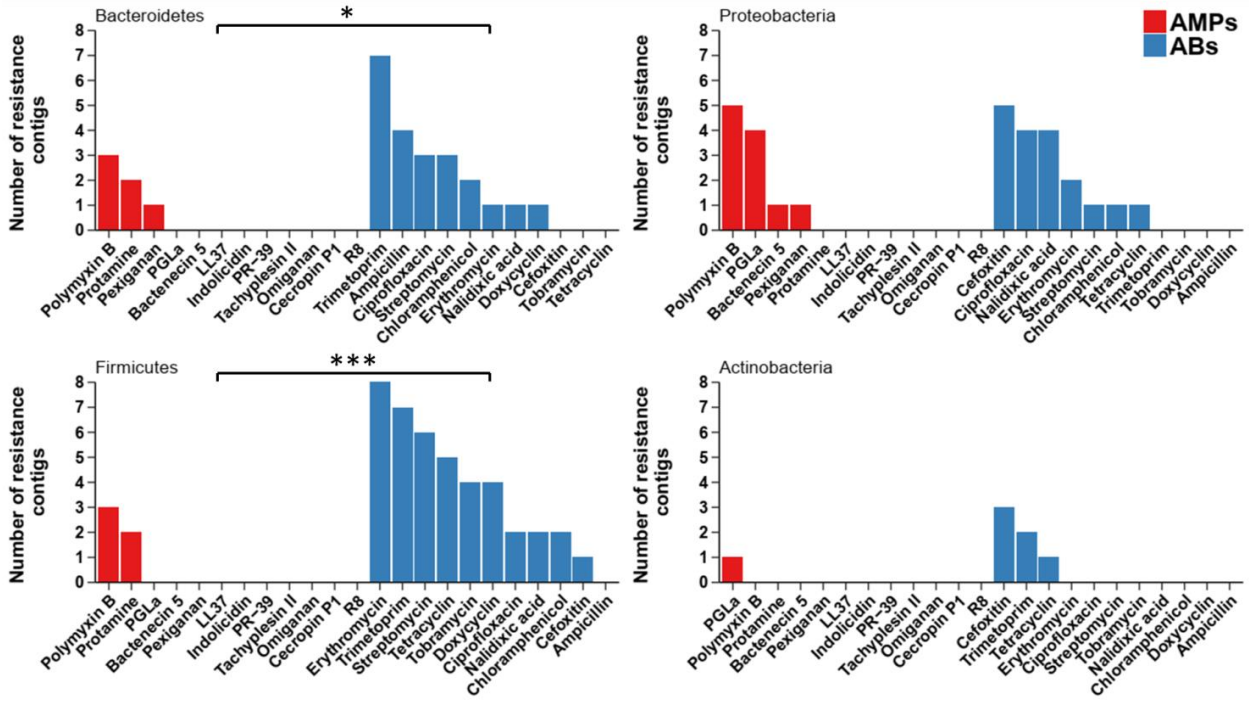
Supplementary Figure 3. The length distributions of known antibiotic- (AB) and AMP-resistance genes are well within the fragment size range of the metagenomic library. Blue and red plots show the length (kilobase pair (kbp)) distribution of known antibiotic- and AMP-resistance genes, respectively, from a representative set of gut microbial genomes (Supplementary Table 2). Green plot shows the length (kbp) distribution of all antibiotic- and AMP-resistance DNA fragments (resistance contigs) identified in our functional metagenomic selections (Supplementary Tables 6 and 10). Each plot shows both the raw data points (left side) and sthe distribution of the data points (violin plot on the right side). Dot and error bars represent the mean value and SD, respectively. Sample sizes were 300, 95 and 251 for AB genes, AMP genes and Resistance contigs, respectively.



