Deterministic effects of pH on 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, Effects of pH, Proton motive force, Multidrug efflux pump

Posted Date: July 19th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1673295/v2

Abstract

The high prevalence of multidrug efflux pump genes in natural ecosystems recently raised concerns. Since most of the efflux pumps were driven by proton-motive-force (PMF), studying whether soil pH is a determinant for the selection of multidrug efflux pump genes and thus shaping soil resistome is of great interest. In this study, we collected soils with pH values ranging from 4.37-9.69 from multiple ecosystems and profiled the composition of antibiotics resistance genes (ARGs) for 36 soil metagenomes and 41 metagenome-assembled genomes by referencing a structured ARG database. We observed the multidrug efflux pump genes enriched in the acidic soil resistome, and it has a significant soil pH dependence. This reflects the natural selection of high soil proton activity on the multidrug efflux pump genes, especially for the PMF-driven inner membrane transferase. In addition, we preliminary indicate the putative microbial participants in pH shaping soil resistome by applying ecological analyzing tools such as composition dissimilarity analysis, correlation networks analysis, structural equation model test and assembly-model fitting. And furtherly quantified the consistency between the deterministically assembled resistome and neutrally assembled microbiome. Our study proposed a novel perspective to understand the influence of edaphic factors on soil resistome, i.e., the natural selection of fitted resistance mechanisms could lead to 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.

Introduction

The spread of antimicrobials resistance genes (ARGs) poses a huge threat to global public health, that cannot be understated (Murray *et al.* 2022). To tackle this global health challenge, it is essential to grasp the microbial interconnections between the environment and humans (i.e. the One Health framework), as bacteria and their gene elements have a high ability in crossing geographical and species boundaries (Berendonk *et al.* 2015, Hernando-Amado *et al.* 2019, Kim and Cha 2021). The soil environment is reported as the major source and reservoir of ARGs (Li *et al.* 2021, Wang *et al.* 2021), where tens of millions of microbes meet and interact via direct contact and/or metabolites (Sokol *et al.* 2022). Soil bacteria and fungi secret antibiotics is one of the typical bio-interactions, it induced the resistance of surrounding microbes (Ghoul and Mitri 2016, Palmer and Foster 2022). In addition, some clinically relevant ARGs are recently proved to be originated from the soil resistome (Forsberg *et al.* 2012, Jiang *et al.* 2017), which has thus led to an increased interest in understanding mechanisms for the generation, evolution and propagation of ARGs in the soil environment.

Although ARGs developed in clinical and natural environments with different driving forces, the rapid evolution and spread of ARGs in past decades were undoubtedly related to the utilization and excretion of antibiotics. In particular, antibiotics exert extensive selective pressures on microbes, so that significant ARG mobilization, horizontal transfer and mutation-based generation occur in scenarios such as hospitals, farmlands and sewage treatment plants (Jadeja and Worrich 2022). However, such selective pressures are indeed uncommon for most natural environments in which antibiotic concentrations were

generally lower than their estimated (or known) minimum selective concentrations in the laboratory, i.e. at ppb or even lower levels (Bengtsson-Palme and Larsson 2016, Larsson and Flach 2022). Instead, the physicochemical properties of the environmental interfaces, and mixing effects of coexisting compounds have more profound impacts on determining the types and abundances of ARGs (Obermeier *et al.* 2021, Yang *et al.* 2021a, b, Zhu *et al.* 2022).

In recent studies through the metagenomic analysis with structured databases and newly developed bioinformatics tools, new insights into the environmental resistome were uncovered (Alcock *et al.* 2019, Boolchandani *et al.* 2019, Yin *et al.* 2018), among which the high prevalence of multidrug resistance (MDR) genes in the soil has raised concerns (Qian *et al.* 2021, Thomas IV *et al.* 2020, Yi *et al.* 2022, Zhu *et al.* 2021). The vast majority of MDR genes are associated with efflux pumps, which are integral membrane transport proteins that function to extrude antibiotics from bacterial cells (Henderson *et al.* 2021). The “multidrug” was named as typically they did not recognize only one class of antibiotics but a wide range of chemically dissimilar compounds, and thus a single efflux pump may provide resistance to a broad range of antimicrobial classes (Alav *et al.* 2021, Henderson *et al.* 2021). For instance, the tripartite AcrAB-TolC complex 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 could already induce high intrinsic resistance to antibiotics (Lomovskaya *et al.* 2007). Moreover, the running force of the efflux pump is not complicated, because of most efflux pumps (e.g. RND: resistance-nodulation-cell division superfamily and MFS: major facilitator superfamily) required only the proton-motive-force (PMF) that is composed of the pH gradient and electrochemical potential inside and outside the cytoplasmic membrane (Klenotic *et al.* 2021). Whether all these physiological functions as a whole benefit the prevalence of multidrug resistance genes; and how they are influenced by edaphic factors, especially by the soil pH (i.e. proton activity), need to be further explored.

