Antimicrobial peptides in the global microbiome: Biosynthetic genes and resistance determinants

Bingfeng Chen[1], Zhenyan Zhang[1], Qi Zhang[1], Nuohan Xu[1], Tao Lu[1], Tingzhang Wang[2], Wenjie Hong[2], Zhengwei Fu[3], Josep Penuelas[4,5], Michael Gillings[6], Haifeng Qian[1],*

1. College of Environment, Zhejiang University of Technology, Hangzhou 310032, P. R. of China.
2. Key laboratory of microbial technology and bioinformatics of Zhejiang Province, Hangzhou 310012, P. R. of China.
3. College of Biotechnology and Bioengineering, Zhejiang University of Technology, Hangzhou 310032, P. R. of China.
4. CSIC, Global Ecology Unit CREAF-CSIC-UAB, Bellaterra, Barcelona 08193, Catalonia, Spain.
5. CREAF, Campus Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Barcelona 08193, Catalonia, Spain.
6. ARC Centre of Excellence in Synthetic Biology, Faculty of Science and Engineering, Macquarie University, NSW 2109, Australia

*Correspondence to: Haifeng Qian (hfqian@zjut.edu.cn)

Abstract: Antimicrobial peptides are a promising new class of antimicrobials that could address the antibiotic resistance crisis, which poses a major threat to human health. These peptides are present in all kingdoms of life, but especially in microorganisms, where they have multiple origins in diverse taxa. To date, there has been no global study on the diversity of antimicrobial peptides, the hosts in which these occur, and the potential for resistance to these agents. Here we investigated the diversity and number of antimicrobial peptides in four main habitats (aquatic, terrestrial, human, and engineered), by analyzing 52,515 metagenome-assembled genomes. The richness of antimicrobial peptides was higher in the human gut microbiome than in other habitats, and most hosts of antimicrobial peptides were habitat-specific. The abundance of genes that confer resistance to antimicrobial peptides varied between habitats and was generally at a low level, except for the built environment and on human skin. The horizontal transfer of potential resistance genes among these habitats was constrained by ecological barriers. We systematically quantified the risk of each resistance determinant to human health and found almost one third pose a risk, especially those that confer resistance to polymyxin B, bacitracin, and defensin. Our results help to identify the biosynthetic potential of antimicrobial peptides in the global microbiome, further identifying peptides where there is a low risk of developing resistance.

.

Keywords: antibiotic; antimicrobial peptide; resistance; metagenome-assembled genomes; health risk; biosynthesis

The prevalence of antibiotic-resistant pathogens exacerbates difficult-to-treat infections in the clinic and poses a serious threat to human health. Infection-related deaths owing to drug-resistant pathogens are projected to account for the highest proportion of deaths worldwide by 2050. Many currently available antibiotics are ineffective due to the development of antibiotic resistance, leading to a crisis in our ability to control bacterial infections (*1*, *2*). Designing and producing novel effective antimicrobial agents is therefore urgently needed.

Antimicrobial peptides (AMPs) are attractive antibacterial drug candidates for overcoming the crisis of antibiotic resistance, due to their high diversity and broad activity (*3*). AMPs are relatively short, structurally diverse small proteins that usually have amphipathic and cationic host-defense properties against pathogen invasion. They are found in diverse taxa, from mammals to soil microorganisms and aquatic environments (*4*, *5*). Thousands of AMPs have been identified from plant, animal, and microbial sources (*6*), and dozens of AMPs are being evaluated against important pathogens in clinical trials (*7*–*12*). AMPs operate by multiple mechanisms, such as membrane disruption, intracellular osmosis, and immune regulation (*13*). For instance, positively charged AMPs interact with negatively charged substances such as the phospholipids and lipopolysaccharides in bacterial membranes to form lesions and cause cell lysis (*5*).

AMPs are present in all kingdoms of life. Nearly 40% of AMPs in clinical trials, however, are of human origin (*3*), raising legitimate concerns that the evolution of resistance against AMPs may lead to collateral resistance to endogenous human

immunity, which would increase the threat to human health. Therefore, it would be prudent to mine AMPs from other domains of life, and especially from microorganisms, where there is immense phylogenetic and functional diversity.

The very ubiquity of AMPs in the environment potentially provides strong selective pressure to drive resistance (*5*, *14*–*17*), even though the pharmacodynamics and killing mechanisms of AMPs are less conducive to the evolution of resistance than traditional antibiotics (*3*). However, similarly to antibiotic resistance, antimicrobial peptide resistance genes (AMPRGs) could be acquired by new hosts through horizontal gene transfer (*18*). This does greatly increase the chance of pathogens potentially acquiring these genes. Thus, before AMPs are used to replace antibiotics in clinical use (*19*), we should assess the potential threats, screen and develop low threat AMPs to avoid a similar fate to the antibiotic resistance crisis.

The identification of genes that encode AMPs has largely been facilitated by studying cultivable microorganisms. However, most microbes are not yet cultivated, and therefore the potential for AMP biosynthesis in most microorganisms remains largely unknown. Further, assigning these genes to their respective hosts remains challenging. In this work, to avoid cultivation bias, we characterized the potential for AMP biosynthesis in microbes at a global scale by analyzing 52,515 metagenome-assembled genomes (MAGs) constructed by Nayfach et al (*20*) in diverse habitats. By mining these metagenomic data, we discovered widespread AMPs in global microbiomes. Next, we annotated the MAGs with AMPRGs to systematically analyze their distribution patterns and identify their hosts at the global scale. This allowed us to

quantify the threat of AMPRGs to human health via three indicators (human accessibility, mobility, and human pathogenicity). These data will guide the development of low health threat AMPs to address the growing antibiotic crisis.

Results

Global distribution patterns of AMPs and their hosts

We conducted AMP mining in global microbiome data via the combination of Attention (ATT), long short-term memory (LSTM), and Bidirectional Encoder Representations from Transformers (BERT) models constructing by Ma et al (*21*). We only keep the results that the accuracy of the three models is greater than 50% suggested by the above reference. Herein, a total of 4,917 AMPs with a length distribution ranging from 21 to 50 amino acids in 4,697 non-redundant MAGs were identified to explore the global distribution patterns of AMP biosynthetic potential in microbes (Table S1 and Fig 1b). These MAGs were classified into six habitat categories including air, aquatic, terrestrial, human, engineered, and other hosts (Fig. 1a and Table S1). We used only four main habitats (aquatic, terrestrial, human, and engineered, mainly the built environment) in the subsequent analysis because they covered 90% of samples.

In all, 4,364 AMPs were identified from four main habitats based on deep learning predictions. Of four habitats, the human gut had the highest richness of AMPs (Fig. 1c). This conforms to predictions that this habitat has many potential AMP families that are yet to be intensively studied (*22*).

Determining the hosts of AMPs is helpful in mining AMPs with potential applications. Therefore, we identified the AMP hosts by taxonomically annotating each MAG (see "Methods") and then explored the distribution of these hosts across diverse habitats. A great diversity of hosts carried AMPs in various habitats. A total of 85 phyla were identified across all MAGs, but a more limited number of bacterial phyla were dominant (Fig. 1d). The composition of the hosts of AMPs differed greatly between the four habitats at the phylum level. Firmicutes, common in the human gut microbiota (*23*), accounted for a large proportion of hosts in human-associated habitats (53.21%), further reflecting the widespread distribution of AMPs in the human gut. Proteobacteria dominated in engineered (27.71%), aquatic (24.80%), and terrestrial (20.97%) habitats. The hosts of AMPs in aquatic and terrestrial environments were more diverse than those in engineered and human habitats, perhaps as a consequence of higher species diversity in these habitats, allowing more diverse sources of AMPs (*24*). *Lachnospiraceae* was the main host of AMPs (Fig. 1e), which is abundant in the unperturbed adult microbiota and promotes colonization resistance against intestinal pathogens (*25*), pointing to this family as a potentially source of AMPs, suggesting that AMPs might participate in pathogen inhibition. The distribution pattern of most hosts of AMPs indicated habitat specificity, with only a few hosts shared between two or more habitats.