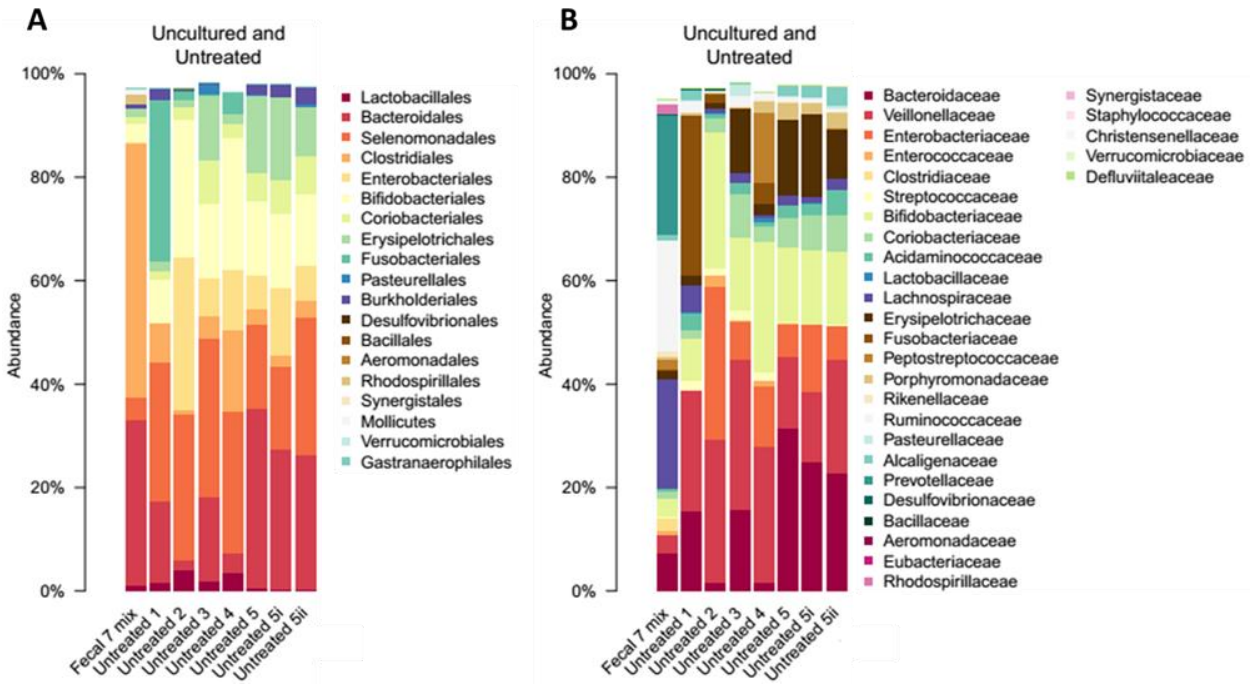
Supplementary Figure 4. Presence (filled bars) and absence (empty bars) of the known AMP- (red bars) and antibiotic-resistance gene families (ABs, blue bars) on the metagenomic contigs identified in our functional selections (Supplementary Table 6). Genes were assigned to gene families (orthogroups) which were classified into major functional categories (see Methods). A gene family was considered present if at least one resistance gene from its orthogroup had a significant sequence similarity hit on the DNA fragments (see Methods). We considered a resistance gene family absent if the orthogroup did not result in any hits on the metagenomic contigs from the functional selections. *n*=114 for AMPs, *n*=199 for ABs.



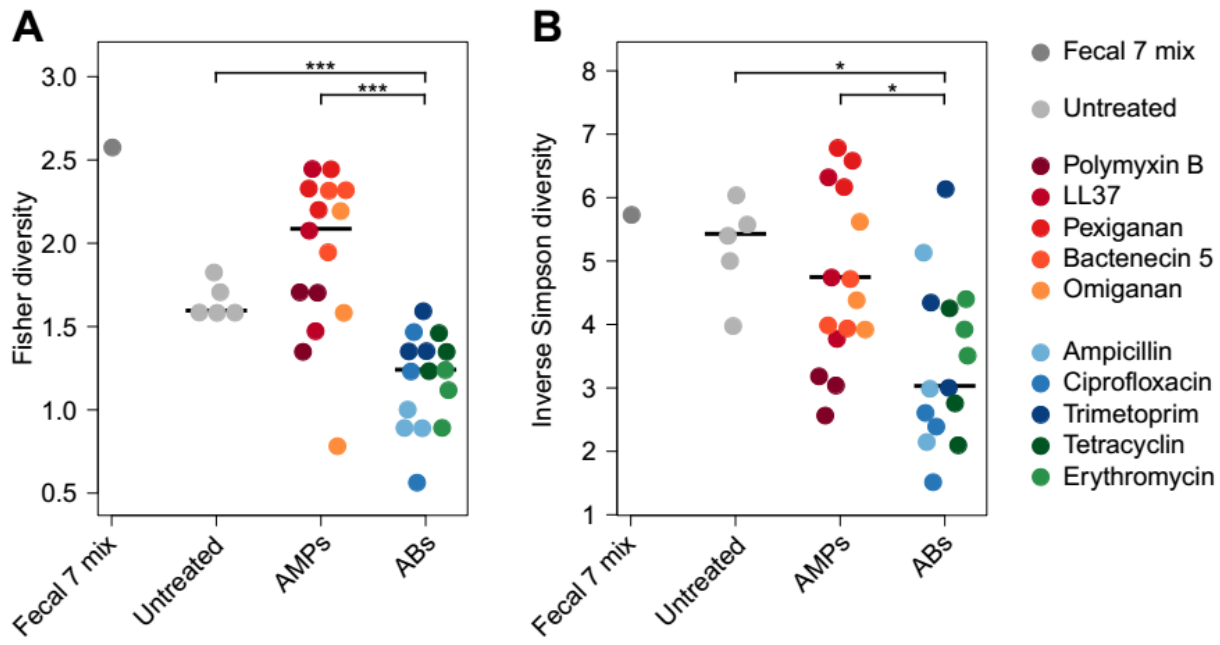
Supplementary Figure 5. Distribution of the nucleotide sequence identities between the AMP- (red bars) and antibiotic-resistance contigs (ABs, blue bars) originating from the metagenomic libraries of the uncultured microbiota (Supplementary Table 6) and the genome sequences from the Human Microbiome Project (HMP) database[24] (see Methods). 28 % of the contig sequences did not result in significant alignment (marked with NA) in the nucleotide sequence similarity search against the genome sequences from the HMP database. Significant alignment is defined by an *e*-value $<10^{-10}$ from NCBI *blastn* algorithm (see Methods). *n*=119 for ABs, *n*=34 for AMPs.



