# Deterministic effects of pH in shaping soil resistome revealed by metagenomic analysis

**Zishu Liu**
College of Environmental and Resource Sciences, Zhejiang University, Hangzhou
**Yuxiang Zhao**
College of Environmental and Resource Sciences, Zhejiang University, Hangzhou
**Baofen Zhang**
Zhejiang Hangzhou Ecological Environment Monitoring Center, Hangzhou
**Jiaqi Wang**
College of Environmental and Resource Sciences, Zhejiang University, Hangzhou
**Lizhong Zhu**
College of Environmental and Resource Sciences, Zhejiang University, Hangzhou
**Baolan Hu** ( blhu@zju.edu.cn )
College of Environmental and Resource Sciences, Zhejiang University, Hangzhou

**Research Article**

**Keywords:** Antibiotic resistance, Soil resistome, pH, Proton activity, Multidrug efflux pump, Background ARG

**Posted Date:** May 31st, 2022

**DOI:** https://doi.org/10.21203/rs.3.rs-1673295/v1

## Abstract

# Background

Research in environmental resistome has been greatly facilitated by the use of the comprehensive database of antibiotic resistance genes (ARGs). Through this approach, the high prevalence of multidrug efflux pump genes was recently found in various ecosystems and has raised concerns. Since most of the efflux pumps were driven by proton-motive-force (PMF), studying whether soil pH, i.e. proton activity, is a determinant in selecting multidrug efflux pump genes and thus shaping soil resistome is of great interest.

# Results

Based on the distribution of soil pH in China, we collected soil samples from forest, cropland and grassland in four typical climatic areas with pH values ranging from 4.37–9.69. ARG profiles of 36 soil metagenomes and 41 metagenome-assembled genomes were obtained by using the SARG database and the ARG-OAPs pipeline. As the result, 264 ARG subtypes associated with 20 classes of antibiotics were identified, while multidrug resistance (69.72% of whole resistome abundances) and efflux pump (82.49%) were the dominant ARG category and resistance mechanism. Independent of sampling location, the abundance of multidrug resistance gene was decreased linearly with increasing pH over the given range, which contributed to almost the entire variation in ARG abundances. In addition, soil pH also determined the diversity, i.e. evenness, of ARGs, and thus explained 75.2% of the resistome variation in the variance partitioning analysis. On this basis, we investigated the relationship between ARGs and bacterial phyla using consistency and correlation analysis, as well as tested their assembly mechanisms using a neutral model, and found ARGs were naturally selected despite the neutral assembly of the bacterial community.

# Conclusions

Our results revealed the deterministic effect of pH on shaping soil resistome, and infer that this is related to the selection of PMF-driven multidrug efflux pumps under high proton activity. Such natural changes in background ARGs deserve further attention.

## Background

The spread of antimicrobials resistance poses a huge threat to global public health, that cannot be understated. In 2019, it is estimated that 4.95 million deaths are associated with bacterial antibiotic resistance, which is an approximately sevenfold increase over the past decade. To tackle this global health challenge, it is essential to grasp the microbial interconnections among the environment and humans (i.e. the One Health framework [1–3]), due to the high ability of bacteria and genes in crossing geographical and species boundaries.

Undoubtedly, the rapid evolutionary expansion and spread of antibiotics resistance genes (ARGs) in past decades are related to the utilization and excretion of antibiotics and other antimicrobial agents. As its induced selective pressure has promoted the mobilization, horizontal transfer and mutation-based generation of a large range of ARGs in the scenario like hospital, farmland and sewage treatment plant [4]. But, calm down a little, such selective pressure is not common in the natural environment where concentrations of antibiotics are often lower than their estimated (or known) minimum selective concentrations in the laboratory [5, 6]. Hence the important driver of the evolution and transmission of ARGs in the environment is still supposed to be the interactions between microbes and their living environments, and have never ceased since ancient times [5, 7].

The soil environment is regarded as the major reservoir of ARGs [8, 9], where numerous microbes and their metabolites meet and interact with each other (e.g. biotic competition and antagonism [10, 11]) and thus induce the generation of resistance genes. The abundance of antibiotic producers, such as microbes in Actinobacteria (bacterial phylum) and Ascomycota (fungal phylum) was reported to be significantly correlated with the abundance and diversity of ARGs in soil [12], as well as many clinically relevant ARGs were also proved (or suspected) to be originated from the soil resistome [13, 14]. Therefore, profiling soil resistome and its formation mechanisms are needed, as better characterizing the background type and amount of ARGs is vital to assess the actual impact of future human activities and thus facilitate the prevention of resistance spreading [15].

In a recent study of the global topsoil microbiome, Bahram et al. used the abundance of ARGs as the evidence to infer strong inter-kingdom antagonism, which in turn suggests the great potential of metagenomic analysis in studying environmental resistome [12]. However, due to the different focus, the relationship between ARGs (i.e., corresponding antibiotic classes and resistance mechanisms) and environmental variables has not been explored in detail in their study. Indeed, it is of great interest to categorize ARGs with hierarchical structure in the study of resistome, since different ARGs encode proteins conferring distinct resistance mechanisms (e.g. regulation, efflux pump, drug inactivation, etc.) in response to varied antibiotics with specific inhibitory effects and molecular targets [16, 17]. To this end, the metagenomic analysis stands out and provides hierarchical profiles of environmental resistome [18–22] through the use of comprehensive databases and newly developed bioinformatic analysis tools [23–25]. As the consequence, new insights into the environmental resistome have recently emerged, among which the high prevalence of multidrug resistance genes in soil ecosystems has raised concerns [15, 20, 26–29].

The vast majority of multidrug resistance genes are associated with multidrug efflux pumps, which are integral membrane transport proteins that function to extrude antibiotics from bacterial cells [30]. They are named “Multidrug” because typically recognize not only one class of antibiotics but a wide range of chemically dissimilar compounds, and thus a single pump may provide resistance to a broad range of antimicrobial classes (well-reviewed in [30, 31]). For instance, the efflux pump complex AcrAB-TolC confers clinically significant resistance to more than six categories of antibiotics including β-lactams, quinolones, macrolides, and tetracycline. Thus, a simple combination of efflux pumps and low permeable

cell membrane would already induce high intrinsic resistance for many antibiotics [32]. Like other resistance mechanisms, multidrug efflux pumps have evolved in bacteria long before the widespread use of antibiotics, and thus many bacterial efflux pumps are versatile and have native functions that are unrelated to antibiotic resistance. For example, address plant-derived toxins, aromatic hydrocarbons and heavy metals, as well as participate in the primary metabolite efflux and acid stress tolerance [30]. Moreover, the operation of the efflux pump is not complicated as far as the driving force is concerned. Because most of the efflux pumps (e.g. RND: resistance-nodulation-cell division superfamily and MFS: major facilitator superfamily) require only the proton-motive-force (PMF) that is composed of the pH gradient and electrochemical potential inside and outside the cytoplasmic membrane [33]. Whether all these physiological functions as a whole benefit the bacterial host and thus lead to a high prevalence of multidrug resistance genes; and how they are influenced by edaphic factors (e.g. pH, temperature, C/N ratio, etc.) needs further investigation.

Mineral matrices, pedogenesis process, climatic conditions, and land-use practices can significantly alter the physicochemical as well as biological properties of the soil. Therefore, to investigate the effects of edaphic factors and microbial contributions on the variance of soil resistome, we used soils harvested from multiple sources i.e. forest, grassland and cropland in temperate monsoon, temperate continental, subtropical monsoon and plateau mountain climatic areas. Metagenomes and metagenome-assembled genomes (MAGs) were profiled by using the ARGs-OAP pipeline and referring SARG database [24]. In particularly to address the following objectives: (i) what are the background types and amounts of ARGs in natural soils; (ii) whether soil pH has a deterministic effect on the selection of multidrug efflux pump genes; and if yes (iii) what are the corresponding patterns of microbial community composition behind. The results of this study will further improve our knowledge of soil resistome and provide a novel perspective for understanding the mechanisms underlying the formation of geographical differences in ARGs.

## Results

# ARG profiles of natural soils

In this study, a total of 36 soil metagenomes were studied with well-characterized metadata (details see Additional file 1: S1 and Additional file 2: Table S1). All in all, 264 ARG subtypes potentially associated with 20 classes of antibiotics were identified from $4.46 \times 10^{11}$ bases of cleaned short reads (statistics for the raw data, Clean Data, ARG types and subtypes data were listed in Additional file 2: Table S2-S9).

As expected, the multidrug resistance gene was detected in all samples and was the most abundant ARG type with an average relative abundance of 69.72% (Fig. 1A). It was followed by ARGs for macrolides-lincosamides-streptogramines (MLS, 9.41%), vancomycin (6.46%), bacitracin (4.11%) and fosmidomycin (3.49%). These five ARG types accounted for more than 90% of all ARG abundances and were detected with frequencies of 97.2–100% in 36 metagenomes (Additional file 2: Tab. S5). Unclassified ARGs such as cAMP-regulatory proteins and cob(I)alamin adenosyltransferase gene mutants were identified with an

average relative abundance of 2.13%, they were engaged in the regulation of bacterial energy metabolism and the single carbon cycle, rather than conferring resistance to a particular class of antibiotics. Moreover, ARGs for antibiotics tetracycline, quinolone, β-lactam, rifamycin, polymyxin and aminoglycoside were found to be widely distributed in the soil environment with detection frequencies greater than 91.7%, although their average relative abundances were all below 1%. Overall, a broad spectrum of antibiotic resistance is naturally harbored by soil microbes, while the abundance of ARGs varied dramatically among different classes of antibiotics.