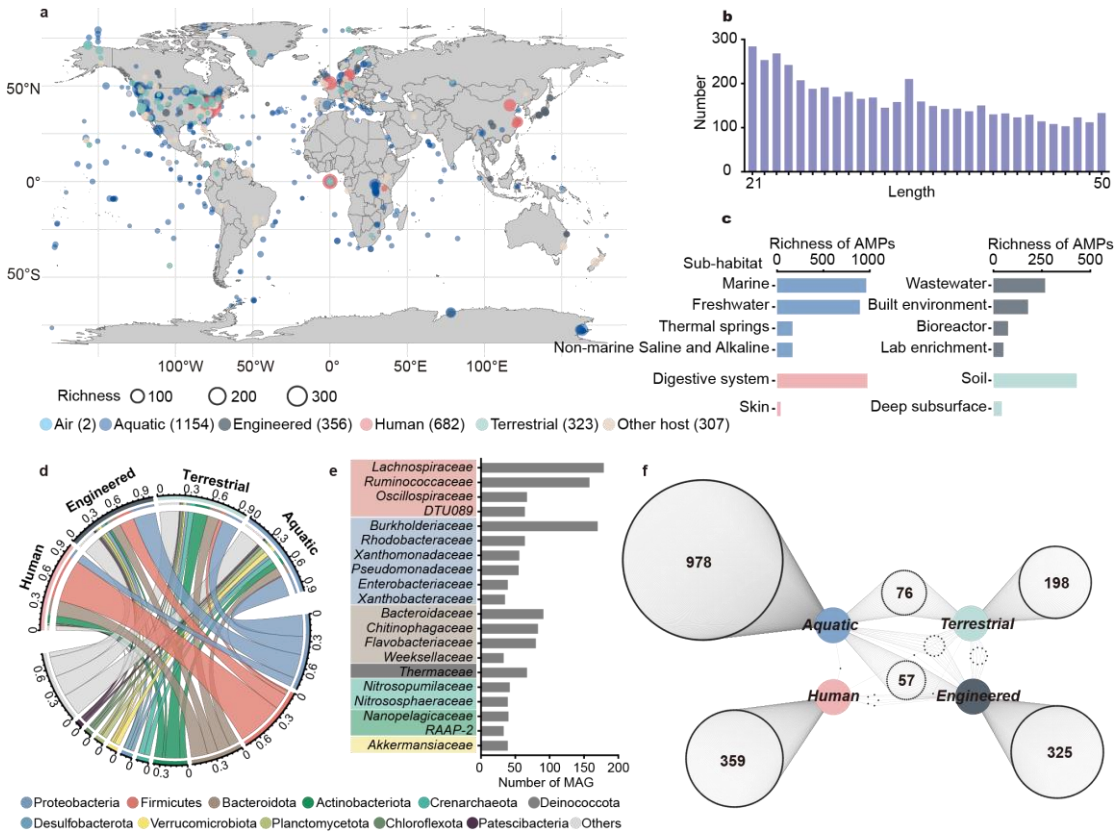

Figure 1. Global distribution patterns of antimicrobial peptides (AMPs) and their hosts. a, Geographical distribution of samples with AMPs, showing richness in various habitats. The size of the dots represents the richness of the AMPs at each sampling site and the colors represents different habitats. The number of samples is in parentheses. Sampling sites without clear geographic coordinates were represented as 0° latitude and 0° longitude. **b,** Length distribution of 4,917 AMPs from the metagenome-assembled genomes (MAGs) in our study. **d,** Composition of AMP hosts in the four main habitats at the phylum level. **e,** Number of MAG in each family (most common 20). **f,** Venn network of shared and specific AMP hosts at the species level in various habitats.

Global patterns of genes conferring resistance to AMPs

A total of 129 genes predicted to confer resistance to AMPs (AMPRGs) were identified from 4,434 non-redundant MAGs. These predictions were based on the AMP resistance

gene database we constructed. This database identified eight different types of resistance phenotypes (Fig. 2a and Table S2; see "Methods"). Of the 129 AMPRGs, 76 were predicted to confer resistance to only one class of AMP, while 53 AMPRGs were predicted to confer resistance to multiple AMP classes.

The abundance of AMPRGs varied across habitats and was generally low in all sub-habitats, except for the built environment and on human skin (Fig. 2b). The abundance of AMPRGs was highest on human skin, but the number of AMPRGs per metagenome sample in the human digestive system was higher than that on human skin ($p < 0.001$, one-way ANOVA test; Fig. 2c). In addition, the high abundance of AMPRGs in the built environment (mainly urban subways) (Fig. 2b) could increase the possibility of AMPRG transmission, in a situation similar to the dissemination of ARGs (*26*). Given that several types of AMPRG might confer resistance to the same AMP or a single AMPRG might confer resistance to multiple AMPs, we summed the relative abundance of AMPRGs in each AMP category for each sample. This revealed an uneven abundance distribution among the various AMPRG types (Fig. 2d). AMPRGs conferring resistance to multiple AMPs were at highest abundance in sub-habitats, except that AMPRGs for colistin and polymyxin B were dominant in both the built environment and on human skin (Fig. 2d). Use of AMPs in these two environments could raise the threat of AMPRGs becoming widespread and potentially cause AMP failure during treatment of infections. For all 129 identified AMPRGs, almost a quarter were specific to one habitat, while one third of AMPRGs were shared among all four

main habitats. These latter AMPRGs were mainly predicted to confer resistance to multiple AMPs and polymyxin B (Fig. 2e).

Because there was a high abundance and diversity of AMPRGs in habitats dominated by humans, we divided all samples into groups of high and low population density based on the threshold of 58 people per square kilometer (*27*) to explore the impact of anthropogenic activities on the distribution of AMPRGs. Almost all AMPRGs could be detected in the high population density environments. These were mainly AMPRGs conferring multiple or polymyxin resistance. Only one AMPRG was specific to the low population density environment (Fig. 2f, h). The abundance of AMPRGs in the high-density environments was significantly higher than that in the low-density environments ($p < 0.001$, two-tailed Welch's *t*-test; Fig. 2g). These results suggest that anthropogenic activities drive increases in the abundance and relative diversity of the AMPRGs.

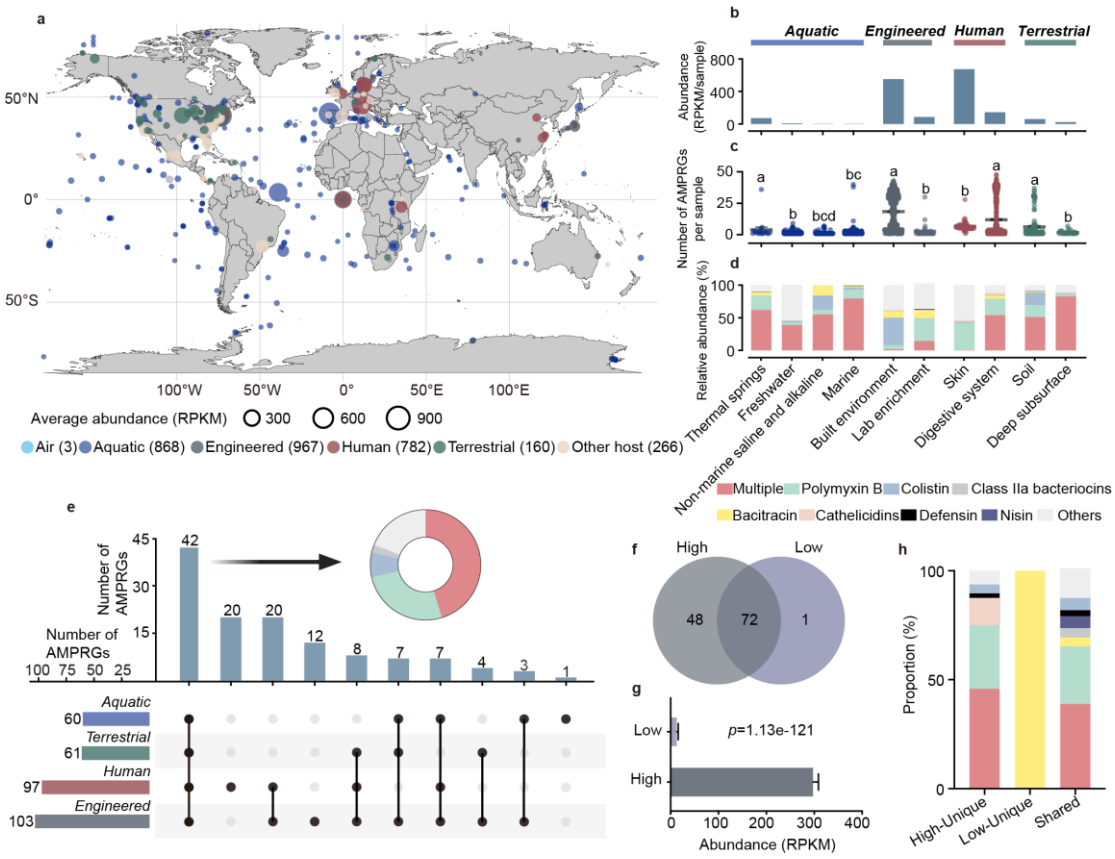

Figure 2. **Global distribution patterns of antimicrobial peptide resistance genes (AMPRGs). a**, Geographic distribution of samples with AMPRG abundance in various habitats. The size of the dot represents the average abundance of the AMPRGs at each sampling site and the colors represent different habitats. The number of samples is shown in parentheses. Sampling sites without clear geographic coordinates were represented as 0° latitude and 0° longitude. **b** to **d**, Abundance (b), number (c), and proportion (d) of AMPRGs grouped by sub-habitats. Abundance was calculated as Reads Per Kilobase per Million mapped reads (RPKM). Different letters indicate group differences between groups determined using a one-way ANOVA test ($p < 0.05$). **e**, Upset plot of shared and specific AMPRGs in the four main habitats. The circle represents the classification of AMPRGs shared by the four habitats. **f**, Venn plot of

188 shared and specific AMPRGs in low- or high-intensity human activities. **g**, The
189 abundance of AMPRGs in low- or high-intensity human activities. The *p* value
190 represents the statistical significance (two-tailed Welch's *t*-test). **h**, The classification
191 of shared and specific AMPRGs in low- or high-intensity human activities.
192
193 **Distribution of AMPRGs host and mobile genetic elements (MGEs)**
194 After taxonomic profiling, 27 bacterial phyla carrying AMPRGs were identified across
195 all MAGs (Table S2). *Gammaproteobacteria* were the main carrier of AMPRGs, unlike
196 AMPs, in all habitats (Fig. 3a), with *Enterobacteriaceae* carrying the most AMPRGs,
197 followed by *Staphylococcaceae* (Fig. 3b). Some AMPRG types tended to be carried by
198 specific bacteria, showing host-specific patterns at the species level. For example, the
199 host of bacitracin resistance genes were almost always *Enterobacteriaceae*.
200 *Enterobacteriaceae* also carried other types of AMPRGs, indicating that this taxon was
201 an important reservoir of AMPRGs, much as it is also an important reservoir of
202 antibiotic resistance. Most AMPRG-carrying hosts were unique to particular habitats,
203 especially in engineered and human environments, with a few hosts shared by two or
204 three habitats, showing habitat specificity, further illustrated by the large differences in
205 host composition among the 42 shared AMPRGs in the four habitats (Figs. 3c and S1).
206 Only one host was found in all habitats. Moreover, AMPRG hosts in human habitats
207 were more prone to be shared with the engineered environment than with other habitats.
208 Different habitats contained the same dominant host of AMPRGs, suggesting that
209 AMPRGs confer site-specific advantages on hosts.