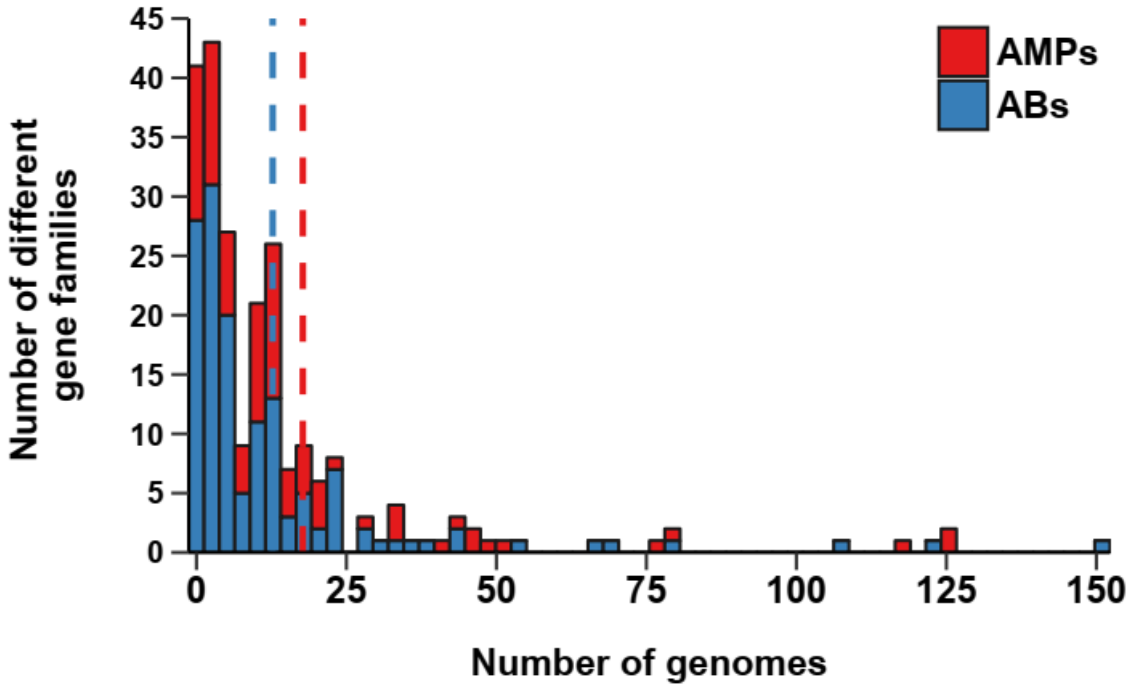
Supplementary Figure 6. Phylum-level distribution of resistance contigs originating from the functional selection of the uncultured microbiota with different AMPs (red bars, *n*=23) and antibiotics (ABs, blue bars, *n*=87) (Supplementary Table 6). Asterisks indicate significant difference between AMPs and ABs (p=0.02 and $4.2*10^{-5}$ from two-sided negative binomial regression for Bacteroidetes and Firmicutes, respectively).