Mineral matrices, pedogenesis process, climatic conditions, and land-use practices can significantly alter the physicochemical as well as biological properties of the soil (Chen *et al.* 2022). Therefore, to investigate the effects of edaphic factors 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 referring SARG database (Yin *et al.* 2018). 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 shaping soil resistome, as most MDR running with PMF; and if yes (iii) are there any marker microorganisms exist and what is the linkage behind them to ARGs. 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.

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 (Hu *et al.* 2021). In brief, these soil samples 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, and then 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. Thus 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 (Staff 2017). In addition, detailed information on 17 abiotic factors per each sample, including climatic and physicochemical parameters like altitude, on site temperature, C/N ratio, elements contents of carbon, nitrogen, sulfur, etc, which were listed together in Additional file 2: Tab. S1. Parameters were measured as described in the previous study (Hu *et al.* 2021).

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 1.24×10^{10} bases were obtained on average per sample. After removing adapter reads, quality control was processed by using Trimmomatic (Ver. 0.36) (Bolger *et al.* 2014) 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 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 (Yin *et al.* 2018), 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 (Yin *et al.* 2018), 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 (Hu *et al.* 2021), 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.

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 (Li *et al.* 2015), 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 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 (Kang *et al.* 2019). The quality of the binned genomes (i.e. bins) was accessed by CheckM software (Parks *et al.* 2015). 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 (Yin *et al.* 2018).

Taxonomic assignment of high-quality bins was done phylogenetically with anvi'o v6.2 (Buchfink *et al.* 2015, Parks *et al.* 2018). 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”.

Results

ARG profiles of natural soils

In this study, a total of 36 soil metagenomes were studied with well-characterized metadata. All in all, 264 ARG subtypes potentially associated with 20 classes of antibiotics were identified from 4.46×10^{11} bases of cleaned short reads (Additional file 2: Table S2-S9).

As presumed, the multidrug resistance gene was detected in all samples and was the most abundant ARG type with an average relative abundance of 69.72% (Figure 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%.

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 among all resistance mechanisms (Figure 1B), and far exceeding the 2.56% of 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 abundance of ARGs per sample (Figure 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 (Figure 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, the abundance of ARG 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 Group1 to 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 (Figure 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 (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 variance partitioning analysis (VPA). And as a consequence, four most 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 Figure 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 Figure 2D for the total abundance of ARGs; Figure 2E for the abundance of multidrug resistance; Figure 2F for ARG richness as the number of ARG types per sample; and Figure 2G for ARG evenness with Shannon index (H) and Pielou index (J). 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, Additional file 1: S6), while the regression for ARG richness per sample was not significant (p value = 0.764). This means examined abundance and diversity indicators of ARGs, except richness, responded linearly to pH changes over the studied range of 4.37-9.69.

ARGs for inner membrane transferase predominated resistome in low pH soils

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

In summary, a total of 89 detected ARG subtypes were associated with the efflux mechanism (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 (Figure 3A) and genes encoded inner membrane transferase in special, i.e. *mdtBC*, *mexF* and *macB*, were predominating.

According to the abundance accumulation curve (Figure 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 (Figure 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 resistome and microbiome changes in natural soils

To reveal the relationship between microbiome and resistome, the consistency of compositional changes was examined at first (Figure 4). The dissimilarity of microbial community was analyzed at the phylum level with the same workflow as for resistome, as the result (Figure 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 (Figure 2C) and ARG subtype level (Figure 4B).

On this basis, the consistency of microbiome and resistance group changes was quantified by Procrustes analysis (Figure 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), resulting 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).

The divergence between resistome and microbiome could also be supported by testing community assembly with a neutral model (Additional file 1: S7) and correlation analysis between ARGs and abio-/biotic parameters (Additional file 1: S8). Although microbial communities investigated in this study were neutrally assembled at the phylum level (fitted R2=0.727, Figure S7), the resistome was assembled by deterministic selection (R2=-0.142, not fitted to the model). As well as, the change of resistome in the soil was more closely related to the environmental factors rather than the phylum composition (Figure S8 and Table S8.2). To underpin these conclusions, a structural equation model was tested (SEM, 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 (Figure S9.1).