The analysis of MGEs by the MAG approaches is largely invalid because these MGEs can have different coverage and nucleotide compositions, they are challenging to metagenomically bin correctly (*28*). Therefore, we collected 47,169 bacterial genomes in NCBI database from aquatic, engineered, human, and terrestrial habitats. Finally, a total of 13,091 bacterial genomes were annotated to the AMPRGs and 6,001 of them were identified as pathogens (Table S3). Then, we selected 5 kb upstream and downstream around each AMPRGs to determine the most closely related MGEs most likely to be involved in AMPRGs transfer from a bioinformatic perspective (*29*). A total of 1439 MGEs were identified in 4,808 bacterial genomes, most of which were transposases (918/1439) (Table S4). The terrestrial and engineered habitats possessed many specific MGEs (Fig. S2). Only *Gammaproteobacteria*, *Alphaproteobacteria*, *Betaproteobacteria*, *Firmicutes*, *Bacteroidota*, and *Acinobacteriota* carried MGEs, with *Gammaproteobacteria* carrying a large proportion (63.4%) (Fig. S3a). Notably, 72.7% of AMPRGs transported between bacteria within the *Gammaproteobacteria* (Fig. 3d). The AMPRG hosts in the various habitats were also notably dominated by *Gammaproteobacteria* (Fig. 3a). The potentially mobile AMPRGs conferring multiple AMP resistance (40.3%) and polymyxin B (26.4%) resistance were dominant in *Gammaproteobacteria* (Fig. S3b). Of the 129 evaluated AMPRGs, 75 (58.1%) were identified as potentially mobile AMPRGs in diverse habitats, mainly the engineered and human environments (Table S5).

Unique environmental conditions in different ecosystems can affect the distribution of MGEs (*30, 31*), altering the composition of AMPRG-carrying hosts in

diverse habitats. The AMPRG hosts in various habitats in our study, however, were all dominated by *Gammaproteobacteria*; we assumed that prevalence of all evaluated AMPRGs carried by MGEs in all metagenomic samples was less than 25% due to ecological barriers, even though they are carried by more MGEs (Fig. S4). Moreover, the mobile AMPRGs in bacterial genomes were shared between fewer bacterial species, suggesting fewer transfer events per AMPRG (Fig. 3e). A total of 54.5% (54) of mobile AMPRGs were shared in nonpathogens and pathogens (Fig. S5), suggesting that pathogens may acquire AMP resistance by horizontal gene transfer from nonpathogens.

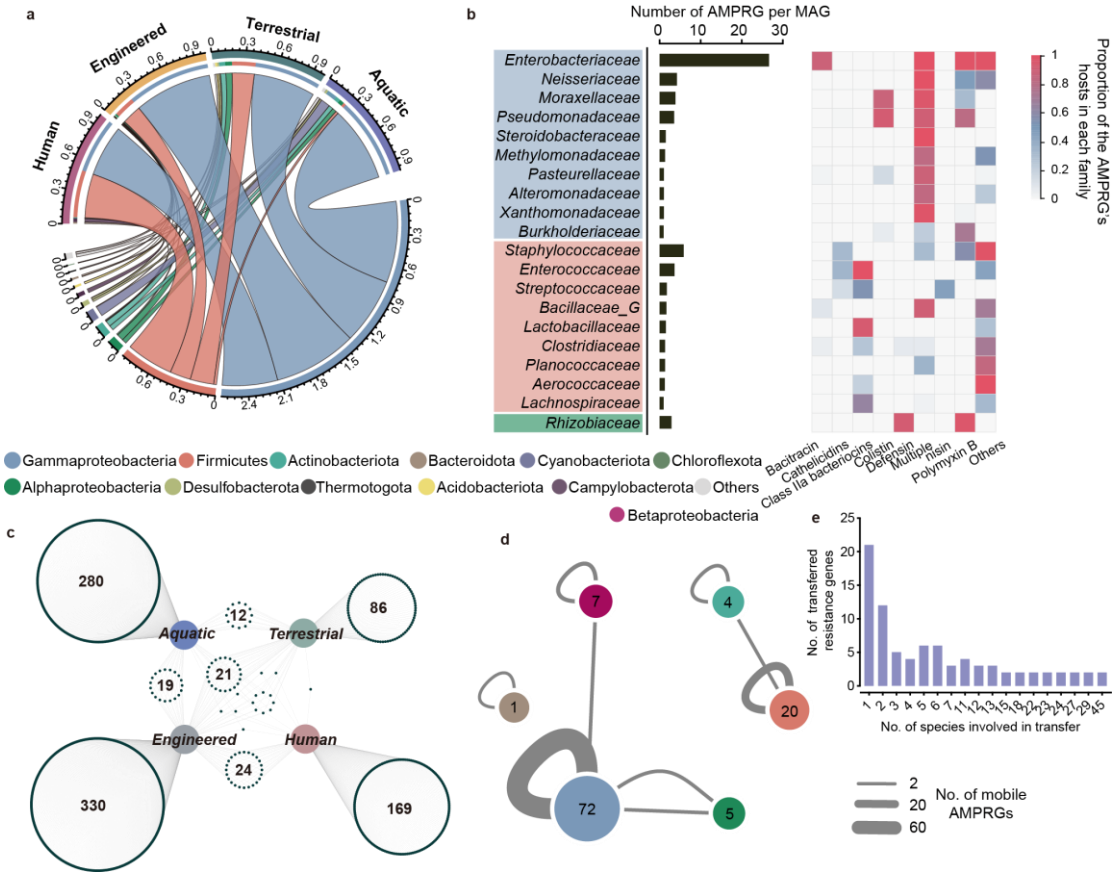

Figure 3. **Hosts and mobility of AMPRGs**. **a**, Composition of AMPRG hosts in four main habitats at the phylum level (and classes in *Proteobacteria*). **b**, Number of AMPRGs per MAG in each family (most common 20) plotted against a heatmap for

the different families of AMPRGs. **c**, Venn network of shared and specific hosts containing AMPRGs at the species level in various habitats. **d**, Network representation of the mobile AMPRGs shared between hosts. Node size and numbers indicate the total number of mobile AMPRGs in each host. Straight and curved lines represent genes that were shared between hosts at different phyla (classes in *Proteobacteria*) and the same phylum (classes in *Proteobacteria*), respectively. Line thickness represents the number of mobile AMPRGs shared among the hosts. Only *Gammaproteobacteria*, *Alphaproteobacteria*, *Betaproteobacteria*, *Firmicutes*, *Bacteroidota*, and *Acinobacteriota* carried MGEs. **e**, Unique mobile AMPRGs were transferred between different numbers of species, only values with at least two transferred AMPRGs were shown.

Determination of the threat of AMPRGs to human health

We quantified the threat of each AMPRG to human health based on its associated pathogenicity (proportion of pathogenic hosts), human accessibility, and mobility (number of MGEs) (see "Methods"), in a similar fashion to the threat evaluation of antibiotic resistance genes (*27*). Indicators of human pathogenicity and accessibility covered more than 75% of the AMPRGs, with more than half being potentially mobile (Fig. 4a and Table S6). Ultimately, 44.96% AMPRGs posed a significant threat to human health based on our evaluation framework (Fig. 4b). We ranked all the AMPRGs according to their threat and divided the AMPRGs with risks > 0 into four categories (Q1: top 25%, Q2: 25-50%, Q3: 50-75% and Q4: bottom 25%). The multiple AMPs

and polymyxin B resistance genes had a wide distribution in the four risk ranks (Fig. 4c), and the average risks for Human Cathepsin G and multiple AMPs resistance genes were much higher than others (Fig. 4d). We also calculated the threat for each sample in various habitats, and the results indicated that the samples from engineered environments had the highest threat, followed by human habitats (Fig. 4e).

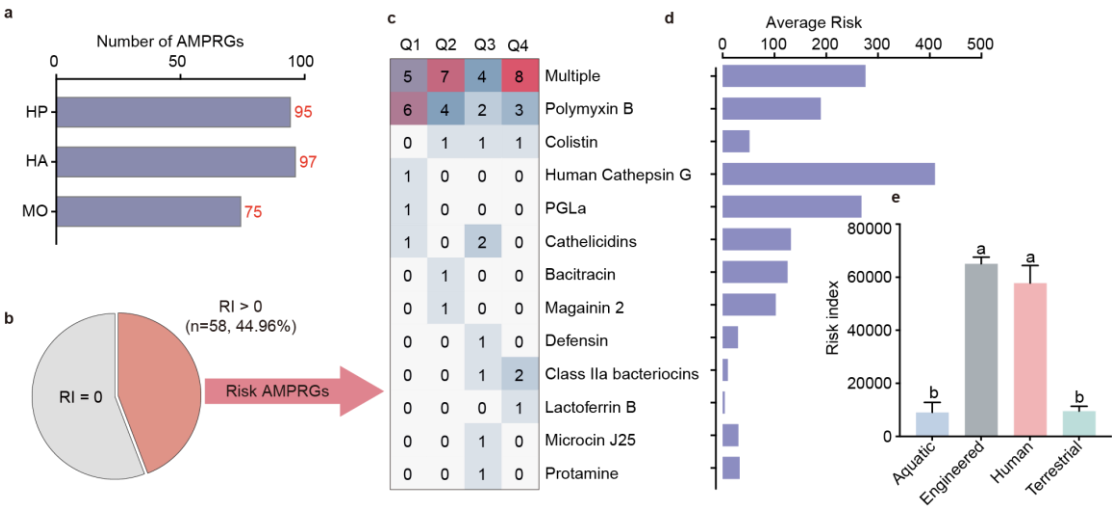

Figure 4. **Health threat evaluation of AMPRGs to human health**. **a**, Number of AMPRGs with risk index (RI) > 0 in each indicator, including human pathogenicity (HP), human accessibility (HA), and mobility (MO). **b**, 44.96% of AMPRGs had a risk index RI > 0. The AMPRGs with an RI > 0 were partitioned three categories based on the rank of their RI: (Q1: top 25%, Q2: 25-50%, Q3: 50-75% and Q4: bottom 25%). **c**, The composition of risk AMPRGs in different AMPRG types. **d**, The average risk of the AMPRGs in each class. **e**, RI of each sample in the four major habitats. The engineered environments (mainly built environment) had the highest risk of antimicrobial peptide resistance. The RI of each sample was calculated as the combination of abundance and the RI of the AMPRGs. Data were showed as mean ± standard error of mean (SEM). Different letters indicate differences between groups

284 determined by a one-way ANOVA (p <0.05).