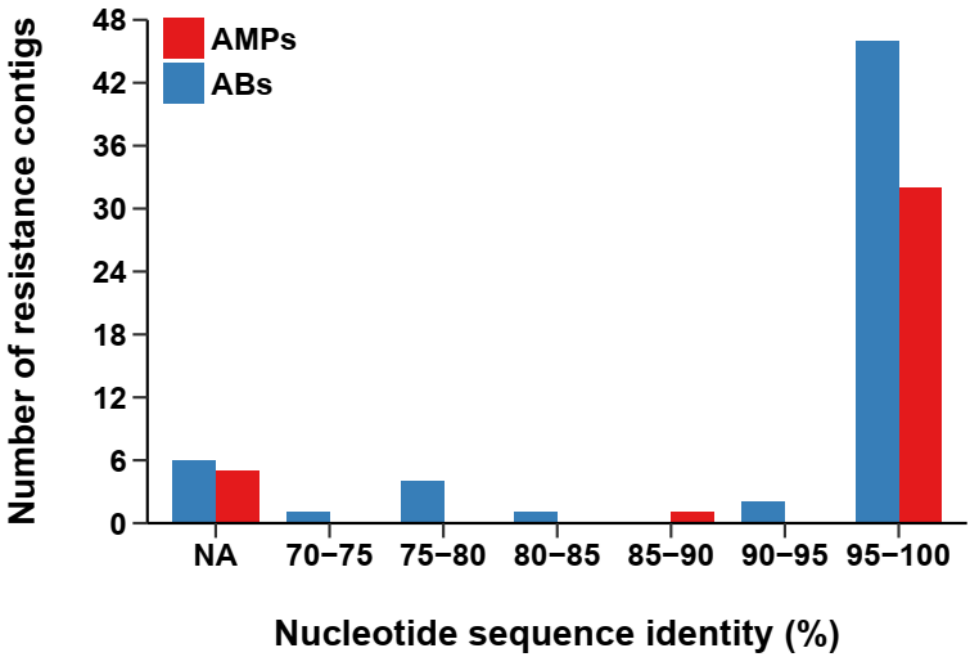
Supplementary Figure 7. Per cent abundances of bacterial orders (A) and families (B) in the uncultured (Fecal 7 mix) and anaerobically cultured (Untreated 1-5, 5i, 5ii) gut microbial samples from seven unrelated healthy individuals. At order level, the cultivable proportion was 64-86 %, while at the family level it was 65-74 %. These results are consistent with the cultivation efficiency reported in a previous study[25]. “Untreated 1-5” samples are biological replicates started from different aliquots of the same frozen samples in independent cultivation experiments. “Untreated 5i” and “Untreated 5ii” are technical replicates of the “Untreated 5” sample started from the same sample and grown at the same time in the same experiment. Data are presented in Supplementary Table 7.



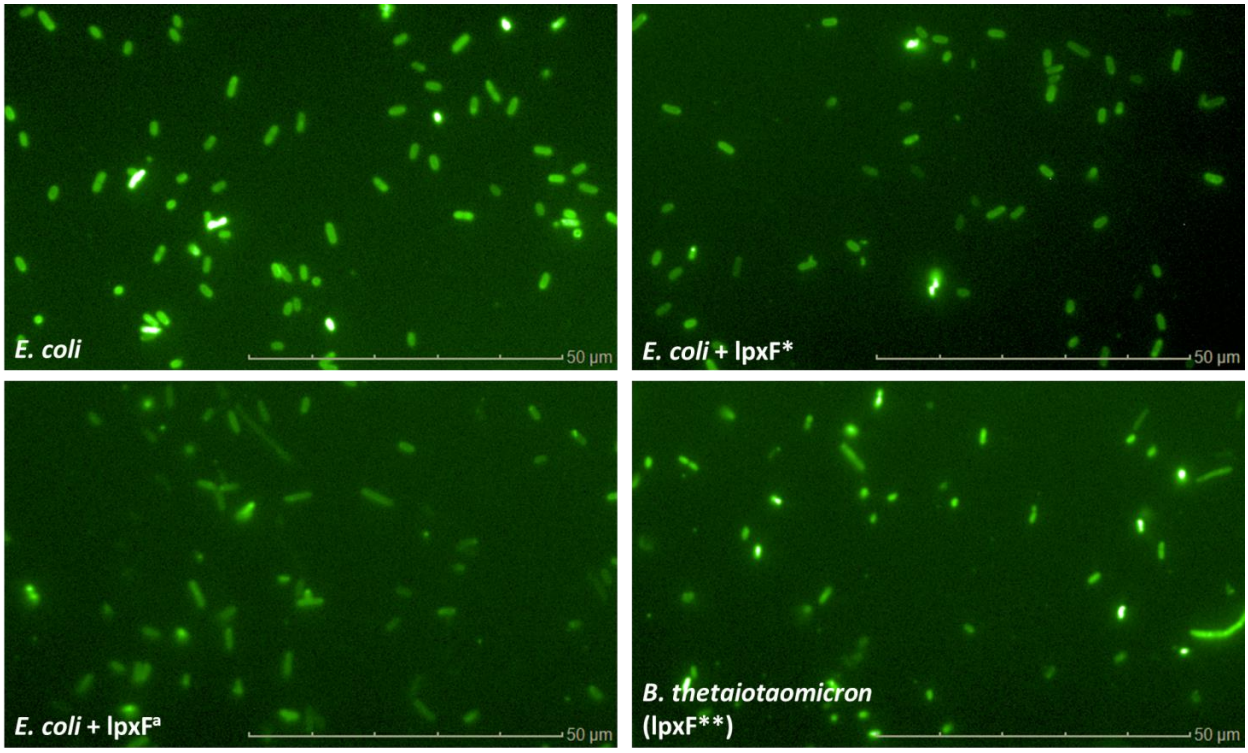
Supplementary Figure 8. Diversities of the cultured microbiota and the original faecal sample (Fecal 7 mix). Fisher (A)[64] and Inverse Simpson (B)[65] alpha diversity indices were calculated at the bacterial family level. Alpha diversity reflects the number of different taxa and distribution of abundances. Asterisks indicate significant difference from two-sided Mann-Whitney U test. p=0.0005 and 0.00001 for Untreated vs ABs and AMPs vs ABs, respectively in figure A, p=0.0085 and 0.0053 for Untreated vs ABs and AMPs vs ABs, respectively in figure B. Central horizontal bars represent median values. Sample sizes were 1, 5, 15 and 15 for Fecal 7 mix, Untreated, AMPs and ABs, respectively both for figure A and B.