Concerning the resistance mechanism that ARGs conferred, most ARG subtypes were belonging to the drug inactive mechanism, i.e. 114 subtypes including β-lactamase, acetyltransferase, erythromycin esterase, etc.; and efflux pump, i.e. 90 subtypes representing transporter, fusion protein, and porins. Although the efflux pump did not have the greatest number in subtypes, its mean relative abundance was 82.49% the highest value among all resistance mechanisms (Fig. 1B), and far exceeding the 2.56% for drug inactive. Furthermore, 30 subtypes and 18 subtypes of ARGs were found to be related to the target-alteration enzyme and the target-protection protein, which contributed 4.80% and 0.76% of all ARGs, respectively. In addition, 12 genes for the regulating protein accounted for another 9.39% of all ARGs (e.g. *vanSR* for vancomycin resistance; *ompR*, *marR* and *arlR* for multidrug).

# Essential role of pH in shaping soil resistome

Hierarchical clustering analysis was performed based on the dissimilarity of the resistome composition between samples (Fig. 2A). Based on the prediction of the optimal number of clusters (Additional file 1: S2), five groups of samples, i.e., Group1-5, were determined, which were independent of the sampling sites. Afterward, a clear upward trend in pH was found by the summarization of pH values per group (Fig. 2B), as the pH value increased from 5.17 ± 0.59 for Group1 to 5.93 ± 0.49 for Group2, 7.51 ± 0.44 for Group3, 7.89 ± 0.26 for Group4 and 9.04 ± 0.66 for Group5. Meanwhile, ARG abundance per antibiotic per sample was visualized in Figure S2 with the same grouping rule (Additional file 1: S3). Different from the rise of pH, it was a downward trend in the total abundance of ARGs per sample from Group1to Group4 and was accompanied by a decrease in multidrug abundance (Additional file 1: S3). But the mean total abundance of Group5 was slightly greater than it of Group4 because of the generally high abundance of vancomycin resistance genes in Group5 (i.e. 5.88 times high than the mean of all samples, Figure S3 A).

Resistome dissimilarity was analyzed with the abundance data of ARG types and visualized by using the NMDS plot (Fig. 2C, Bray Curtis index, stress = 0.04), where each point represented a soil resistome and was colored according to the determined groupings. Points well clustered in colors means the inter-group variation of resistomes was distinct, and the significance of such variation could be supported by the pairwise Adonis analysis (also known as PERMANOVA: permutational multivariate analysis of variance, Additional File1: S4). To further investigate which parameter was affecting soil resistome, contributions of environmental factors (Additional file 2: Table S1) and phylum compositions (Additional file 2: Table S10) on driving resistome changes were examined by correlation analysis and variance partitioning analysis (VPA) successively. And as a consequence, four influential factors were screened out (Additional file: S5). They were soil pH, sulfur content, and relative abundances of *Actinobacteria* and *Acidobacteria*

(fitted against resistome changes and plotted in Fig. 2C). These four factors together explained 83.8% of the variation in resistome compositions, while a partition of 75.2% was relevant to pH, it thus indicated pH to be the most explanatory factor for the changes in soil resistome (Additional file 1: S5).

To this end, regression analyses of pH were performed against both the abundance and diversity of ARGs per sample, results were shown in Fig. 2D for the total abundance of ARGs; Fig. 2E for the abundance of multidrug resistance; Fig. 2F for ARG richness as the number of ARG types per sample; and Fig. 2G for ARG evenness with Shannon index (H) and Pielou index (J). Although the regression for richness was not significant (*p* value = 0.764), the obtained coefficients of determination (adjusted R-squared, R2) were as high as 0.67, 0.80, 0.76, and 0.79 for the total abundance of ARGs, the abundance of multidrug, Shannon index and Pielou index, respectively (*p* value < 0.001, more details about significance testing were given in Additional file 1: S6). This means all examined abundance and diversity indicators, except richness, responded linearly to pH changes over a given range of 4.37–9.69.

# ARGs for inner membrane transferase predominated resistome in acidic soil

As described much of the ARGs abundance corresponded to the efflux mechanism (Fig. 1B), which encompassed most multidrug resistance genes (only except *ompF* and *omp36* two subtypes for porin) and in 34 samples accounted for more than half the abundance of all ARGs (in total 36 samples). It is thus reasonable to investigate the efflux pump genes at the subtype level.

A total of 89 detected ARG subtypes were associated with the efflux mechanism (summarized in Additional file 2: Table S12), which represented efflux pump complexes mainly from the RND superfamily (42 subtypes), the MFS superfamily (33 subtypes) and ABC superfamily (ATP-binding-cassette transporter, 6 subtypes). Among them, *mdtABC-tolC, mexEF-oprN* and *macAB-tolC* were found to be the most abundant gene components for efflux pumps (Fig. 3A) and were predominated by genes encoded inner membrane transferase in special (i.e. *mdtBC*, *mexF* and *macB*). According to the abundance accumulation curve (Fig. 3B), these three gene components had contributed 63% of all efflux genes, while the other seven of the listed top 10 gene components accounted for another 17.1%. In addition, unclassified efflux pump genes occupied 15.4% (5 subtypes without specific gene names in the database) and the others (56 subtypes) shared the last 3.5% proportion. Moreover, the abundance of the three most abundant gene components of the efflux pump was plotted against the defined groups (Fig. 3C). The *mdtBC*, *mexF* and *macB* genes, which encoded inner membrane permeases, were highly abundant in Group1 and their abundances declined from Group1 to Group5. Especially for *mdtBC*, from the RND family, they were hyper-dominant in the Gruop1 and Group 2 (i.e. acidic groups) but almost decreased to none in Group4 and Group5 (i.e. alkaline groups).

# Different patterns of the microbiome and resistome changes in natural soils

To reveal the relationship between microbiome and resistome, the consistency of compositional changes was examined at first (Fig. 4). The dissimilarity of microbial community was analyzed at the phylum level with the same workflow as for resistome, as the result (Fig. 4A), Group2 overlapped with the region of Group1 and Group3 on the plot. Further statistical analysis revealed that the coefficient of difference (R2) between Group2 and Group3 was as small as 0.16 ($p < 0.05$, Adonis analysis, Additional file 1: S4), while the difference between Group2 and Group1 was even not significant (R2 = 0.11, $p = 0.157$). This indistinguishability of samples between groups was different from the observation for resistome at ARG type level (Fig. 2C) and ARG subtype level (Fig. 4B). On this basis, the consistency of microbiome and resistance group changes was quantified by Procrustes analysis (Fig. 4A the right plot). It showed that some of the originally dispersed empty points (i.e. samples visualized in the transformed ordination defined upon resistome composition) tended to converge toward the center of the coordinates when they pointed to solid points (i.e. samples visualized in the ortho-ordination defined upon phylum composition), resulted in the overlap of different grouping areas as mentioned above. Meanwhile, the calculated consistency coefficient $M^2$ (the smaller value the higher consistency) between changes in phylum and ARG types was equal to 0.453 ($p < 0.001$), and such consistency was relatively weaker than the comparison for ARG types and ARG subtypes ($M^2 = 0.233$, $p < 0.001$). This assumption could also be supported by testing community assembly with a neutral model (Additional file 1: S7), whereas microbial communities investigated in this study were neutrally assembled at the phylum level (fitted R2 = 0.727), but the resistome was not (R2=-0.142, not fitted to the model).

# Relationships among environmental factors, microbes and ARGs

To further understand the differences in the patterns of microbiome and resistome changes, the contribution of environmental factors to their relationship should be considered. For this purpose, correlation analysis was performed by using data from 20 ARG types (corresponding to antibiotic class), 17 environmental factors (abiotic parameters), and all 61 identified phyla. Only the significant correlations (i.e. |SparCC-correlation value| > 0.6, $p < 0.05$) were used to determine modules (i.e. cluster of interconnected nodes) and construct the correlation network.

As the result, a network with two modules (i.e. Module #1 and #2) was displayed in Fig. 5, which associated five ARG types, 17 environmental factors and 15 phyla together. Module #1 was named the acid module because it was only containing pH-negative correlated ARGs, i.e. multidrug and fosimidomycin, and phyla of *Acidobacteria*, *Verrucomicrobia*, *Dependentiae* (formerly TM6), and WPS-2 (recently proposed as *Eremiobacterota* [34]). In addition, the acid module contains 12 environmental factors, including pH, C/N ratio, contents of carbon, nitrogen and sulfur, etc., with a total of 100 edges and an average connectness of 4.16 (i.e. total number of edges to the total number of nodes). In contrast, Module #2 named the alkali module that contained *Actinobacteria* the unique pH-positive correlated phylum, which linked with ARGs for tetracycline and vancomycin. However, there were no ARGs directly linked with pH. The alkali module was containing a total of 11 nodes (two ARGs, three environmental

factors and six phyla) and 39 edges, with an average connectness of 3.55. Thus, the acid module was more nodal and connected than the alkali module.