285

286 **Discussion**

287 The rapid spread of resistance to conventional antibiotics in pathogens has led to one
288 of the most urgent global health concerns. The development of novel antimicrobial
289 agents such as AMPs is the most promising component of the defensive strategy against
290 these pathogens. Soil, human, and marine microbiota contain a large number of bacteria
291 competing for nutrients. This competition could drive the evolution of new AMPs and
292 AMPRGs (*32*). Any of these microbiotas could be reservoirs of potentially useful AMPs
293 (*33*).

294 Traditional techniques for culturing microbes, however, limit the exploitation of
295 unculturable microorganisms that represent sources of biodiversity of unexploited
296 AMPs (*34*). We focused on mining the potential for novel AMPs in microbes.
297 Previously, tools like sequence alignment-based BLAST or hidden Markov models for
298 identifying conserved motifs and domains were typical methods for identifying proteins
299 with similar functions. However, these approaches are difficult to work with in the
300 identification of short peptides like AMPs, especially those that lack significant
301 homology. Many machine learning approaches have been developed recently to address
302 this challenge. Ma et al (*21*) successfully used three different natural language
303 processing (NLP) models (ATT, LSTM, and BERT) for AMP-mining in human gut
304 microbiomes. Herein, we applied the unified pipeline combined with these models to
305 predict AMPs in the global microbiome, resulting in 4,917 AMPs. Please note all AMPs

306 predicted in this study were potential AMPs, so it is necessary to verify their functions
307 in future experiments. The AMPs were identified in diverse habitats, especially in the
308 human digestive system (mainly the human gut). The diversity of AMPs probably
309 reflects the ability of species to adapt to the unique microbial environment of the niche
310 they occupy, as a single mutation can greatly change the biological activity of each
311 peptide. Complex microbial ecosystems such as the human gut are locations where
312 long-term competition and co-evolution could generate antimicrobial compounds such
313 as AMPs (*35*). Therefore, the human gut microbiome could serve as a source of AMPs
314 that could be used to address the antibiotic resistance crisis.

315 Some AMPs in the current databases have been experimentally verified, but many
316 of the AMPs in these databases are based on prediction by homology. This is a key
317 knowledge gap, and more information is required to understand how differences in
318 sequence diversity affects biological functions of presumptive AMPs, how these AMPs
319 might then exert their activity against different taxa, and the functional mechanisms by
320 which they operate. Therefore, matching AMPs with their corresponding resistance
321 genes is a dilemma. Further study of the functions of presumptive AMPs is therefore
322 necessary. Classical *in vitro* assays can evaluate AMPs specific functions (*3*). However,
323 results *in vitro* assays can be strongly influenced by assay conditions, such as pH, salt
324 concentrations, nutrients and osmotic stress, and might not accurately reflect the AMPs
325 activity *in vivo*. Synergistic effects among different AMPs *in vivo* may also lower the
326 possibility of resistance evolving (*36*). Some AMPs form pores on the bacterial
327 membrane that then enable other AMPs to enter and attack the cells (*37*), implying that

synergistic and efficacious AMP cocktails from several of these structural classes could be deployed. In summary, the function of AMPs should be tested *in vivo* to more accurately account for the mechanisms of action of AMPs.

An unprecedented view of the global distribution of resistance genes for antimicrobial peptides was generated in this study. A total of 129 AMPRGs were identified in 4434 MAGs. By examining various sub-habitats of each habitat, we demonstrated that AMPRGs were most abundant in the built environment and on human skin. Our results indicate that anthropogenic activities have promoted the spread of AMPRGs. AMPRG abundance, however, was generally low in other sub-habitats.

It is still unclear why AMPs are widespread in different habitats, yet resistance genes are at low abundance. The pharmacology of AMPs might contribute to a lower likelihood of developing resistance (*38*). Unlike antibiotics, most AMPs function by modifying bacterial cell membranes (*39*), and can kill bacteria within minutes (*40*). This could prevent bacterial growth and accumulation of potential resistance mutations, especially if the action is all or nothing. Such a situation might mean that there is only a very narrow window of sub-inhibitory concentrations where genetic variants can be selected. Additionally, to resist antimicrobial peptide attack, microorganisms might have to alter the lipid composition and/or organization of their cell membranes, which could be very costly. AMP resistance is often generated via nonspecific mechanisms, and so fewer mutational pathways are available by which resistance to AMPs may evolve (*41*).

We also identified the hosts of AMPRGs in diverse habitats. *Gammaproteobacteria* were the dominant carrier of AMPRGs in all habitats, and the distribution pattern of hosts was habitat-specific at the species level. *Gammaproteobacteria* also carried the most mobile genetic elements associated with AMPRGs. We reasoned that the lateral transfer of AMPRGs among microorganisms could be constrained by taxonomic barriers.

AMPRGs act primarily by modifying cell membranes. Since membrane formation is the result of networks of interacting proteins that are highly interconnected, it might be difficult for lateral gene transfer to modify these complex cellular subsystems (*42*, *43*). However, even so, we do not conclude that AMPs in the clinic or for clinical use are resistance free. For example, once the colistin resistance gene became plasmid borne, it has spread rapidly around the globe (*44*). Consequently, clinical use of AMPs could accelerate resistance dissemination by selecting for complex mobile genetic elements, as has been seen in the antibiotic crisis.

We should learn from the current antibiotic crisis, and prioritize the risks posed by known AMPRGs before widespread adoption of AMPs as therapeutic agents. In doing so, we might avoid the premature appearance of resistance to this promising class of antibacterial agents. With this in mind, we quantified the threat of each of the AMPRGs we detected, based on their human accessibility, mobility and pathogenicity associated with their hosts. Near half of the 129 AMPRGs we detected were determined to pose some risk of compromising future AMP therapies. Most of these genes conferred resistance to multiple AMPs and polymyxin B. For these AMPRGs, we could apply

AMP cocktails to minimize the effect of resistance selection or adjuvants to maximize the cost of developing AMP resistance (*18*).

Our study provides a view of the AMP biosynthetic potential in the global microbiome, which will be conducive to mining the resources of AMPs and minimizing the risk of the development of resistance to endogenous host defenses. Apparently low levels of AMP resistance and the potential for phylogenetic barriers to horizontal transfer for AMPRGs do not mean that risks to human health will not occur. However, by learning the lessons of antibiotic resistance, we might be able to prevent the appearance of resistance to AMPs in the future.

Methods

Data sets

Nayfach et al (*20*) constructed a database of 52,515 MAGs recovered from environmentally diverse metagenomes. We downloaded these MAGs with permission from https://genome.jgi.doe.gov/GEMs. The database also provided information for the taxonomies and habitats of MAGs.

We constructed an AMP resistance genes database, including the genes with experimentally confirmed influence on AMP susceptibility, collected from the references provided by Kintses et al (*45*), and the genes associated with AMP resistance in the Comprehensive Antibiotic Resistance Database (CARD) (*46*). The remaining genes in CARD not associated with AMP resistance were used as a comparable data set.

We collected 47,169 bacterial genomes in NCBI database from aquatic (12,037), engineered (11,204), human-associated (12,007), and terrestrial (11,921) habitat to evaluate the mobility and human pathogenicity of AMPRGs more accurately.

Prediction of AMPs in MAGs

Ma et al (*21*) used AMP sequences from four AMP databases, APD (*6*), ADAM(*47*), CAMP(*48*), and LAMP(*49*), that contained many AMPs from different species as input data and combined several deep-learning-based natural language processing models, whose performance they optimized to design a pipeline for AMP identification. The combination of ATT, LSTM, and BERT models showed the highest precision. Therefore, we used the combination of the above three models to identify AMPs in 52,515 MAGs.

Annotation and calculation of the abundance of AMPRGs in the MAGs

The AMPRGs were annotated with the AMP resistance genes we constructed with strict parameters (e-value $\leq 10^{-10}$, $\geq$70% identity at the protein level, and 60% query coverage) using Diamond (v2.0.11.149) (*50*). Reads in each MAG were mapped to AMPRGs using BWA (v0.7.13) (*51*) and the unmapped reads were removed using Samtools (v1.3.1) (*52*). The abundance of AMPRGs in each MAG was calculated as Reads Per Kilobase per Million mapped reads based on the number of mapped reads and length of genes. To identify the hosts of AMPRGs, we taxonomically annotated the MAGs containing AMPRGs using the methods described by Nayfach et al (*20*). Pathogenic

hosts were identified using the A-to-Z database (*53*). We manually reclassified AMPRGs based on the AMPs to which they confer resistance, including polymyxin B, colistin, class IIa bacteriocins, bacitracin, cathelicidins, defensin and nisin. AMPRGs providing resistance to more than one AMP class were classified as multiple resistance determinants.

Annotation of the MGEs in the MAGs

We extracted 5000 bp (*29*) upstream and downstream of the AMPRGs identified in the bacterial genomes to annotate the MGEs. We used BLASTN (e-value $\leq 10^{-10}$, identity 80%, query coverage 80%) to annotate inserted sequences (ISs) with the ISfinder database (*54*). We annotated transposases and integrases with the NCBI reference sequences that were clustered using CD-HIT (v4.7) (*55*), with a threshold ≥90%, an e-value $\leq 10^{-10}$, and minimum amino acid identity of 60% over 60% query coverage (*56*) using Diamond (v2.0.11.149).