Supplementary Figure 9. The diversity of previously described AMP- (red bars) and antibiotic-resistance gene families (ABs, blue bars) in a representative set of gut microbial genomes[15]. The red and blue dashed lines represent the mean values for AMPs and ABs, respectively. For details see Methods ("*Comparing the prevalence of AMP- and antibiotic-resistance genes in gut microbial genomes*" section). (p=0.037 from two-sided negative binomial regression, n=87 for AMPs and n=140 for ABs).

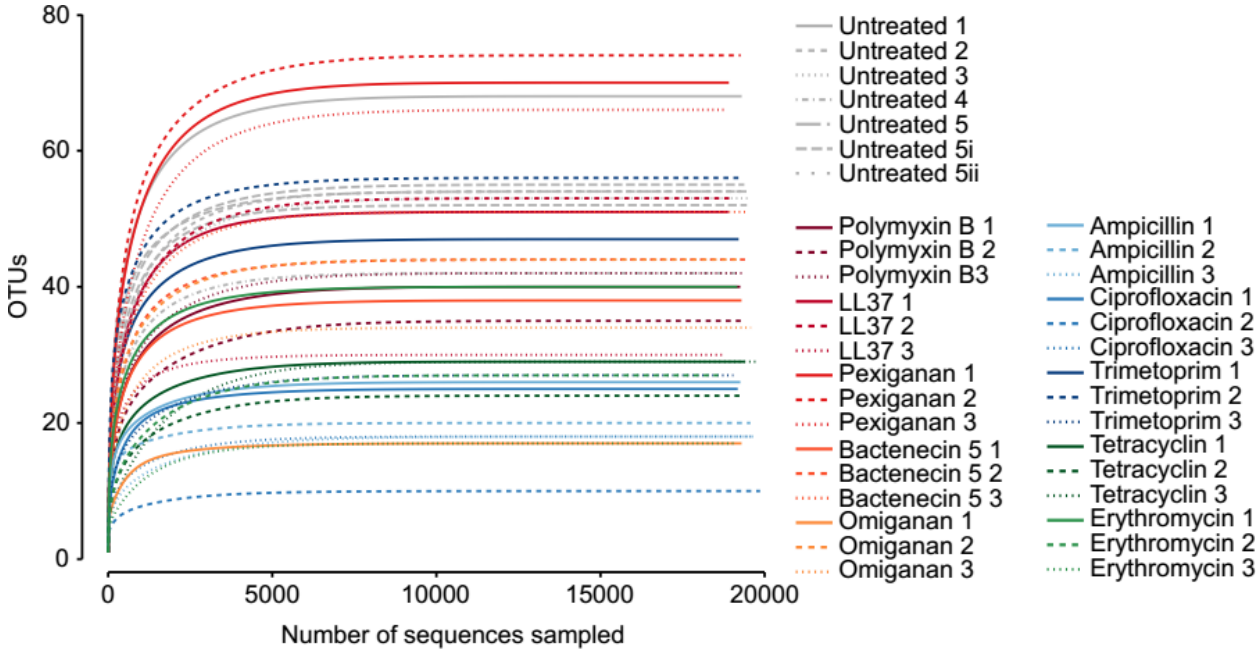


Supplementary Figure 10. Distributions of the nucleotide sequence identities between the AMP- (red bars) and antibiotic-resistance contigs (AB, blue bars) originating from the cultured microbiota and the genome sequences from the Human Microbiome Project (HMP) database (see Methods). *n*=60 for ABs, *n*=38 for AMPs (Supplementary Table 10). 11% of the contig sequences did not result in significant alignment (marked with NA) in the nucleotide sequence similarity search against the genome sequences from the HMP database. Significant alignment is defined by an *e*-value $<10^{-10}$ from NCBI *blastn* algorithm (see Methods).