On this basis, the significant correlations were counted per each ARG type (Table 1), to quantify the contribution of environmental factors and microbial composition in selecting ARGs. Overall, 42 significant correlations were determined with ARGs, among which 18 were related to environmental factors, as well as 14 and 10 were positively and negatively correlated with bacterial phyla, respectively. Correspondingly, for each type of ARG, except vancomycin, the counts of abiotic correlations with environmental factors were generally higher than the positive correlations with bacterial phyla. From this perspective, the change of resistome in the soil was more closely related to the environmental factors than the phylum composition. To underpin this conclusion, a structural equation model was tested (Additional file1: S9), where a significant relationship between ARGs and environmental factors was weighted as 0.679, higher than the 0.173 for it between ARGs and bacterial phyla.

Table 1
Counts of significant correlations (|SparCC-correlation| > 0.6, $p < 0.05$) per ARGs. Total significant correlations were counted per each ARG type (counts), correlations between environmental factors, positive and negative correlations between bacterial phyla were also counted and recorded as both counts and percentages of all correlations per ARG types; MLS means the ARG type for Macrolide–Lincosamide–Streptogramin resistance.

| ARGs | All significant correlations | Correlations with environmental factor (counts\|%) | Positive correlations with phylum (counts \|%) | Negative correlations with phylum (counts \|%) |
|---|---|---|---|---|
| multidrug | 17 | 9\|52.9% | 6\|35.3% | 2\|11.8% |
| MLS | 6 | 4\|66.7% | 2\|33.3% | 0 |
| fosmidomycin | 9 | 3\|37.5% | 2\|25.0% | 3\|37.5% |
| vancomycin | 9 | 1\|11.1% | 3\|33.3% | 5\|55.6% |
| tetracycline | 2 | 1\|50.0% | 1\|50.0% | 0 |

# The ARG profiles of MAGs show the other side of the coin

As a complement to the analysis of shotgun data, the ARGs were also profiled at the genome level. In this study, a total of 41 high-quality MAGs were generated from the de novo assembly per sample, and their ARG profiles were displayed in Fig. 6. These MAGs were phylogenetically classified into one archaeal phylum (i.e. *Thermoproteota*) and 10 bacterial phyla by using the Genome Taxonomy Database (GTDB). And 18 ARG types were detected with 194.0 ± 76.0 ARG-carrying ORFs per MAG (containing 252 ARG subtypes, Additional file 2: Table S11). In terms of the richness of ARG per MAG, the detected ARG types were generally consistent with those detected in the short reads. ARGs for bacitracin, β-lactam, fosmidomycin, MLS, multidrug, polymycin, sulfamide, tetracycline and vancomycin were both highly detected in MAGs and short reads, with detection frequencies ranging from 95.1–100%. However, low detected ARGs for kasugamycin, puromycin and carbomycin in shotgun data were not found in MAGs anymore.

On the other hand, in terms of the relative abundance of ARG per MAG (i.e. proportion of counts per ARG in sum counts of all ARGs per MAG), resistance genes for multidrug were not the most abundant ARG anymore (mean relative abundance = 11.96%) and, as an alternative, ARGs for vancomycin, MLS and bacitracin had greater abundance (mean relative abundance = 26.10%, 24.91% and 12.88%, respectively). In addition to this, the ARG profiles in MAGs still show a phylum-to-phylum variation at the subtypes level (Additional file 1: S10) although the high abundant genes (i.e. *macB*, *arnA*, *bcrA* and *vanRSD*) were shared by all bacterial MAGs.

# Discussion

In this study, to examine the soil resistome we profiled 36 metagenomes and 41 high-quality MAGs by referring to the SARG database along with a recommended pipeline [24]. The comprehensive information of ARGs was investigated with both the antibiotic classes and functional genes, and thus compositional characteristics of background ARG in natural soils and potential microbial hosts were revealed. ARG profiles generated from the shotgun data and MAGs were mutually supportive. While the short-read-based approach provided more sensitive detections of ARGs, de novo assembled genome analysis would better reflect the co-occurrence and relative abundance of potentially functional ARGs as well as the identity of their microbial host [15], although low abundant ARGs might be lost during the assembly process. It is indeed the case in our results that ARGs for nine classes of antibiotics were both frequently detected in MAGs and short reads (Fig. 6 and Addition file 1: Figure S3), meanwhile genes that were low detected in short reads have vanished from MAGs (i.e. genes for kasugamycin, puromycin and carbomycin). However, with regards to the relative abundance of ARGs, results of short reads and MAGs show some contradictions. In MAGs genes like *bcrA* (confer resistance to bacitracin), *macB* (extrude macrolide and peptide), and *vanSR* (regulon for vancomycin resistance) were more abundant than the short-read-popular multidrug resistance genes like *mdtBC* and *mexF*. (Addition file 2: Table S11). For environmental samples, such contradictions due to gene assembly seem to occur from time to time, as the mapped sequences of ARGs are only a small fraction of all reads in a sample. In a recent review, Coelho et al. summarized 13174 publicly available metagenomes and found MAGs generally capture a proportion less than 5% of all genes in a soil sample [35]. In addition, the ability of binning method to correctly recover ARGs on mobile genetic elements was systemically problemed by the variable copy number and sequence composition of the gene elements [36]. Therefore, to have a consistent perspective, we mainly discuss the results obtained from short reads in this study.

As we mentioned in the introduction, it is unsurprising that ARGs are ubiquitous in the soil [8], since many antibiotics are natural secondary metabolites of soil microbes and were produced as chemical arms in the competition for survival [37]. ARGs serve as a genetic blueprint for fortifications and is necessary for the survival of microbes. However, what are the physiological and ecological nature behind the predominance of multidrug resistance genes (or more broadly, efflux pump genes), in the soil resistome is still worth exploring. In this study genes for efflux pump mechanisms show dominancy in soil resistome in terms of either total abundance, average relative abundance or detection frequency (Fig. 2 and Additional file 2: Table S4-S9). The dominancy could reflect their higher fitness in natural soils.

As Larson et al. concluded the concentration of a single antibiotic in the environment is generally insufficient to cause a selective pressure, and even if the antibiotic is accidentally exposed due to human factors, it will degrade within a short period [5]. Therefore, the real situation is supposed to be the coexistence of multiple naturally secreted antimicrobials at low concentrations, where multidrug efflux pumps were reasonably selected by microbes due to their functional versatility [30]. For instance, MdtABC-TolC was the efflux pump with the highest average abundance detected in this study (Fig. 3), it is associated with the efflux of antibiotics such as β-lactams and novobiocin as well as with the exclusion of copper and zinc and the formation of biofilms [31]. Similarly, MexEF-OprN could pump out aromatic hydrocarbons and population-sensing signaling molecules; MacAB-TolC extrudes lipopolysaccharides and peptides as non-antibiotic resistant functions [30]. These functions are closely related to interfacial colonization and nutrient acquisition and are themselves a compensatory mechanism for the cost of resistance expression of microbes in the soil environment. Furthermore, consistent with our observation, previous studies also reported the prevalence of multidrug resistance genes in natural and anthropogenically affected soil [15, 26]. In a study of the Savannah River Site [26], a high abundance of multidrug resistance genes were found in all 24 acidic (i.e. pH in 3.98–4.38) and heavy metal polluted soils. If in that situation it can still be explained by heavy metal co-selection at contaminated sites, the prevalence of multidrug resistance genes in 26 rainforest and pasture soil samples could only be attributed to the natural selection of soil microbes [15].

Based on the prevalence of multidrug efflux pump genes, the deterministic shaping role of soil pH in the resistome was highlighted. In terms of the abundance of ARGs, the total abundance was found to be highest when the soil pH was the lowest (Figure S3 in Additional file1) and showed a linearly decreasing trend with the increase of pH (Fig. 2D). Moreover, all the changes in total ARG abundance seem to be caused by the changes in multidrug resistance genes (Fig. 2E and Additional file1: S6, similar slopes in regression). Especially for the genes related to inner membrane permease, i.e. *mdtBC*, they were hyper dominant in groupings of acidic soils (Fig. 3) but almost decreased to none in groupings with soil pH higher than neutral, reflecting a dependence on the low pH (i.e. strong proton activity). MdtB and MdtC have been proved to be the PMF driven multidrug transporter since both of them had five charged and polar amino acid residues which conserved with AcrB and were essential for the proton transportation [38]. Therefore, there is a biochemical basis for the selection of low soil pH for MdtBC. In addition, based on the calculation of the diversity index of ARGs, we found the evenness of ARGs linearly increased with pH, although the richness of ARGs is almost constant. This indicates a shift from the predominance of a few types of ARGs to the coexistence of multiple types, accompanied by an increase in soil pH.