Acquisition of population data

We collected the population density of each sampling site from the SEDAC Population Estimation Service (*56*) and classified all samples into high (>58 people/km^2) and low (≤58 people/km^2) population densities based on the global average density (*27*). We only considered the samples for which clear coordinate information was available.

Determination of AMPRG threat to human health

We quantified the threat of AMPRGs to human health based on three indicators, as calculated in our previous study (*27*): human accessibility, mobility, and human pathogenicity. The accessibility of AMPRGs to humans was the ability to transfer from the environment to humans, calculated as the average abundance and prevalence of the AMPRGs in habitats associated with humans. AMPRG mobility was considered the ability to transfer between hosts by horizontal transfer, calculated as the sum of the number of MGEs (ISs, transposases, and integrases) detected in each bacterial genome. The pathogenicity of an AMPRG was determined using the A-to-Z database (*53*) and defined as the proportion of pathogenic hosts containing it. Finally, we ranked all the AMPRGs by their threat to human health and divided the AMPRGs with risks >0 into four categories (Q1, Q2, Q3 and Q4).

Statistical analysis and visualization

The data are presented as means ± standard errors of the means. A one-way ANOVA was performed in IBM SPSS Statistics v25.0.0 (IBM Corp., Armonk, NY, USA). Values were considered to differ significantly at $p < 0.05$. Two-tailed Welch's *t*-tests were performed in Excel Analysis Tools (Microsoft Corporation, Redmond, USA). Maps of the geographic distribution of the samples were constructed using the ggmap and ggplot2 packages in R (v4.2.1). Venn diagrams, upset plots, and heatmaps were generated using the OmicStudio tools at https://www.omicstudio.cn. Venn network plots of shared and specific hosts containing AMPs and AMPRGs in diverse habitats

were visualized using Cytoscape (v3.9.0). Other plots were visualized using GraphPad Prism (v9.1.1).

Data availability

All metagenome-assembled genomes used in this study are available in IMG/M portal. The bacterial genomes are available in NCBI database.

Code availability

All the scripts and codes for gene annotation, abundance calculation, and visualization used in this study are available online at https://github.com/bingfeng-micro.

References

1. C. Lepore, L. Silver, U. Theuretzbacher, J. Thomas, D. Visi, The small-molecule antibiotics pipeline: 2014-2018. *Nat. Rev. Drug Discov.* **18**, 739–740 (2019).
2. J. O'Neill, Tackling drug-resistant infections globally: final report and recommendations (2016).
3. B. P. Lazzaro, M. Zasloff, J. Rolff, Antimicrobial peptides: Application informed by evolution. *Science*. **368**, eaau5480 (2020).
4. M. Zasloff, Antimicrobial peptides of multicellular organisms. *Nature*. **415**, 389–395 (2002).
5. D. I. Andersson, D. Hughes, J. Z. Kubicek-Sutherland, Mechanisms and consequences of bacterial resistance to antimicrobial peptides. *Drug Resist. Updat.*

26, 43–57 (2016).

6. G. Wang, X. Li, Z. Wang, APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res.* **44**, D1087–D1093 (2016).

7. H. B. Koo, J. Seo, Antimicrobial peptides under clinical investigation. *Pept. Sci.* **111**, e24122 (2019).

8. X. Xu, R. Lai, The chemistry and biological activities of peptides from amphibian skin secretions. *Chem. Rev.* **115**, 1760–1846 (2015).

9. M. Simmaco, G. Mignogna, D. Barra, Antimicrobial peptides from amphibian skin: What do they tell us? *Pept. Sci.* **47**, 435–450 (1998).

10. J. M. Conlon, The contribution of skin antimicrobial peptides to the system of innate immunity in anurans. *Cell Tissue Res.* **343**, 201–212 (2011).

11. P. Nicolas, D. Vanhoye, M. Amiche, Molecular strategies in biological evolution of antimicrobial peptides. *Antimicrob. Pept. II*. **24**, 1669–1680 (2003).

12. J. M. Conlon, M. Mechkarska, Host-defense peptides with therapeutic potential from skin secretions of frogs from the family pipidae. *Pharmaceuticals*. **7**, 58–77 (2014).

13. A. A. Bahar, D. Ren, Antimicrobial Peptides. *Pharmaceuticals*. **6**, 1543–1575 (2013).

14. J. S. Gunn, S. S. Ryan, J. C. Van Velkinburgh, R. K. Ernst, S. I. Miller, Genetic and functional analysis of a PmrA-PmrB-regulated locus necessary for lipopolysaccharide modification, antimicrobial peptide resistance, and oral virulence of Salmonella entericaSerovar Typhimurium. *Infect. Immun.* **68**, 6139–

6146 (2000).

15. S. Stumpe, R. Schmid, D. L. Stephens, G. Georgiou, E. P. Bakker, Identification of OmpT as the protease that hydrolyzes the antimicrobial peptide protamine before it enters growing cells of *Escherichia coli*. *J. Bacteriol.* **180**, 4002–4006 (1998).

16. V. Nizet, Antimicrobial peptide resistance mechanisms of human bacterial pathogens. *Curr. Issues Mol. Biol.* **8**, 11–26 (2006).

17. J. Z. Kubicek-Sutherland, H. Lofton, M. Vestergaard, K. Hjort, H. Ingmer, D. I. Andersson, Antimicrobial peptide exposure selects for *Staphylococcus aureus* resistance to human defence peptides. *J. Antimicrob. Chemother.* **72**, 115–127 (2017).

18. P. K. Jangir, L. Ogunlana, R. C. MacLean, Evolutionary constraints on the acquisition of antimicrobial peptide resistance in bacterial pathogens. *Trends Microbiol.* **29**, 1058–1061 (2021).

19. C. D. Fjell, J. A. Hiss, R. E. W. Hancock, G. Schneider, Designing antimicrobial peptides: form follows function. *Nat. Rev. Drug Discov.* **11**, 37–51 (2012).

20. S. Nayfach, S. Roux, R. Seshadri, D. Udwary, N. Varghese, F. Schulz, D. Wu, D. Paez-Espino, I.-M. Chen, M. Huntemann, K. Palaniappan, J. Ladau, S. Mukherjee, T. B. K. Reddy, T. Nielsen, E. Kirton, J. P. Faria, J. N. Edirisinghe, C. S. Henry, S. P. Jungbluth, D. Chivian, P. Dehal, E. M. Wood-Charlson, A. P. Arkin, S. G. Tringe, A. Visel, H. Abreu, S. G. Acinas, E. Allen, M. A. Allen, L. V. Alteio, G. Andersen, A. M. Anesio, G. Attwood, V. Avila-Magaña, Y. Badis, J. Bailey, B.

Baker, P. Baldrian, H. A. Barton, D. A. C. Beck, E. D. Becraft, H. R. Beller, J. M. Beman, R. Bernier-Latmani, T. D. Berry, A. Bertagnolli, S. Bertilsson, J. M. Bhatnagar, J. T. Bird, J. L. Blanchard, S. E. Blumer-Schuette, B. Bohannan, M. A. Borton, A. Brady, S. H. Brawley, J. Brodie, S. Brown, J. R. Brum, A. Brune, D. A. Bryant, A. Buchan, D. H. Buckley, J. Buongiorno, H. Cadillo-Quiroz, S. M. Caffrey, A. N. Campbell, B. Campbell, S. Carr, J. Carroll, S. C. Cary, A. M. Cates, R. A. Cattolico, R. Cavicchioli, L. Chistoserdova, M. L. Coleman, P. Constant, J. M. Conway, W. P. Mac Cormack, S. Crowe, B. Crump, C. Currie, R. Daly, K. M. DeAngelis, V. Denef, S. E. Denman, A. Desta, H. Dionisi, J. Dodsworth, N. Dombrowski, T. Donohue, M. Dopson, T. Driscoll, P. Dunfield, C. L. Dupont, K. A. Dynarski, V. Edgcomb, E. A. Edwards, M. S. Elshahed, I. Figueroa, B. Flood, N. Fortney, C. S. Fortunato, C. Francis, C. M. M. Gachon, S. L. Garcia, M. C. Gazitua, T. Gentry, L. Gerwick, J. Gharechahi, P. Girguis, J. Gladden, M. Gradoville, S. E. Grasby, K. Gravuer, C. L. Grettenberger, R. J. Gruninger, J. Guo, M. Y. Habteselassie, S. J. Hallam, R. Hatzenpichler, B. Hausmann, T. C. Hazen, B. Hedlund, C. Henny, L. Herfort, M. Hernandez, O. S. Hershey, M. Hess, E. B. Hollister, L. A. Hug, D. Hunt, J. Jansson, J. Jarett, V. V. Kadnikov, C. Kelly, R. Kelly, W. Kelly, C. A. Kerfeld, J. Kimbrel, J. L. Klassen, K. T. Konstantinidis, L. L. Lee, W.-J. Li, A. J. Loder, A. Loy, M. Lozada, B. MacGregor, C. Magnabosco, A. Maria da Silva, R. M. McKay, K. McMahon, C. S. McSweeney, M. Medina, L. Meredith, J. Mizzi, T. Mock, L. Momper, M. A. Moran, C. Morgan-Lang, D. Moser, G. Muyzer, D. Myrold, M. Nash, C. L. Nesbø, A. P. Neumann, R. B.