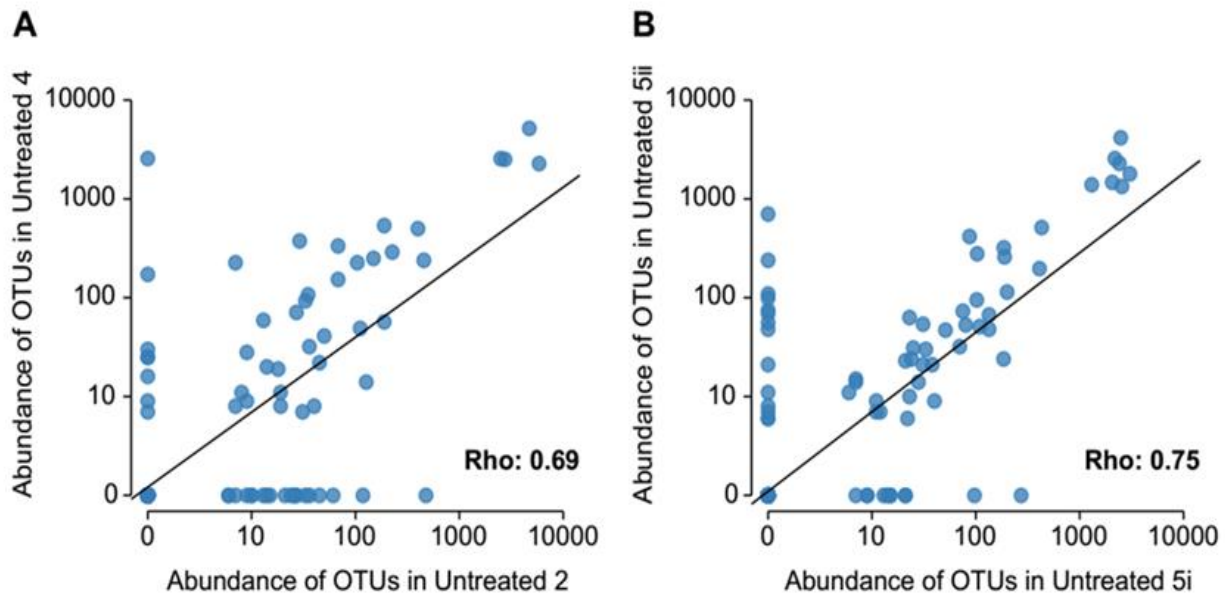


Supplementary Figure 11. Microscopic images of cells incubated with FITC-labeled poly-L-lysine (PLL) polycationic molecules. The brighter the cells are the more PLL is bound, indicating a more negative outer membrane surface charge. For details see Methods, "Surface charge measurement" section. The quantitative analysis of the fluorescent signals is presented in Figure 4b.

Abbreviations: *E. coli* = BW25113 Δ*lpxM; E. coli* + *lpxF*^{a} = *E. coli* expressing an *lpxF* ortholog from *Parabacteroides merdae*. This gene was identified in our metagenomic screen (denoted as *lpxF*^{a} in Table 1). *E. coli* + *lpxF* = E. coli* expressing the *LpxF* of *Francisella novicida. B. thetaiotaomicron* (*lpxF***) = *Bacteroides thetaiotaomicron* strain, which intrinsically expresses the *lpxF BT1854* gene (*lpxF***). In BW25113 Δ*lpxM* strain the *lpxM* gene is deleted and therefore this strain has pentaacylated lipid-A molecules instead of the hexaacylated ones[30]. We used this strain to allow the phenotypic comparison of the LpxF^{a} identified in our screen to the LpxF from *Francisella novicida*, which can be expressed only in *ΔlpxM E. coli* that has pentaacylated LPS molecules (*n*=4 biological replicates).



Supplementary Figure 12. Rarefaction Operational taxonomic units (OTU or 'species') richness curves. Curves are plots of the number of OTUs as a function of the number of sequences, for different samples. At a sample size of 10,000 sequences, all curves reach their plateau, indicating that sufficient sequences were used to estimate the number of OTUs in different samples. Sequences were merged at the level of 97% sequence identity to generate OTUs (see Methods).