In general, the pH shaping rule we found is consistent with the observation of Bahram et al. in a recent study where global topsoil samples (7560 subsamples in 189 sites from 12 ecosystems) were investigated and a negative correlation between ARGs abundance and soil pH was found [12]. However, the deterministic role of pH in shaping soil resistome is more evident in our study (see the VPA result in Additional file1: S5), which could due to the differences in pH of the observed soil. In particular, their pH values ranged from 2.5 to 7.5 and were mainly concentrated in acidic conditions, while our pH values are evenly distributed between 4.37 and 9.69 (Additional file1: S2). Bahram et al. demonstrated that higher

fungal biomass and stronger fungal-bacterial antagonism in acidic soils would lead to a higher abundance of ARGs, this should be a reliable nature. Fungi are well-known natural producers of many antimicrobials, not just natural β-lactam antibiotics, but also ergot alkaloids, fungal polyketides and kinds of antimicrobial peptides [39]. Therefore, in low pH soils where fungal antimicrobials are present in a large amount, PMF-driven multidrug efflux pumps (such as MdtBC and MexF in our case) would be carried by bacterial hosts to counteract multiple antimicrobials stresses is intuitively correct, due to their versatility and the easier access to protons. When soil pH rises it is another situation, where fungal antimicrobials are reduced and access to protons was relatively limited, thus the abundance of ARGs, especially for multidrug genes, gradually decreased. Then the abundance increasing in non-multidrug resistance genes will be foregrounded, such as the ARGs for vancomycin and tetracycline in this study (Figure S3 in Additional file1: S3).

To further explore the relationship between microbes and ARGs in soil, a strict correlation analysis was performed in this study, where indirect correlations were deduced from the network (Fig. 5). And we found ARGs and correlated bacterial phyla have separated into an acid module and an alkali module. In the alkali module, the unique pH-positive-correlated phylum *Actinobacteria* was also positively associated with the tetracycline and vancomycin genes, this is consistent with the “producer hypothesis” [13, 14]. That microbes in *the Actinobacteria* phylum (i.e. *Streptomyces* spp.) are natural producers of tetracycline and vancomycin, and thus are the origins of tetracycline and vancomycin resistance genes. While no direct correlation between pH and ARGs was indicated the changes in tetracycline and vancomycin resistance genes were primarily driven by the natural selection of the *Actinobacteria* phylum. A similar conclusion was also given by Qian et al., who demonstrated environmental factors select bacterial populations and hence enrich the ARGs they carry [15]. However, in the acid module of our network, environmental factors directly connected with ARGs and their counts were generally higher than the positive correlations between ARGs and bacterial phyla, thus a stronger linkage was found between ARGs and environmental factors. To further confirm, we used SEM to examine relationships among environmental factors, microbes, and ARGs (Additional file1: S9), consistent with the result of correlation analysis, the causality between ARGs and environmental factors was higher weighted than it between ARGs and bacterial phyla. Combining these results, we assumed that in acidic soils ARGs corresponding to high fitness functions may be directly selected by environmental factors rather than through a particular bacterial population, indeed the selected ARGs might be carried by more than one bacterial population. For instance, *mdtBC* and *mexF* genes were carried by more than half MAGs and have no specific taxonomic affiliation (Figure S10.1 in Additional file1: S10).

Until here a panorama of pH-driven changes in soil resistome becomes clear. Unlike the neutral assembly of bacterial phylum composition (Additional file1: S7), soil pH variation has a deterministic effect on shaping resistome. The specific manifestations are 1) pH decrease (until 4.5): the total abundance of ARGs increased to meet the putative stronger fungal-bacterial antagonism, multidrug efflux pumps are preferred due to their versatility and stronger fitness in low pH soils and thus were selected; and 2) pH increase (until 9.5): decreased strength of inter-kingdom competition, thus the lower total abundance of

ARGs, lower abundance of multidrug genes, when *Actinobacteria* enriched the abundance of corresponding ARGs like for vancomycin and tetracycline increased.

## Conclusions

In this study, we used a comprehensive database SARG to obtain hierarchically structured ARG profiles of soil metagenomes for four climatic areas and three types of land use. We observed the dominancy of multidrug efflux pump genes in the soil resistome, and it has a significant soil pH dependence but is independent of the sampling position. This reflects the natural selection of high environmental proton activity on the multidrug efflux pumps, especially for the PMF-driven inner membrane transferase. By applying ecological analysis tools such as composition dissimilarity, α-diversity, correlation networks, and assembly-model fitting we preliminary revealed the microbial participants in pH shaping soil resistome, and furtherly quantified the consistency between the deterministically assembled resistome and neutrally assembled microbial community. Our study proposed a novel perspective to understand the influence of edaphic factors on soil resistome, i.e., the natural selection on functional proteins (confer resistance mechnisms) has the ability to lead bottom-up selections on ARGs regardless of the microbial context. And such natural changes in resistome are herein suggested to be considered when assessing the actual impact of future human activities on the spread of ARGs.

## Materials And Methods

# Experimental design and sample description

The metagenomic sequence data used herein were derived from soil samples used in a previous study assessing the dominancy of comammox *Nitrospira* in soil nitrification [40]. In brief, these soil metagenomes were harvested from the forest, grassland and cropland, three kinds of ecosystems, which have developed in climate areas temperate monsoon, temperate continental, subtropical monsoon and plateau mountain, respectively. The site locations of samples used in this study are depicted in Figure S1(Additional file 1: S1). Soils (0–60 cm) were sampled by using the tubular soil sampler with a 5 cm diameter. The sampled soil column was spaced equally into six segments in situ to represent soils from 0–10 cm, 10–20 cm, 20–30 cm, 30–40 cm, 40–50 cm, and 50–60 cm depths, respectively.

Sampled soils per each site were well mixed at depths of 0–10 cm, 10–30 cm, and 30–60 cm, to mimic the soil horizons of organic surface (O), surface soil (A), and subsoil (B) for the follow-up investigation. A total of 36 samples were thus obtained for metagenomic analysis and 16s rRNA gene analysis. The measured pH of each sample ranged from 4.37 to 9.69, covering diverse denominations from extremely acidic soils to strongly alkaline soils (classified according to the rule of the U.S. Department of Agriculture [41]). In addition, detailed information on 17 abiotic factors, including climatic and physicochemical parameters for each sample, measured as described in the previous study [40], were listed together in Additional file 2: Tab. S1.

# DNA sequencing and quality control

DNA of the soil microbiome was extracted with a DNeasy PowerSoil Kit (Ref. 12888-100, Qiagen, Germany) following the introduction of manufacture. The concentration and purity of extracted DNA were comparatively measured using Nanodrop One (ThermoFisher Scientific, USA) and Qubit 2.0 (ThermoFisher Scientific, USA) at the same time. Gene libraries were generated by using the Next® Ultra™ DNA Library Prep Kit (New England Biolabs, USA) and sequenced on an Illumina Hiseq X-ten platform with 150-bp paired-end strategy by a sequencing server (Guangdong Magigene Biotechnology LTD, China). Approximately, 10 GB of raw data and $1.24 \times 10^{10}$ bases were obtained on average per sample. After removing adapter reads, quality control was processed by using Trimmomatic (Ver. 0.36 [42]) to discard bases with a quality score < 20 and length < 50 base pair. In summary, $1.03 \pm 0.30 \times 10^{10}$ clean bases per sample with the Q20 rate of 100% were generated as the Clean Data and used in later ARGs identification. More details about the raw data and Clean Data were given as Additional file 2: Tab. S2.

# Identification of ARGs in shotgun data

Clean Data per each sample were subsampled into small files of 20 million bases before the ARGs annotation. Afterward, using USEARCH (-ublast) to blast subsampled sequences with SARG v2.0 database (i.e. 12307 sequences at gene level) to detect ARG-like sequences. The ARG-like sequences were then identified by BLASTP against the SARG v2.0 database (i.e. 1208 subtypes at protein level) to classify sequences into the ARG types relating to antibiotic class, and the ARG subtypes within that antibiotic class. Threshold values used for the identification were set as recommended by the ARG-OAPs 2.0 pipeline [24], while as alignment length cut-off was 75 nucleotides; e-value cut-off was $10^{-7}$; identity was 80%. The abundance per each ARG subtype was then determined as the total number of identified ARG-like sequences in this subtype per million cleaned sequences (i.e. parts per million, ppm). The abundance of each ARG type was then the sum of the corresponding subtypes. The abundance of ARG per each antibiotic class and subtype was summarized in Additional file 2, as Tab. S4 and Tab. S8, respectively. The relative abundance of each ARG per sample was then determined as the percentage of the abundance of the ARG to the total abundance of each sample. Relative abundances of ARGs were shown in Additional file 2: Tab. S5 for ARG types and Tab. S9 for ARG subtypes, respectively. According to the recommended protocol [24], normalized abundances of ARGs were given in Additional file 2 as Tab. S6 for the abundance normalized to 16s rRNA genes and Tab. S7 for the abundance normalized to cell numbers.

# Soil bacterial community

Bacterial taxonomic classification and abundance quantification were performed with 16S rRNA gene amplicon sequencing with the previously published protocol [40], while primers 515F and 806R were used to target the V4 region. The quality of the raw data was controlled using Fastp (version 0.14.1). Afterward, 24662 operational taxonomic units (OTU) were determined by using UPARSE (Edgar 2013). OTUs were annotated against the SILVA database (Ver.132) by the usearch command (-sintax, with a default threshold = 0.8). After rarefaction, a normalized OTU table was obtained with 14527 reads per sample. In this study, bacterial community composition was investigated at the phylum level, and the relative abundance per phylum per sample was given in Additional file 2: Tab. S10.