Neumann, D. Noguera, T. Northen, J. Norton, B. Nowinski, K. Nüsslein, M. A. O'Malley, R. S. Oliveira, V. Maia de Oliveira, T. Onstott, J. Osvatic, Y. Ouyang, M. Pachiadaki, J. Parnell, L. P. Partida-Martinez, K. G. Peay, D. Pelletier, X. Peng, M. Pester, J. Pett-Ridge, S. Peura, P. Pjevac, A. M. Plominsky, A. Poehlein, P. B. Pope, N. Ravin, M. C. Redmond, R. Reiss, V. Rich, C. Rinke, J. L. M. Rodrigues, W. Rodriguez-Reillo, K. Rossmassler, J. Sackett, G. H. Salekdeh, S. Saleska, M. Scarborough, D. Schachtman, C. W. Schadt, M. Schrenk, A. Sczyrba, A. Sengupta, J. C. Setubal, A. Shade, C. Sharp, D. H. Sherman, O. V. Shubenkova, I. N. Sierra-Garcia, R. Simister, H. Simon, S. Sjöling, J. Slonczewski, R. S. Correa de Souza, J. R. Spear, J. C. Stegen, R. Stepanauskas, F. Stewart, G. Suen, M. Sullivan, D. Sumner, B. K. Swan, W. Swingley, J. Tarn, G. T. Taylor, H. Teeling, M. Tekere, A. Teske, T. Thomas, C. Thrash, J. Tiedje, C. S. Ting, B. Tully, G. Tyson, O. Ulloa, D. L. Valentine, M. W. Van Goethem, J. VanderGheynst, T. J. Verbeke, J. Vollmers, A. Vuillemin, N. B. Waldo, D. A. Walsh, B. C. Weimer, T. Whitman, P. van der Wielen, M. Wilkins, T. J. Williams, B. Woodcroft, J. Woolet, K. Wrighton, J. Ye, E. B. Young, N. H. Youssef, F. B. Yu, T. I. Zemskaya, R. Ziels, T. Woyke, N. J. Mouncey, N. N. Ivanova, N. C. Kyrpides, E. A. Eloe-Fadrosh, A genomic catalog of Earth's microbiomes. *Nat. Biotechnol.* **39**, 499–509 (2021).

21. Y. Ma, Z. Guo, B. Xia, Y. Zhang, X. Liu, Y. Yu, N. Tang, X. Tong, M. Wang, X. Ye, J. Feng, Y. Chen, J. Wang, Identification of antimicrobial peptides from the human gut microbiome using deep learning. *Nat. Biotechnol.* **40**, 921–931 (2022).

22. H. Sberro, B. J. Fremin, S. Zlitni, F. Edfors, N. Greenfield, M. P. Snyder, G. A.

568 Pavlopoulos, N. C. Kyrpides, A. S. Bhatt, Large-scale analyses of human
569 microbiomes reveal thousands of small, novel genes. *Cell*. **178**, 1245-1259.e14
570 (2019).
571 23. M. Arumugam, J. Raes, E. Pelletier, D. Le Paslier, T. Yamada, D. R. Mende, G.
572 R. Fernandes, J. Tap, T. Bruls, J.-M. Batto, M. Bertalan, N. Borruel, F. Casellas,
573 L. Fernandez, L. Gautier, T. Hansen, M. Hattori, T. Hayashi, M. Kleerebezem, K.
574 Kurokawa, M. Leclerc, F. Levenez, C. Manichanh, H. B. Nielsen, T. Nielsen, N.
575 Pons, J. Poulain, J. Qin, T. Sicheritz-Ponten, S. Tims, D. Torrents, E. Ugarte, E.
576 G. Zoetendal, J. Wang, F. Guarner, O. Pedersen, W. M. de Vos, S. Brunak, J. Doré,
577 M. Antolín, F. Artiguenave, H. M. Blottiere, M. Almeida, C. Brechot, C. Cara, C.
578 Chervaux, A. Cultrone, C. Delorme, G. Denariaz, R. Dervyn, K. U. Foerstner, C.
579 Friss, M. van de Guchte, E. Guedon, F. Haimet, W. Huber, J. van Hylckama-Vlieg,
580 A. Jamet, C. Juste, G. Kaci, J. Knol, K. Kristiansen, O. Lakhdari, S. Layec, K. Le
581 Roux, E. Maguin, A. Mérieux, R. Melo Minardi, C. M'rini, J. Muller, R. Oozeer,
582 J. Parkhill, P. Renault, M. Rescigno, N. Sanchez, S. Sunagawa, A. Torrejon, K.
583 Turner, G. Vandemeulebrouck, E. Varela, Y. Winogradsky, G. Zeller, J.
584 Weissenbach, S. D. Ehrlich, P. Bork, MetaHIT Consortium (additional members),
585 Enterotypes of the human gut microbiome. *Nature*. **473**, 174–180 (2011).
586 24. L. Flandroy, T. Poutahidis, G. Berg, G. Clarke, M.-C. Dao, E. Decaestecker, E.
587 Furman, T. Haahtela, S. Massart, H. Plovier, Y. Sanz, G. Rook, The impact of
588 human activities and lifestyles on the interlinked microbiota and health of humans
589 and of ecosystems. *Sci. Total Environ.* **627**, 1018–1038 (2018).

25. M. T. Sorbara, E. R. Littmann, E. Fontana, T. U. Moody, C. E. Kohout, M. Gjonbalaj, V. Eaton, R. Seok, I. M. Leiner, E. G. Pamer, Functional and Genomic Variation between Human-Derived Isolates of Lachnospiraceae Reveals Inter- and Intra-Species Diversity. *Cell Host Microbe*. **28**, 134-146.e4 (2020).

26. D. Danko, D. Bezdan, E. E. Afshin, S. Ahsanuddin, C. Bhattacharya, D. J. Butler, K. R. Chng, D. Donnellan, J. Hecht, K. Jackson, K. Kuchin, M. Karasikov, A. Lyons, L. Mak, D. Meleshko, H. Mustafa, B. Mutai, R. Y. Neches, A. Ng, O. Nikolayeva, T. Nikolayeva, E. Png, K. A. Ryon, J. L. Sanchez, H. Shaaban, M. A. Sierra, D. Thomas, B. Young, O. O. Abudayyeh, J. Alicea, M. Bhattacharyya, R. Blekhman, E. Castro-Nallar, A. M. Cañas, A. D. Chatziefthimiou, R. W. Crawford, F. De Filippis, Y. Deng, C. Desnues, E. Dias-Neto, M. Dybwad, E. Elhaik, D. Ercolini, A. Frolova, D. Gankin, J. S. Gootenberg, A. B. Graf, D. C. Green, I. Hajirasouliha, J. J. A. Hastings, M. Hernandez, G. Iraola, S. Jang, A. Kahles, F. J. Kelly, K. Knights, N. C. Kyrpides, P. P. Łabaj, P. K. H. Lee, M. H. Y. Leung, P. O. Ljungdahl, G. Mason-Buck, K. McGrath, C. Meydan, E. F. Mongodin, M. O. Moraes, N. Nagarajan, M. Nieto-Caballero, H. Noushmehr, M. Oliveira, S. Ossowski, O. O. Osuolale, O. Özcan, D. Paez-Espino, N. Rascovan, H. Richard, G. Rätsch, L. M. Schriml, T. Semmler, O. U. Sezerman, L. Shi, T. Shi, R. Siam, L. H. Song, H. Suzuki, D. S. Court, S. W. Tighe, X. Tong, K. I. Udekwu, J. A. Ugalde, B. Valentine, D. I. Vassilev, E. M. Vayndorf, T. P. Velavan, J. Wu, M. M. Zambrano, J. Zhu, S. Zhu, C. E. Mason, N. Abdullah, M. Abraao, A. Adel, M. Afaq, F. S. Al-Quaddoomi, I. Alam, G. E. Albuquerque, A. Alexiev, K. Ali, L. E.

Alvarado-Arnez, S. Aly, J. Amachee, M. G. Amorim, M. Ampadu, M. A.-F. Amran, N. An, W. Andrew, H. Andrianjakarivony, M. Angelov, V. Antelo, C. Aquino, Á. Aranguren, L. F. Araujo, H. F. Vasquez Arevalo, J. Arevalo, C. Arnan, L. E. Alvarado Arnez, F. Arredondo, M. Arthur, F. Asenjo, T. S. Aung, J. Auvinet, N. Aventin, S. Ayaz, S. Baburyan, A.-M. Bakere, K. Bakhl, T. F. Bartelli, E. Batdelger, F. Baudon, K. Becher, C. Bello, M. Benchouaia, H. Benisty, A.-S. Benoiston, J. Benson, D. Benítez, J. Bernardes, D. Bertrand, S. Beurmann, T. Bitard-Feildel, L. Bittner, C. Black, G. Blanc, B. Blyther, T. Bode, J. Boeri, B. Boldgiv, K. Bolzli, A. Bordigoni, C. Borrelli, S. Bouchard, J.-P. Bouly, A. Boyd, G. P. Branco, A. Breschi, B. Brindefalk, C. Brion, A. Briones, P. Buczansla, C. M. Burke, A. Burrell, A. Butova, I. Buttar, J. Bynoe, S. Bönigk, K. O. Bøifot, H. Caballero, X. W. Cai, D. Calderon, A. Cantillo, M. Carbajo, A. Carbone, A. Cardenas, K. Carrillo, L. Casalot, S. Castro, A. V. Castro, A. Castro, A. V. B. Castro, S. Cawthorne, J. Cedillo, S. Chaker, J. Chalangal, A. Chan, A. I. Chasapi, S. Chatziefthimiou, S. R. Chaudhuri, A. K. Chavan, F. Chavez, G. Chem, X. Chen, M. Chen, J.-W. Chen, A. Chernomoretz, A. Chettouh, D. Cheung, D. Chicas, S. Chiu, H. Choudhry, C. Chrispin, K. Ciaramella, E. Cifuentes, J. Cohen, D. A. Coil, S. Collin, C. Conger, R. Conte, F. Corsi, C. N. Cossio, A. F. Costa, D. Cuebas, B. D'Alessandro, K. E. Dahlhausen, A. E. Darling, P. Das, L. B. Davenport, L. David, N. R. Davidson, G. Dayama, S. Delmas, C. K. Deng, C. Dequeker, A. Desert, M. Devi, F. S. Dezem, C. N. Dias, T. R. Donahoe, S. Dorado, L. Dorsey, V. Dotsenko, S. Du, A. Dutan, N. Eady, J. A. Eisen, M. Elaskandrany, L. Epping, J. P. Escalera-