Supplementary Figure 13. OTUs abundances correlate across biological (A) and technical (B) replicates of cultivated, untreated samples, showing good reproducibility of the replicates. Spearmann's rho is 0.69 (A) and 0.75 (B), respectively. Points represent individual OTUs that were generated by merging 16S rRNA sequences at the level of 97% sequence identity (see Methods). Untreated 2 and 4 samples are biological replicates, while Untreated 5i and 5ii are technical replicates. Biological replicates originate from independent cultivation experiments starting from different aliquots of the same frozen samples. Technical replicates originate from parallel cultivations starting from the same sample and grown at the same time. Sample size (number of OTUs) is 362 for both figure A and B.

Supplementary Tables

Supplementary Table 5. List and characteristics of antimicrobials used in this study. The origin of the antimicrobial peptides is indicated in brackets and the peptides marked with asterisks are involved in clinical trials.

Antimicrobials	Mode of action	Used in
ANTIMICROBIAL PEPTIDES (AMPs)		
Cecropin P1 (parasitic nematode)	pore-forming activity	Functional metagenomics, culturing the gut microbiota
Bactenecin 5 (cow)	intracellular targets	Functional metagenomics, culturing the gut microbiota
Indolicidin (cow)	pore-forming activity, inhibition of DNA synthesis	Functional metagenomics
LL37 (human)	pore-forming activity, membrane depolarization	Functional metagenomics, culturing the gut microbiota
Omiganan* (synthetic, indolicidin analogue)	pore-forming activity, inhibition of DNA synthesis	Functional metagenomics, culturing the gut microbiota
Pexiganan* (synthetic, magainin analogue)	pore-forming activity	Functional metagenomics, culturing the gut microbiota
PGLa (frog)	pore-forming activity	Functional metagenomics
Polymyxin B (bacterial)	pore-forming activity, membrane depolarization	Functional metagenomics, culturing the gut microbiota
PR-39 (pig)	inhibition of DNA- and protein synthesis	Functional metagenomics
Protamine (salmon)	intracellular targets	Functional metagenomics
R8 (synthetic)	pore forming activity	Functional metagenomics
Tachyplesin II (horseshoe crab)	pore forming activity, membrane depolarization	Functional metagenomics

ANTIBIOTICS (ABs)

Ampicillin	inhibition of cell wall synthesis	Functional metagenomics, culturing the gut microbiota
Cefoxitin	inhibition of cell wall synthesis	Functional metagenomics
Ciprofloxacin	inhibition of DNA synthesis	Functional metagenomics, culturing the gut microbiota
Cloramphenicol	inhibition of protein synthesis	Functional metagenomics
Doxycycline	inhibition of protein synthesis	Functional metagenomics
Erythromycin	inhibition of protein synthesis	Functional metagenomics, culturing the gut microbiota
Nalidixic acid	inhibition of DNA synthesis	Functional metagenomics
Streptomycin	inhibition of protein synthesis	Functional metagenomics
Tetracycline	inhibition of protein synthesis	Functional metagenomics, culturing the gut microbiota
Tobramicin	inhibition of protein synthesis	Functional metagenomics
Trimethoprim	inhibition of folate synthesis	Functional metagenomics, culturing the gut microbiota

Supplementary Table 8. Selection conditions used for the anaerobic cultivation experiments. Those AMP and antibiotic treatment concentrations were chosen where the average proportion of the resistant colonies originating from three selection concentrations per AMP and antibiotic ranged from 0.01 to 0.1% of the total cultivable colony number in the absence of any antimicrobial. Trimethoprim represented an exception, where the resistant fraction was slightly higher, but the concentration could not be raised further because of solubility issues.

Antimicrobial peptide (AMP) / antibiotic (AB)	Selection concentration (µg/ml)	Treatment type	Relative resistance level of the gut microbiota*	Resistant colonies (percentage of the cultivable cell population)
Bactenecin 5	666.7	AMP	6.7	0.02
LL37	1400	AMP	10	0.02
Omiganan	250	AMP	8.3	0.04
Pexiganan	166.7	AMP	16.7	0.01
Polymyxin B	280	AMP	1400	0.1
Ampicillin	700	AB	46.7	0.1
Ciprofloxacin	75	AB	833.3	0.03
Erythromycin	1600	AB	40	0.04
Tetracycline	48	AB	26.7	0.1
Trimetoprim	1200	AB	1500	0.49

Supplementary Table 9. Characteristics of metagenomic libraries constructed from AMP- and antibiotic-resistant microbiota cultures with established microbial compositions.