# Construction and visualization of the correlation network

Sparse correlations for compositional data (i.e. SparCC, [43]) were utilized to deduce the potential correlation networks with the values for phylum composition, resistome composition, and environmental factors. SparCC, logarithmically scaled variances to calculate correlations, was calculated by R package “SpiecEasi” [43]. In addition, 100 times bootstrap was performed to test the reliability of the *p*-values to remove indirect correlations. Significant correlations were determined with the threshold value like |SparCC-correlation| > 0.60, $p < 0.05$. The correlation network was visualized in Gephi v0.9.3 [44]. At the network level, the closeness, the number of connected edges and the number of effective partners per node were calculated by using the R package “bipartite” [45], and the calculation result was given in Additional file 1: S8. To determine the reliability of topological features for the constructed network, 100 randomly rewired networks were generated by using Maslov-Sneppen procedure [46] and the significance of the difference between the constructed network and the random network was tested before follow-up analysis.

# Identification of ARGs in high quality binned genomes per sample

Binned microbial genomes per sample were constructed upon de novo assembly with Clean Data by using MEGAHIT[47], and processing parameters were set as k-min 35, k-max 115, and k-step 20. The statistics of de novo assembled scaffolds were given in Additional file 2: Tab. S3. Afterward, scaffolds were strictly dissociated at unknown bases to clean chimera, and the scaftigs were extracted as the contiguous sequences with a length greater than 500 bases and without unknown bases. The scaftigs per each sample were then taken for metagenomic binning by using MetaBAT2 with the default parameter set [48]. The quality of the binned genomes (i.e. bins) was accessed by CheckM software [49]. Only the high-quality bins, with estimated completeness greater than 50% and contamination less than 10%, were used for the subsequent ARGs annotation and taxonomic assignment. ARG profiles per each high-quality bin were identified by BLASTP against the SARG v2.0 database with e-value less than $10^{-7}$, similarity greater than 80% and query coverage over 70% as recommended [24].

Taxonomic assignment of high-quality bins was done phylogenetically with anvi’o v6.2 [50, 51]. Alignments for the single-copy core genes of the ribosomal protein per each high-quality bin was produced by searching GTDB with DIAMOND and successively used to construct maximum-likelihood trees in FastTree v2.1.5 with default settings and then visualized in the iTOL v3.

# Statistical analyses

If not otherwise specified all statistical analyses were done in R (V4.1.3) with packages “vegan”, “Hmisc”, and “psych”. The hierarchical agglomerative classification was done with the ‘hclust’ function in “stats” package. Dissimilarity analysis and Procrustes analysis for compositions of microbiome and resistome were done with “vegan” package based on Bray-Curtis index, and significant differences in compositions among groupings were tested with the Adonis method. In addition, changes in abundance of ARGs were

fitted to environmental factors and relative abundance of phylum by using the ‘envfit’ function, and the variance partitioning analysis was performed with the ‘varpart’ function. The structural equation model was conducted by the package “plspm”.

# Abbreviations

ARG
antibiotic resistance gene
MFS
major facilitator superfamily
MLS
macrolides-lincosamides-streptogramines
NMDS
non-metric multidimensional scaling
PMF
proton-motive-force
RND
resistance-nodulation-cell division superfamily
SparCC
Sparse correlations for compositional data
VPA
variance partitioning analysis

# Declarations

### Ethics approval and consent to participate

The manuscript does not report data collected from humans and animals.

### Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and analyzed during the current study are available at CNCB-NGDC (https://ngdc.cncb.ac.cn/) with BioProject accession: PRJCA009649.

### Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China (22193061) and National Key Research and Development Program of China (2020YFC1806903). Hangzhou Science and Technology Bureau, Agricultural and Social Development Scientific Research Project (20191203B66). Open Project of State Key Laboratory of Urban Water Resource and Environment, Harbin Institute and Technology (No. ES202118).

Authors' contributions

Z.L., B.H. and L.Z. developed the research topic. J.W., Y.Z and Z.L. performed the experiment and collected the data. Z.L. prepared figures 1-4 and table 1, Y.Z. prepared figures 5-6. Z.L. and B.H. worte the main manuscript text. All authors reviewed and approved the final manuscript for submission.

Acknowledgments

We would like thank Dr. Jiajie Hu, Qin Weng and Junxian Zhao for their efforts in the sample preparation. We would also like thank Xiangwu Yao, who enlightened us on the mechanisms of acid stress tolerance in microorganisms.

# References

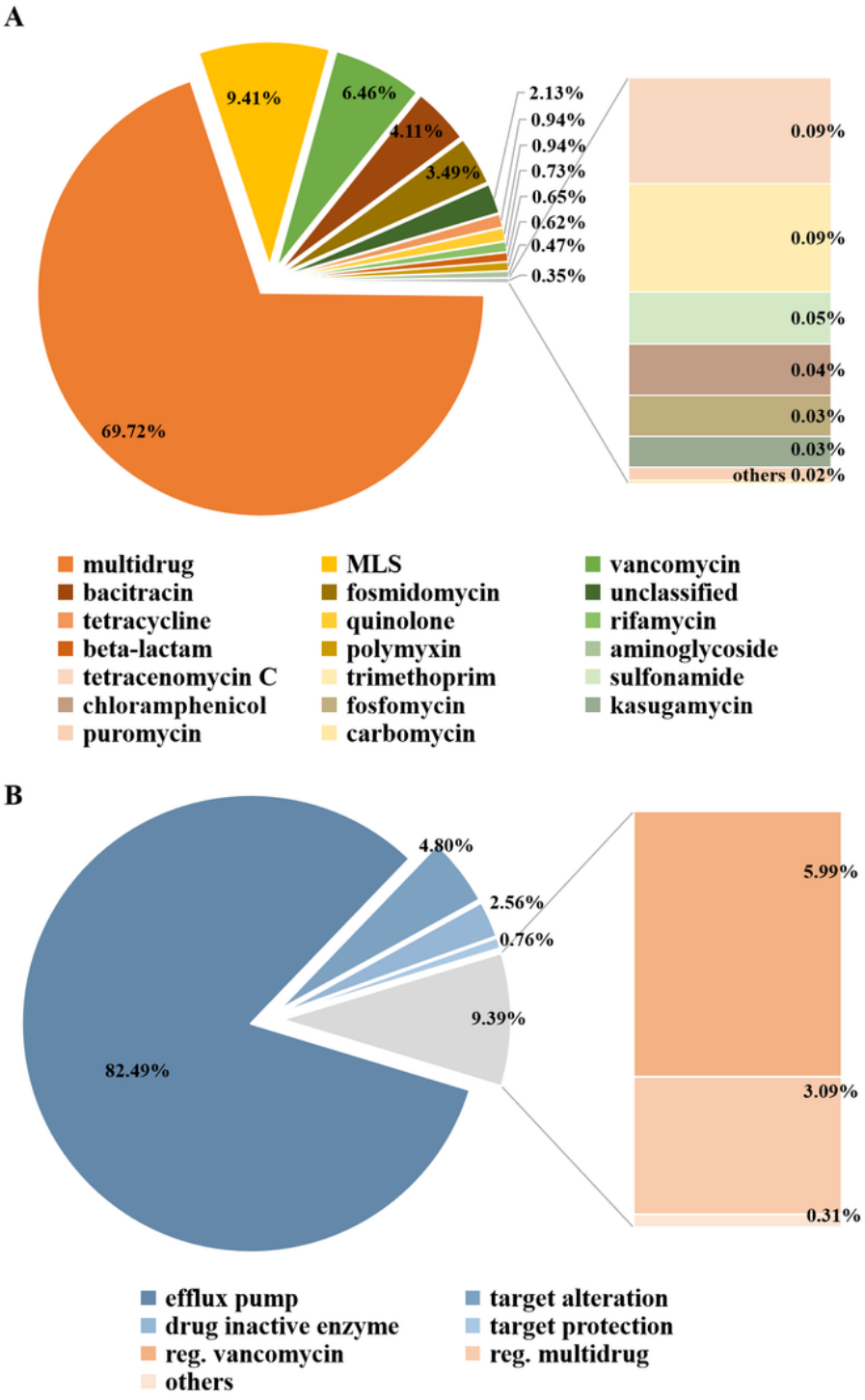
1. S Hernando-Amado,TM Coque,F Baquero,JL Martínez. Defining and combating antibiotic resistance from One Health and Global Health perspectives. Nat Microbiol. 2019; 4(9): 1432–1442.
2. TU Berendonk,CM Manaia,C Merlin,D Fatta-Kassinos,E Cytryn,F Walsh, et al. Tackling antibiotic resistance: the environmental framework. Nat Rev Microbiol. 2015; 13(5): 310–317.
3. D-W Kim,C-J Cha. Antibiotic resistome from the One-Health perspective: understanding and controlling antimicrobial resistance transmission. Exp Mol Med. 2021; 53(3): 301–309.
4. NB Jadeja,A Worrich. From gut to mud: dissemination of antimicrobial resistance between animal and agricultural niches. Environ Microbiol. 2022;
5. DGJ Larsson,C-F Flach. Antibiotic resistance in the environment. Nat Rev Microbiol. 2021;
6. J Bengtsson-Palme,DGJ Larsson. Concentrations of antibiotics predicted to select for resistant bacteria: Proposed limits for environmental regulation. Environ Int. 2016; 86: 140–149.
7. VM D'Costa,CE King,L Kalan,M Morar,WWL Sung,C Schwarz, et al. Antibiotic resistance is ancient. Nature. 2011; 477(7365): 457–461.
8. F Wang,Y-H Fu,H-J Sheng,E Topp,X Jiang,Y-G Zhu, et al. Antibiotic resistance in the soil ecosystem: A One Health perspective. Curr Opin Environ Sci Health. 2021; 20: 100230.
9. W Li,ZS Liu,BL Hu,LZ Zhu. Co-occurrence of crAssphage and antibiotic resistance genes in agricultural soils of the Yangtze River Delta, China. Environ Int. 2021; 156
10. JD Palmer,KR Foster. Bacterial species rarely work together. Science. 2022; 376(6593): 581–582.
11. M Ghoul,S Mitri. The ecology and evolution of microbial competition. Trends Microbiol. 2016; 24(10): 833–845.