Antezana, C. L. Ettinger, I. Faiz, L. Fan, N. Farhat, E. Faure, F. Fauzi, C. Feigin, S. Felice, L. P. Ferreira, G. Figueroa, A. Fleiss, D. Flores, J. L. Velasco Flores, M. A. S. Fonseca, J. Foox, J. C. Forero, A. Francis, K. French, P. Fresia, J. Friedman, J. J. Fuentes, J. Galipon, M. Garcia, L. Garcia, C. García, A. Geiger, S. M. Gerner, S. L. Ghose, D. P. Giang, M. Giménez, D. Giovannelli, D. Githae, S. Gkotzis, L. Godoy, S. Goldman, G. H. Gonnet, J. Gonzalez, A. Gonzalez, C. Gonzalez-Poblete, A. Gray, T. Gregory, C. Greselle, S. Guasco, J. Guerra, N. Gurianova, W. Haehr, S. Halary, F. Hartkopf, J. J. A. Hastings, A. Hawkins-Zafarnia, N. H. Hazrin-Chong, E. Helfrich, E. Hell, T. Henry, S. Hernandez, P. L. Hernandez, D. Hess-Homeier, L. E. Hittle, N. X. Hoan, A. Holik, C. Homma, I. Hoxie, M. Huber, E. Humphries, S. Hyland, A. Hässig, R. Häusler, N. Hüsser, R. A. Petit, B. Iderzorig, M. Igarashi, S. B. Iqbal, S. Ishikawa, S. Ishizuka, S. Islam, R. Islam, K. Ito, S. Ito, T. Ito, T. Ivankovic, T. Iwashiro, S. Jackson, J. Jacobs, M. James, M. Jaubert, M.-L. Jerier, E. Jiminez, A. Jinfessa, Y. De Jong, H. W. Joo, G. Jospin, T. Kajita, A. S. Ahmad Kassim, N. Kato, A. Kaur, I. Kaur, F. de Souza Gomes Kehdy, V. S. Khadka, S. Khan, M. Khavari, M. Ki, G. Kim, H. J. Kim, S. Kim, R. J. King, K. Knights, G. KoLoMonaco, E. Koag, N. Kobko-Litskevitch, M. Korshevniuk, M. Kozhar, J. Krebs, N. Kubota, A. Kuklin, S. S. Kumar, R. Kwong, L. Kwong, I. Lafontaine, J. Lago, T. Y. Lai, E. Laine, M. Laiola, O. Lakhneko, I. Lamba, G. de Lamotte, R. Lannes, E. De Lazzari, M. Leahy, H. Lee, Y. Lee, L. Lee, V. Lemaire, E. Leong, M. H. Y. Leung, D. Lewandowska, C. Li, W. Liang, M. Lin, P. Lisboa, A. Litskevitch, E. M. Liu, T. Liu, M. A. Livia, Y. H. Lo, S. Losim, M. Loubens,

J. Lu, O. Lykhenko, S. Lysakova, S. Mahmoud, S. A. Majid, N. Makogon, D. Maldonado, K. Mallari, T. M. Malta, M. Mamun, D. Manoir, G. Marchandon, N. Marciniak, S. Marinovic, B. Marques, N. Mathews, Y. Matsuzaki, V. Matthys, M. May, E. McComb, A. Meagher, A. Melamed, W. Menary, K. N. Mendez, A. Mendez, I. M. Mendy, I. Meng, A. Menon, M. Menor, R. Meoded, N. Merino, C. Meydan, K. Miah, M. Mignotte, T. Miketic, W. Miranda, A. Mitsios, R. Miura, K. Miyake, M. D. Moccia, N. Mohan, M. Mohsin, K. Moitra, M. Moldes, L. Molina, J. Molinet, O.-E. Molomjamts, E. Moniruzzaman, S. Moon, I. de Oliveira Moraes, M. Moreno, M. S. Mosella, J. W. Moser, C. Mozsary, A. L. Muehlbauer, O. Muner, M. Munia, N. Munim, M. Muscat, T. Mustac, C. Muñoz, F. Nadalin, A. Naeem, D. Nagy-Szakal, M. Nakagawa, A. Narce, M. Nasu, I. G. Navarrete, H. Naveed, B. Nazario, N. R. Nedunuri, T. Neff, A. Nesimi, W. C. Ng, S. Ng, G. Nguyen, E. Ngwa, A. Nicolas, P. Nicolas, A. Nika, H. Noorzi, A. Nosrati, H. Noushmehr, D. N. Nunes, K. O'Brien, N. B. O'Hara, G. Oken, R. A. Olawoyin, J. Q. Oliete, K. Olmeda, T. Oluwadare, I. A. Oluwadare, N. Ordioni, J. Orpilla, J. Orrego, M. Ortega, P. Osma, I. O. Osuolale, O. M. Osuolale, M. Ota, F. Oteri, Y. Oto, R. Ounit, C. A. Ouzounis, S. Pakrashi, R. Paras, C. Pardo-Este, Y.-J. Park, P. Pastuszek, S. Patel, J. Pathmanathan, A. Patrignani, M. Perez, A. Peros, S. Persaud, A. Peters, A. Phillips, L. Pineda, M. P. Pizzi, A. Plaku, A. Plaku, B. Pompa-Hogan, M. G. Portilla, L. Posada, M. Priestman, B. Prithiviraj, S. Priya, P. Pugdeethosal, C. E. Pugh, B. Pulatov, A. Pupiec, K. Pyrshev, T. Qing, S. Rahiel, S. Rahmatulloev, K. Rajendran, A. Ramcharan, A. Ramirez-Rojas, S. Rana, P. Ratnanandan, T. D.

Read, H. Rehrauer, R. Richer, A. Rivera, M. Rivera, A. Robertiello, C. Robinson, P. Rodríguez, N. A. Rojas, P. Roldán, A. Rosario, S. Roth, M. Ruiz, S. E. Boja Ruiz, K. Russell, M. Rybak, T. S. Sabedot, M. Sabina, I. Saito, Y. Saito, G. A. Malca Salas, C. Salazar, K. M. San, J. Sanchez, K. Sanchir, R. Sankar, P. T. de Souza Santos, Z. Saravi, K. Sasaki, Y. Sato, M. Sato, S. Sato, R. Sato, K. Sato, N. Sayara, S. Schaaf, O. Schacher, A.-L. M. Schinke, R. Schlapbach, C. Schori, J. R. Schriml, F. Segato, F. Sepulveda, M. S. Serpa, P. F. De Sessions, J. C. Severyn, H. Shaaban, M. Shakil, S. Shalaby, A. Shari, H. Shim, H. Shirahata, Y. Shiwa, R. Siam, O. Da Silva, J. M. Silva, G. Simon, S. K. Singh, K. Sluzek, R. Smith, E. So, N. Andreu Somavilla, Y. Sonohara, N. Rufino de Sousa, C. Souza, J. Sperry, N. Sprinsky, S. G. Stark, A. La Storia, K. Suganuma, H. Suliman, J. Sullivan, A. A. M. Supie, C. Suzuki, S. Takagi, F. Takahara, N. Takahashi, K. Takahashi, T. Takeda, I. K. Takenaka, S. Tanaka, A. Tang, Y. Man Tang, E. Tarcitano, A. Tassinari, M. Taye, A. Terrero, E. Thambiraja, A. Thiébaut, S. Thomas, A. M. Thomas, Y. Togashi, T. Togashi, A. Tomaselli, M. Tomita, I. Tomita, X. Tong, O. Toth, N. C. Toussaint, J. M. Tran, C. Truong, S. I. Tsonev, K. Tsuda, T. Tsurumaki, M. Tuz, Y. Tymoshenko, C. Urgiles, M. Usui, S. Vacant, B. Valentine, L. E. Vann, F. Velter, V. Ventorino, P. Vera-Wolf, R. Vicedomini, M. A. Suarez-Villamil, S. Vincent, R. Vivancos-Koopman, A. Wan, C. Wang, T. Warashina, A. Watanabe, S. Weekes, J. Werner, D. Westfall, L. H. Wieler, M. Williams, S. A. Wolf, B. Wong, Y. L. Wong, T. Wong, R. Wright, T. Wunderlin, R. Yamanaka, J. Yang, H. Yano, G. C. Yeh, O. Yemets, T. Yeskova, S. Yoshikawa, L. Zafar, Y. Zhang,

S. Zhang, A. Zhang, Y. Zheng, S. Zubenko, A global metagenomic map of urban microbiomes and antimicrobial resistance. *Cell*. **184**, 3376-3393.e17 (2021).

27. Z. Zhang, Q. Zhang, T. Wang, N. Xu, T. Lu, W. Hong, J. Penuelas, M. Gillings, M. Wang, W. Gao, H. Qian, Assessment of global health risk of antibiotic resistance genes. *Nat. Commun.* **13**, 1553 (2022).

28. F. Maguire, B. Jia, K. L. Gray, W. Y. V. Lau, R. G. Beiko, F. S. Brinkman, Metagenome-assembled genome binning methods with short reads disproportionately fail for plasmids and genomic islands. *Microb. Genomics*. **6** (2020).

29. Y. Che, Y. Yang, X. Xu, K. Břinda, M. F. Polz, W. P. Hanage, T. Zhang, Reply to Partridge et al.: Complementary bioinformatics and experimental approaches to investigate the transfer of AMR genes. *Proc. Natl. Acad. Sci.* **118**, e2108995118 (2021).