Antimicrobial peptide (AMP) / antibiotic (AB)	Library name	Library size (Gbp)	Average insert size (bp)	Treatment type
Bactenecin 5	BAC5_1	2.2	2700	AMP
Bactenecin 5	BAC5_2	1.8	2600	AMP
LL37	LL37_1	1.736	2800	AMP
LL37	LL37_2	4.833	4833	AMP
Omiganan	OMG_1	0.8	1000	AMP
Omiganan	OMG_2	0.96	1600	AMP
Pexiganan	PEX_1	1.658	1658	AMP
Pexiganan	PEX_2	7.5506	1900	AMP
Polymyxin B	PMB_1	2.4464	3058	AMP
Polymyxin B	PMB_2	1.6	1600	AMP
Ampicillin	AMP_1	1.02	1700	AB
Ampicillin	AMP_2	0.884	1300	AB
Ciprofloxacin	CPR_1	0.594	1100	AB
Ciprofloxacin	CPR_2	1.3325	2050	AB
Erythromycin	ERY_1	0.6	1600	AB
Erythromycin	ERY_2	2.1692	3190	AB
Tetracyclin	TET_1	0.174563	3675	AB
Tetracyclin	TET_2	3.2	4000	AB
Trimetoprim	TRM_1	1.5	2300	AB
Trimetoprim	TRM_2	2.56074	3283	AB

Supplementary Table 11. Phylum-level distributions of the AMP- and antibiotic-resistant microbiota and the transferring resistance contigs originating from them. Count data represent either colony numbers from the culturing experiments estimated by using the 16S rRNA abundance data (Supplementary Table 7) or the number of resistance contigs detected in the functional metagenomics screens. The data was used for logistic regression analyses to determine if the phylum-level representation of resistance contigs are proportional to that of the cultured microbiota (see Methods).

Phylum	Source	Treatment	Count
Proteobacteria	Transferred DNA contig	AMP	18
Proteobacteria	Transferred DNA contig	AB	3
Proteobacteria	Gut bacteria colony number	AMP	5590
Proteobacteria	Gut bacteria colony number	AB	1838
Firmicutes	Transferred DNA contig	AMP	5
Firmicutes	Transferred DNA contig	AB	17
Firmicutes	Gut bacteria colony number	AMP	14734
Firmicutes	Gut bacteria colony number	AB	11402
Bacteroidetes	Transferred DNA contig	AMP	8
Bacteroidetes	Transferred DNA contig	AB	30
Bacteroidetes	Gut bacteria colony number	AMP	1050
Bacteroidetes	Gut bacteria colony number	AB	8077
Actinobacteria	Transferred DNA contig	AMP	0
Actinobacteria	Transferred DNA contig	AB	4
Actinobacteria	Gut bacteria colony number	AMP	647
Actinobacteria	Gut bacteria colony number	AB	2868
Fusobacteria	Transferred DNA contig	AMP	0
Fusobacteria	Transferred DNA contig	AB	0
Fusobacteria	Gut bacteria colony number	AMP	630
Fusobacteria	Gut bacteria colony number	AB	151

Captions for Supplementary Tables

Supplementary Table 1. A comprehensive catalogue of previously reported AMP resistance genes compiled based on literature mining and of antibiotic resistance genes from the CARD database. The table is provided as a separate Excel file.

Supplementary Table 2. Identification of the AMP- and antibiotic-resistance genes in the bacterial genome sequences from which the mobile gene pool was derived. The table is provided as a separate Excel file.

Supplementary Table 3. Characteristics of the transferred AMP- and antibiotic-resistance genes in the mobile gene pool. The table is provided as a separate Excel file.

Supplementary Table 4. Identification of AMP- and antibiotic-resistance genes associated with naturally occurring plasmids and Integrative Conjugative Elements (ICEs) in the human microbiota. The table is provided as a separate Excel file.

Supplementary Table 6. List of resistance contigs identified from the functional metagenomic selections of the uncultured microbiota with 12 AMPs and 11 small-molecule antibiotics (Supplementary Table 5). The table is provided as a separate Excel file.

Supplementary Table 7. Bacterial abundances in the untreated and AMP/antibiotic resistant microbiota at family and order level, respectively. Abundances at order and family levels are presented on separate Excel sheets. The table is provided as a separate Excel file.

Supplementary Table 10. List of resistance contigs identified from the functional selection of the cultured microbiota with 5 AMPs and 5 small-molecule antibiotics (Supplementary Table 5). The table is provided as a separate Excel file.

Supplementary Table 12. Resistance gains in *E. coli* and *S. enterica* provided by a representative set of plasmids carrying AMP- and antibiotic-resistance contigs that were isolated in our metagenomic screens. The table is provided as a separate Excel file.

Supplementary Table 13. List of primers used in this study. The table is provided as a separate Excel file.