12. M Bahram,F Hildebrand,SK Forslund,JL Anderson,NA Soudzilovskaia,PM Bodegom, et al. Structure and function of the global topsoil microbiome. Nature. 2018; 560(7717): 233–237.
13. X Jiang,MMH Ellabaan,P Charusanti,C Munck,K Blin,Y Tong, et al. Dissemination of antibiotic resistance genes from antibiotic producers to pathogens. Nat Commun. 2017; 8(1): 15784.
14. KJ Forsberg,A Reyes,B Wang,EM Selleck,MOA Sommer,G Dantas. The Shared Antibiotic Resistome of Soil Bacteria and Human Pathogens. Science. 2012; 337(6098): 1107–1111.
15. X Qian,S Gunturu,J Guo,B Chai,JR Cole,J Gu, et al. Metagenomic analysis reveals the shared and distinct features of the soil resistome across tundra, temperate prairie, and tropical ecosystems. Microbiome. 2021; 9(1): 108.
16. I Yelin,R Kishony. Antibiotic resistance. Cell. 2018; 172(5): 1136–1136 e1131.
17. JMA Blair,MA Webber,AJ Baylay,DO Ogbolu,LJV Piddock. Molecular mechanisms of antibiotic resistance. Nat Rev Microbial. 2015; 13(1): 42–51.
18. K Moon,JH Jeon,I Kang,KS Park,K Lee,C-J Cha, et al. Freshwater viral metagenome reveals novel and functional phage-borne antibiotic resistance genes. Microbiome. 2020; 8(1): 75.
19. O Mencia-Ares,R Cabrera-Rubio,JF Cobo-Diaz,A Alvarez-Ordonez,M Gomez-Garcia,H Puente, et al. Antimicrobial use and production system shape the fecal, environmental, and slurry resistomes of pig farms. Microbiome. 2020; 8(1): 164.
20. D Dai,C Brown,H Bürgmann,DGJ Larsson,I Nambi,T Zhang, et al. Long-read metagenomic sequencing reveals shifts in associations of antibiotic resistance genes with mobile genetic elements from sewage to activated sludge. Microbiome. 2022; 10(1): 20.
21. Y Che,X Xu,Y Yang,K Břinda,W Hanage,C Yang, et al. High-resolution genomic surveillance elucidates a multilayered hierarchical transfer of resistance between WWTP- and human/animal-associated bacteria. Microbiome. 2022; 10(1): 16.
22. Z Zhang,Q Zhang,T Wang,N Xu,T Lu,W Hong, et al. Assessment of global health risk of antibiotic resistance genes. Nat Commun. 2022; 13(1): 1553.
23. BP Alcock,AR Raphenya,TTY Lau,KK Tsang,M Bouchard,A Edalatmand, et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. 2019; 48(D1): D517-D525.
24. X Yin,X-T Jiang,B Chai,L Li,Y Yang,JR Cole, et al. ARGs-OAP v2.0 with an expanded SARG database and Hidden Markov Models for enhancement characterization and quantification of antibiotic resistance genes in environmental metagenomes. Bioinformatics. 2018; 34(13): 2263–2270.
25. M Boolchandani,AW D'Souza,G Dantas. Sequencing-based methods and resources to study antimicrobial resistance. Nat Rev Genet. 2019; 20(6): 356–370.
26. JC Thomas IV,A Oladeinde,TJ Kieran,JW Finger Jr.,NJ Bayona-Vásquez,JC Cartee, et al. Co-occurrence of antibiotic, biocide, and heavy metal resistance genes in bacteria from metal and radionuclide contaminated soils at the Savannah River Site. Microb Biotechnol. 2020; 13(4): 1179–1200.

27. G Zhu,X Wang,T Yang,J Su,Y Qin,S Wang, et al. Air pollution could drive global dissemination of antibiotic resistance genes. ISME J. 2021; 15(1): 270–281.
28. F Ju,K Beck,X Yin,A Maccagnan,CS McArdell,HP Singer, et al. Wastewater treatment plant resistomes are shaped by bacterial composition, genetic exchange, and upregulated expression in the effluent microbiomes. ISME J. 2019; 13(2): 346–360.
29. J Song,T Li,Z Zheng,W Fu,Z Long,N Shi, et al. Carbendazim shapes microbiome and enhances resistome in the earthworm gut. Microbiome. 2022; 10(1): 63.
30. PJF Henderson,C Maher,LDH Elbourne,BA Eijkelkamp,IT Paulsen,KA Hassan. Physiological functions of bacterial “multidrug” efflux pumps. Chem Rev. 2021; 121(9): 5417–5478.
31. I Alav,J Kobylka,MS Kuth,KM Pos,M Picard,JMA Blair, et al. Structure, assembly, and function of tripartite efflux and Type 1 secretion systems in gram-negative bacteria. Chem Rev. 2021; 121(9): 5479–5596.
32. O Lomovskaya,HI Zgurskaya,M Totrov,WJ Watkins. Waltzing transporters and 'the dance macabre' between humans and bacteria. Nat Rev Drug Discov. 2007; 6(1): 56–65.
33. PA Klenotic,MA Moseng,CE Morgan,EW Yu. Structural and Functional Diversity of Resistance–Nodulation–Cell Division Transporters. Chem Rev. 2021; 121(9): 5378–5416.
34. M Ji,TJ Williams,K Montgomery,HL Wong,J Zaugg,JF Berengut, et al. *Candidatus* Eremiobacterota, a metabolically and phylogenetically diverse terrestrial phylum with acid-tolerant adaptations. ISME J. 2021; 15(9): 2692–2707.
35. LP Coelho,R Alves,ÁR del Río,PN Myers,CP Cantalapiedra,J Giner-Lamia, et al. Towards the biogeography of prokaryotic genes. Nature. 2022; 601(7892): 252–256.
36. F Maguire,B Jia,KL Gray,WYV Lau,RG Beiko,FSL Brinkman. Metagenome-assembled genome binning methods with short reads disproportionately fail for plasmids and genomic Islands. Microb Genom. 2020; 6(10)
37. K Lewis. The science of antibiotic discovery. Cell. 2020; 181(1): 29–45.
38. H-S Kim,D Nagore,H Nikaido. Multidrug efflux pump MdtBC of *Escherichia coli* is active only as a B2C heterotrimer. J Bacteriol. 2010; 192(5): 1377–1386.
39. D Jakubczyk,F Dussart. Selected Fungal Natural Products with Antimicrobial Properties. Molecules. 2020; 25(4): 911.
40. J Hu,Y Zhao,X Yao,J Wang,P Zheng,C Xi, et al. Dominance of comammox Nitrospira in soil nitrification. Sci Total Environ. 2021; 780: 146558.
41. Soil Science Division Staff. Chapter 3: Examination and description of soil profiles. In: Ditzler, C, Scheffe, K, and Monger, H, editors. Soil survey manual, USDA Handbook 18. Washington, DC: Government Printing Office; 2017. p. 83–233.
42. AM Bolger,M Lohse,B Usadel. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30(15): 2114–2120.