30. L. P. Shaw, K. K. Chau, J. Kavanagh, M. AbuOun, E. Stubberfield, H. S. Gweon, L. Barker, G. Rodger, M. J. Bowes, A. T. M. Hubbard, H. Pickford, J. Swann, D. Gilson, R. P. Smith, S. J. Hoosdally, R. Sebra, H. Brett, T. E. A. Peto, M. J. Bailey, D. W. Crook, D. S. Read, M. F. Anjum, A. S. Walker, N. Stoesser, null null, Niche and local geography shape the pangenome of wastewater- and livestock-associated Enterobacteriaceae. *Sci. Adv.* **7**, eabe3868.

31. J. Bethke, A. Davidovich, L. Cheng, A. Lopatkin, W. Song, J. Thaden, V. Fowler, M. Xiao, L. You, Environmental and genetic determinants of plasmid mobility in pathogenic Escherichia coli. *Sci. Adv.* **6**, eaax3173 (2020).

722 32. Arnold Markus F. F., Shabab Mohammed, Penterman Jon, Boehme Kevin L.,
723 Griffitts Joel S., Walker Graham C., Genome-wide sensitivity analysis of the
724 microsymbiont Sinorhizobium meliloti to symbiotically important, defensin-like
725 host peptides. *mBio*. **8**, e01060-17 (2017).
726 33. N. J. Tobias, H. Wolff, B. Djahanschiri, F. Grundmann, M. Kronenwerth, Y.-M.
727 Shi, S. Simonyi, P. Grün, D. Shapiro-Ilan, S. J. Pidot, T. P. Stinear, I. Ebersberger,
728 H. B. Bode, Natural product diversity associated with the nematode symbionts
729 Photorhabdus and Xenorhabdus. *Nat. Microbiol.* **2**, 1676–1685 (2017).
730 34. M. Magana, M. Pushpanathan, A. L. Santos, L. Leanse, M. Fernandez, A.
731 Ioannidis, M. A. Giulianotti, Y. Apidianakis, S. Bradfute, A. L. Ferguson, A.
732 Cherkasov, M. N. Seleem, C. Pinilla, C. de la Fuente-Nunez, T. Lazaridis, T. Dai,
733 R. A. Houghten, R. E. W. Hancock, G. P. Tegos, The value of antimicrobial
734 peptides in the age of resistance. *Lancet Infect. Dis.* **20**, e216–e230 (2020).
735 35. B. Vishnepolsky, M. Grigolava, G. Managadze, A. Gabrielian, A. Rosenthal, D.
736 E. Hurt, M. Tartakovsky, M. Pirtskhalava, Comparative analysis of machine
737 learning algorithms on the microbial strain-specific AMP prediction. *Brief.*
738 *Bioinform.* **23**, bbac233 (2022).
739 36. Yu Guozhi, Baeder Desiree Y., Regoes Roland R., Rolff Jens, Combination effects
740 of antimicrobial peptides. *Antimicrob. Agents Chemother.* **60**, 1717–1724 (2016).
741 37. M. R. Yeaman, N. Y. Yount, Mechanisms of antimicrobial peptide action and
742 resistance. *Pharmacol. Rev.* **55**, 27 (2003).
743 38. G. Yu, D. Y. Baeder, R. R. Regoes, J. Rolff, Predicting drug resistance evolution:

744 insights from antimicrobial peptides and antibiotics. *Proc. R. Soc. B Biol. Sci.* **285**,
745 20172687 (2018).
746 39. A. Rodríguez-Rojas, O. Makarova, J. Rolff, Antimicrobials, stress and
747 mutagenesis. *PLoS Pathog.* **10**, e1004445 (2014).
748 40. G. E. Fantner, R. J. Barbero, D. S. Gray, A. M. Belcher, Kinetics of antimicrobial
749 peptide activity measured on individual bacterial cells using high-speed atomic
750 force microscopy. *Nat. Nanotechnol.* **5**, 280–285 (2010).
751 41. N. Jochumsen, R. L. Marvig, S. Damkiær, R. L. Jensen, W. Paulander, S. Molin,
752 L. Jelsbak, A. Folkesson, The evolution of antimicrobial peptide resistance in
753 Pseudomonas aeruginosa is shaped by strong epistatic interactions. *Nat. Commun.*
754 **7**, 13002 (2016).
755 42. R. Jain, M. C. Rivera, J. A. Lake, Horizontal gene transfer among genomes: The
756 complexity hypothesis. *Proc. Natl. Acad. Sci.* **96**, 3801–3806 (1999).
757 43. O. Cohen, U. Gophna, T. Pupko, The complexity hypothesis revisited:
758 connectivity rather than function constitutes a barrier to horizontal gene transfer.
759 *Mol. Biol. Evol.* **28**, 1481–1489 (2011).
760 44. R. Wang, L. van Dorp, L. P. Shaw, P. Bradley, Q. Wang, X. Wang, L. Jin, Q.
761 Zhang, Y. Liu, A. Rieux, T. Dorai-Schneiders, L. A. Weinert, Z. Iqbal, X. Didelot,
762 H. Wang, F. Balloux, The global distribution and spread of the mobilized colistin
763 resistance gene mcr-1. *Nat. Commun.* **9**, 1179 (2018).
764 45. B. Kintses, O. Méhi, E. Ari, M. Számel, Á. Györkei, P. K. Jangir, I. Nagy, F. Pál,
765 G. Fekete, R. Tengölics, Á. Nyerges, I. Likó, A. Bálint, T. Molnár, B. Bálint, B.

766 M. Vásárhelyi, M. Bustamante, B. Papp, C. Pál, Phylogenetic barriers to
767 horizontal transfer of antimicrobial peptide resistance genes in the human gut
768 microbiota. *Nat. Microbiol.* **4**, 447–458 (2019).
769 46. A. G. McArthur, N. Waglechner, F. Nizam, A. Yan, M. A. Azad, A. J. Baylay, K.
770 Bhullar, M. J. Canova, P. G. De, L. Ejim, L. Kalan, A. M. King, K. Koteva, M.
771 Morar, M. R. Mulvey, J. S. O'Brien, A. C. Pawlowski, L. J. V. Piddock, P.
772 Spanogiannopoulos, A. D. Sutherland, I. Tang, P. L. Taylor, M. Thaker, W. Wang,
773 M. Yan, T. Yu, G. D. Wright, The Comprehensive Antibiotic Resistance Database.
774 *Antimicrob. Agents Chemother.* **57**, 3348–3357 (2013).
775 47. D. Veltri, U. Kamath, A. Shehu, Deep learning improves antimicrobial peptide
776 recognition. *Bioinformatics*. **34**, 2740–2747 (2018).
777 48. F. H. Waghu, R. S. Barai, P. Gurung, S. Idicula-Thomas, CAMPR3: a database on
778 sequences, structures and signatures of antimicrobial peptides. *Nucleic Acids Res.*
779 **44**, D1094–D1097 (2016).
780 49. X. Zhao, H. Wu, H. Lu, G. Li, Q. Huang, LAMP: a database linking antimicrobial
781 peptides. *PloS One*. **8**, e66557 (2013).
782 50. B. Buchfink, K. Reuter, H.-G. Drost, Sensitive protein alignments at tree-of-life
783 scale using DIAMOND. *Nat. Methods*. **18**, 366–368 (2021).
784 51. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler
785 transform. *Bioinformatics*. **25**, 1754–1760 (2009).
786 52. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G.
787 Abecasis, R. Durbin, 1000 Genome Project Data Processing Subgroup, The

Sequence Alignment/Map format and SAMtools. *Bioinformatics*. **25**, 2078–2079 (2009).

53. HARTMANN Science Centre, Relevant Pathogens from A–Z https://www.bode-science-center.com/center/relevant-pathogens-from-a-z.html (2021).

54. P. Siguier, J. Perochon, L. Lestrade, J. Mahillon, M. Chandler, ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* **34**, D32–D36 (2006).

55. L. Fu, B. Niu, Z. Zhu, S. Wu, W. Li, CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. **28**, 3150–3152 (2012).

56. Y. Che, Y. Xia, L. Liu, A.-D. Li, Y. Yang, T. Zhang, Mobile antibiotic resistome in wastewater treatment plants revealed by Nanopore metagenomic sequencing. *Microbiome*. **7**, 44 (2019).

57. SEDAC, Population Estimation Service https://sedac.ciesin.columbia.edu/mapping/popest/pes-v3/ (2021).

Acknowledgements

We appreciate Nayfach et al. for their work on the genomic catalog of Earth's microbiomes and its publication as a public resource. This study was financially supported by the National Key Research and Development Program of China (2017YFD0200503), the Key Research and Development Program of Zhejiang (2022C02029 and 2022C02046), and the National Natural Science Foundation of China (21777144, 21777145, and 21976161). JP acknowledge financial support from the Spanish Government grants PID2019-110521GB-I00 and TED2021-132627B-I00, the Fundación Ramón Areces grant CIVP20A6621, and the Catalan Government grant SGR 2017-1005.

Contributions

Bingfeng Chen designed the study with guidance from Haifeng Qian. Bingfeng Chen wrote the first draft of the manuscript and Haifeng Qian, Josep Penuelas, Zhengwei Fu, and Michael Gillings contributed substantially to revision. Tingzhang Wang and Wenjie Hong contributed to the all-functional annotation of metagenome-assemble genomes. Bingfeng Chen and Zhenyan Zhang, and Tao Lu contributed to the collection and disposal of all data information. Bingfeng Chen, Qi Zhang, and Nuohan Xu performed all the metagenomic analysis. Bingfeng Chen and Zhenyan Zhang performed the visualization of all data and the artistic design of all figures. Haifeng Qian, Zhengwei Fu, Tao Lu, and Josep Penuelas acquired funding for this project.

Ethics declarations

Competing Interest

The authors declare that they have no competing interests.