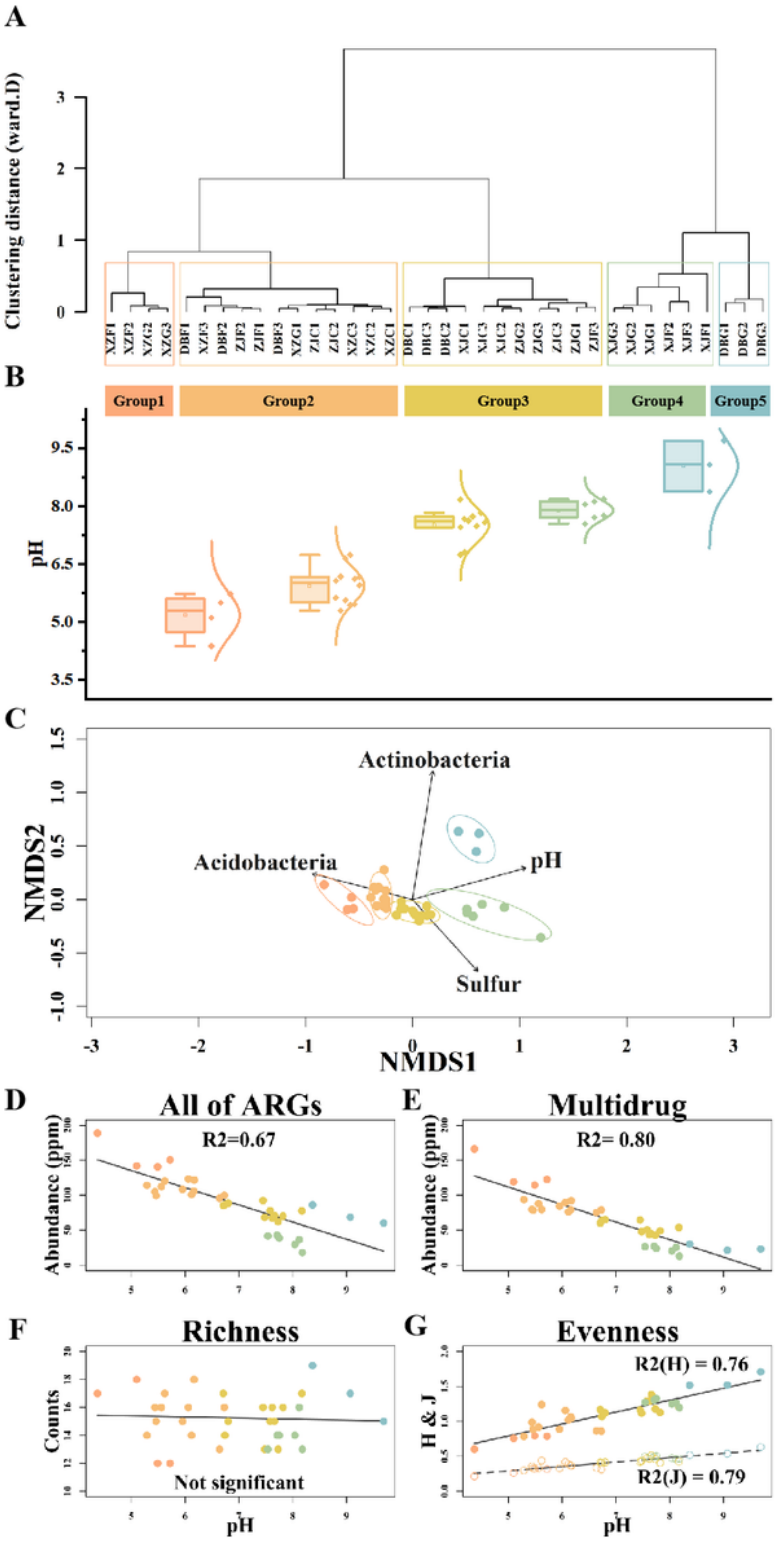
43. ZD Kurtz,CL Müller,ER Miraldi,DR Littman,MJ Blaser,RA Bonneau. Sparse and compositionally robust inference of microbial ecological networks. PloS Comput Biol. 2015; 11(5): e1004226.
44. M Bastian,S Heymann,M Jacomy. Gephi: An Open Source Software for Exploring and Manipulating Networks. ICWSM. 2009; 3(1): 361–362.
45. CF Dormann,J Fründ,N Blüthgen,B Gruber. Indices, graphs and null models: analyzing bipartite ecological networks. Open J Ecol. 2009; 2(1): 7–24.
46. S Maslov,K Sneppen. Specificity and stability in topology of protein networks. Science. 2002; 296(5569): 910–913.
47. D Li,C-M Liu,R Luo,K Sadakane,T-W Lam. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015; 31(10): 1674–1676.
48. DD Kang,F Li,E Kirton,A Thomas,R Egan,H An, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ. 2019; 7: e7359.
49. DH Parks,M Imelfort,CT Skennerton,P Hugenholtz,GW Tyson. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015; 25(7): 1043–1055.
50. DH Parks,M Chuvochina,DW Waite,C Rinke,A Skarshewski,P-A Chaumeil, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol. 2018; 36(10): 996–1004.
51. B Buchfink,C Xie,DH Huson. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015; 12(1): 59–60.

# Figures




## Figure 1

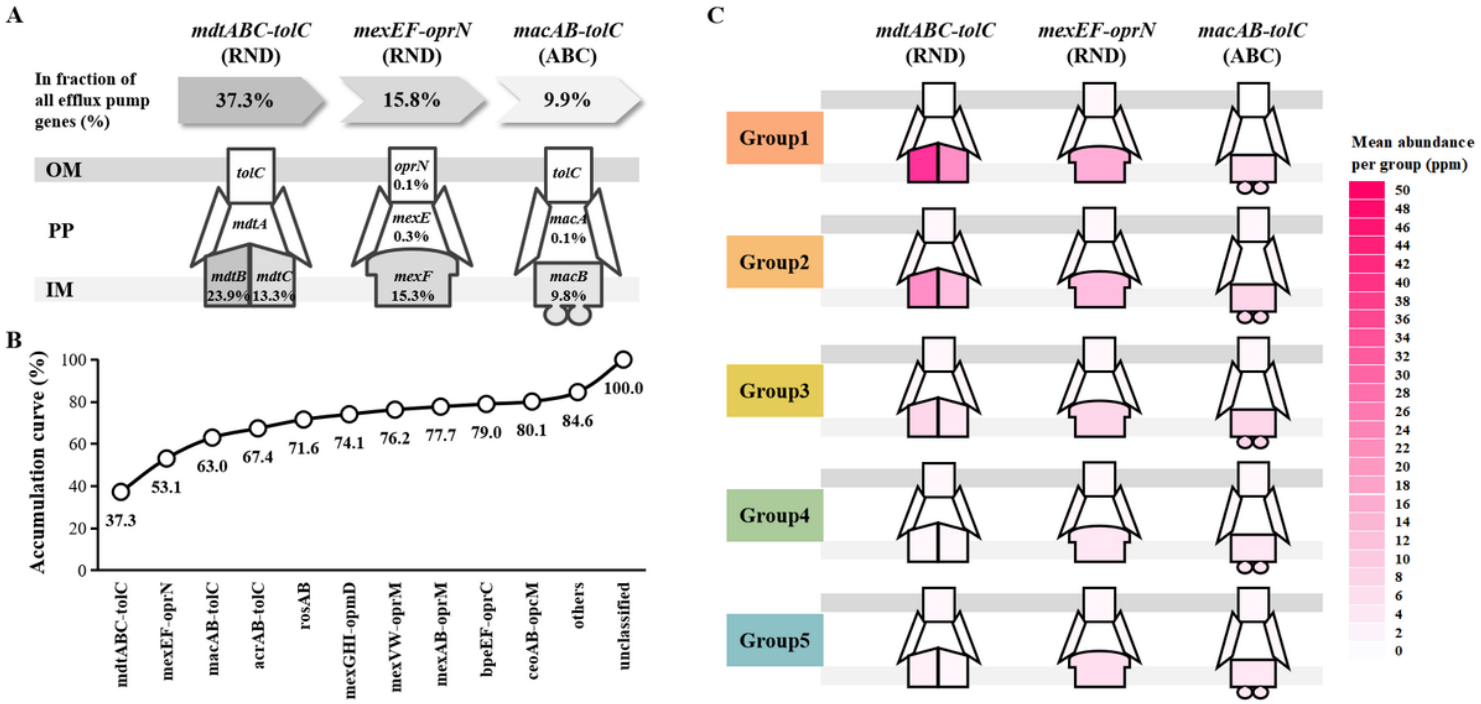
ARG profile of natural soil metagenomes. Mean relative abundance of ARGs per **A)** antibiotic classes and **B)** resistance mechanisms.




## Figure 2

Composition analysis for natural soil resistome with ARGs at antibiotic-class level. **A)** hierarchical agglomerative classification, **B)** summarize the pH values in Group1-5 that determined according the hierarchical clusters, **C)** Dissimilarities of resistome per sample were calculated on the basis of ARGs abundance (Bray-Curtis index) and visualized with a NMDS plot against the influential factors, i.e. soil pH, sulfur content, relative abundance of *Actinobacteria* and *Acidobacteria*, Samples from different groups

(points in colors) were separated from each other, and significance of differences between groups was demonstrated by pairwise ADONIS (Additional file 1: S4), **D-G)** linear regression of pH against abundance of all of ARGs, abundance of multidrug, richness of ARGs per sample, and evenness of ARGs per sample (H stand for Shannon index and J for Pielou index).



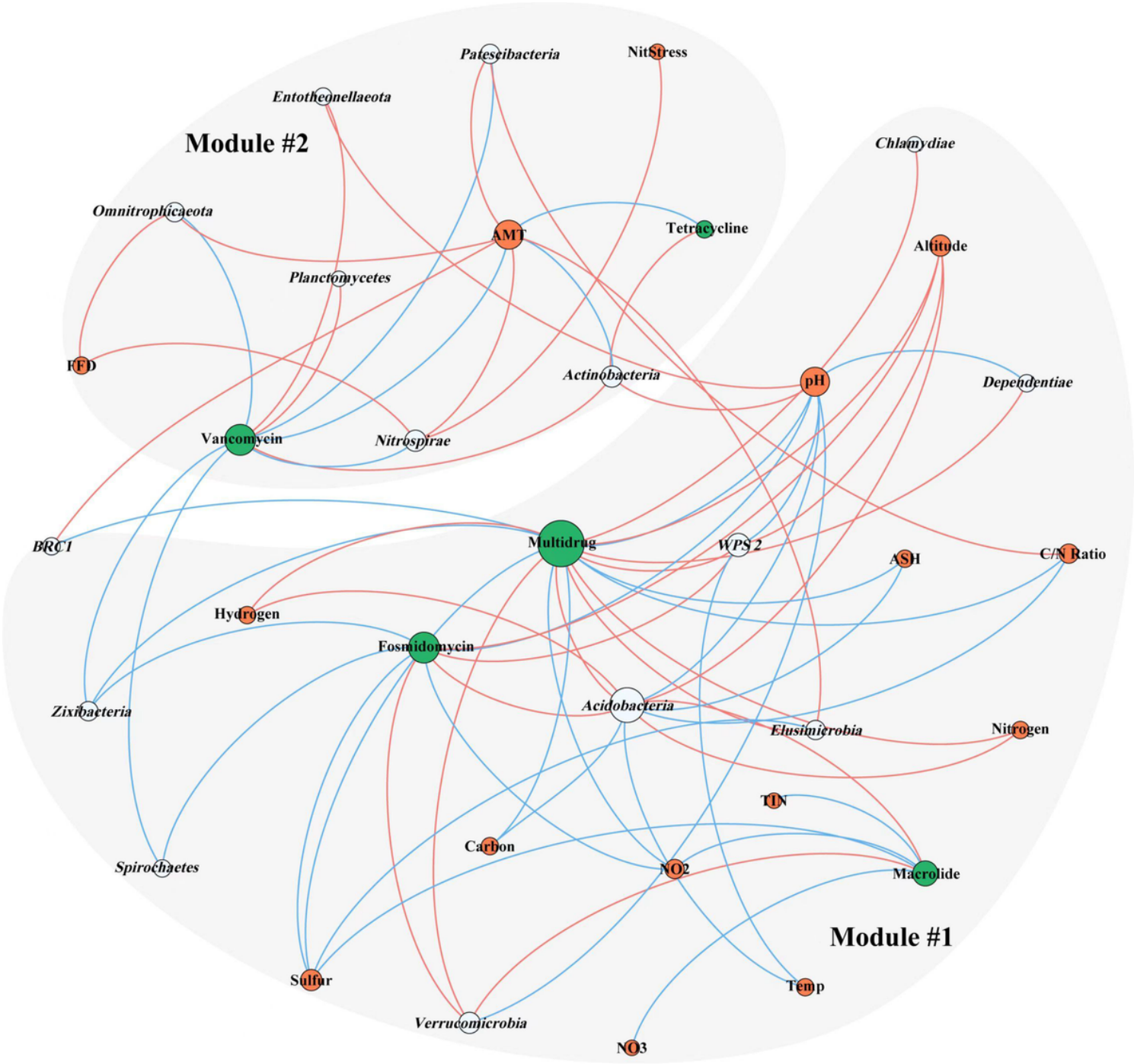

## Figure 3

predominant efflux pumps genes detected in this study, their abundance distribution and protein family affiliations. **A)** structural sketch of protein complexes for three most abundant efflux pumps, their relative gene abundances were labeled and position on cell membrane were indicated (OM: outer membrane, PP: pericellular plasm, IM: inner membrane), **B)** the abundance accumulation curve for the top 10 abundant efflux pumps, **C)** changes in the average gene abundance of *mdtABC-tolC*, *mexEF-oprN* and *macAB-tolC* were visualized in every pH group, and each component of the efflux pump complex were painted separately.

## Figure 4

Consistency of changes in microbial community composition and resistome composition. **A)** Variation of community composition at phylum level per sample and its consistency with the variation of ARGs, **B)** variation of resistome composition at ARG subtype level per sample and its consistency with the variation of ARGs. Solid point shows the position of a soil sample in the ordination based on the composition of phylum or ARG subtype, and empty points (source of arrows) in the transformed

ordination defined upon resistome composition of ARGs. The same grouping was performed according to Figure 2C.




## Figure 5

Legend not included with this version

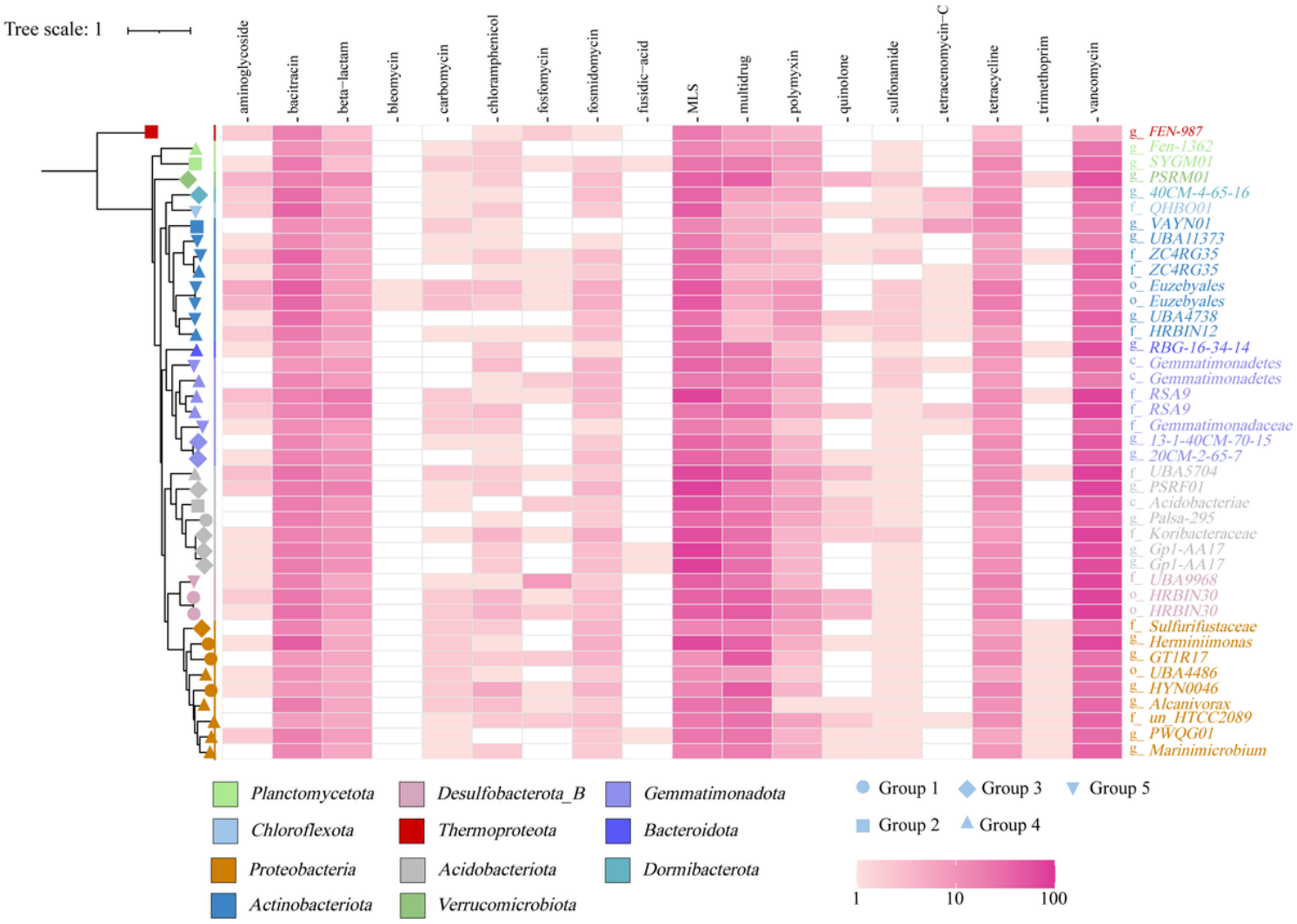



**Figure 6**

ARG profiles of metagenome-assembled genomes (MAGs). A total of 41 high-quality MAGs were generated from the de novo assembly of each sample. Species annotation and phylogenetic analysis of MAGs were done with GTDB, and the results are shown as the phylogenetic tree on the left, as well as the bottom-level recognized taxonomy are labeled on the right. As shown in the legend, the color and shape of the markers represent their phylum and grouping (sample grouping is the same as Figure 2C), respectively. Number of detected ARGs were counted per antibiotic class per each MAGs and visualized as a heatmap with color codes representing the counts of ARGs.

# Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.docx
- Additionalfile2.xlsx