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 mapped 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 Figure 5. These MAGs were phylogenetically classified into one archaeal phylum (i.e. *Thermoproteota*) and 10 bacterial phyla. And ARGs relevant to 18 classes of antibiotics 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% to 100%. However, low detected ARGs for kasugamycin, puromycin and carbomycin in shotgun data were not detected 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

As introduced above, it is not surprising to found ARGs are ubiquitous in the soil, since many antibiotics orientated from natural secondary metabolites of soil microbes, and were yielded by microbes as chemical arms in the competition for survival (Lewis 2020, Wang *et al.* 2021). And ARGs served as genetic blueprints of defensive fortifications that is necessary for the survival of the wild 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 need to be explored. In this study to examine the soil ARGs characteristics, we profiled 36 metagenomes and 41 high-quality MAGs by referring to the SARG database along with a recommended pipeline (Yin *et al.* 2018). The compositional background of ARGs in soils and potential microbial hosts were thus revealed. As a result, ARGs of efflux pump mechanisms dominating soil resistome in terms of either total abundance, average relative abundance or detection frequency were shown (Figure 2 and Additional file 2: Table S4-S9).

The dominancy of MDR genes might reflect their higher fitness in natural soils. As Larsson and Flach (2022) 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. 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 and easy access driving force (Henderson *et al.* 2021). For instance, MdtABC-TolC was the efflux pump with the highest average abundance detected in this study (Figure 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 (Alav *et al.* 2021). Similarly, MexEF-OprN could pump out aromatic hydrocarbons and population-sensing signaling molecules; MacAB-TolC extrudes lipopolysaccharides and peptides as non-antibiotic resistant functions (Henderson *et al.* 2021). The efflux of these compounds is closely related to interfacial colonization and nutrient acquisition of soil bacteria and are themselves a compensatory mechanism for the cost of resistance expression. Furthermore, consistent with our observation, previous studies also reported the prevalence of MDR genes in natural and anthropogenically affected soil (Qian *et al.* 2021, Thomas IV *et al.* 2020, Yi *et al.* 2022). In the study of the Savannah River Site (Thomas IV *et al.* 2020), high abundances 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 were attributed to the natural selection of soil properties (Qian *et al.* 2021).

Based on the prevalence of MDR 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 (Figure 2D). Moreover, all the changes in total ARG abundance seem to be caused by the changes in multidrug resistance genes (Figure 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 (Figure 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 proton transportation (Kim *et al.* 2010). 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.* (2018) 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. 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 (Jakubczyk and Dussart 2020). 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, easy access to protons was relatively limited, thus the fitness of MDR 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 species foundation behind the ARGs variation, a strict correlation analysis was performed in this study, where indirect correlations were deduced from networks (Figure S8). 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” (Forsberg *et al.* 2012, Jiang *et al.* 2017). 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. A similar conclusion was also given by Qian *et al.* (2021), who demonstrated environmental factors select bacterial populations and hence enrich the ARGs they carried. 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.

Consistent with the conclusion from correlation analysis, the causality between ARGs and environmental factors was higher weighted than it between ARGs and bacterial phyla in the result of SEM test (Figure S9.1). Especially for considering the significant effect of pH on the variation of ARGs, we thus assumed that in acidic soils ARGs corresponding to high fitness efflux pumps may be directly enriched with the easy proton access rather than through the enlarger of a particular bacterial population. Consistently, we found *mdtBC* and *mexF* genes were carried by more than half MAGs including bacteria from phylum

Acidobacteria, *Gemmatimonadota*, *Proteobacteria, Verrucomicrobiota* etc., and no specific taxonomic affiliation were found (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 4.37): 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, relatively enriched the abundance of ARGs for vancomycin and tetracycline increased.

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

Competing interests

The authors declare that they have no competing interests.

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.

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; Y.Z. prepared figures 5. Z.L. and B.H. worte the main manuscript text. All authors reviewed and approved the final manuscript for submission.

Acknowledgments

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

References

Alav I., Kobylka J., Kuth M.S., Pos K.M., Picard M., Blair J.M.A. *et al.* (2021). Structure, assembly, and function of tripartite efflux and Type 1 secretion systems in gram-negative bacteria. *Chem Rev* **121:** 5479-5596.

Alcock B.P., Raphenya A.R., Lau T.T.Y., Tsang K.K., Bouchard M., Edalatmand A. *et al.* (2019). CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res* **48:** D517-D525.

Bahram M., Hildebrand F., Forslund S.K., Anderson J.L., Soudzilovskaia N.A., Bodegom P.M. *et al.* (2018). Structure and function of the global topsoil microbiome. *Nature* **560:** 233-237.

Bengtsson-Palme J., Larsson D.G.J. (2016). Concentrations of antibiotics predicted to select for resistant bacteria: Proposed limits for environmental regulation. *Environ Int* **86:** 140-149.

Berendonk T.U., Manaia C.M., Merlin C., Fatta-Kassinos D., Cytryn E., Walsh F. *et al.* (2015). Tackling antibiotic resistance: the environmental framework. *Nat Rev Microbiol* **13:** 310-317.

Bolger A.M., Lohse M., Usadel B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30:** 2114-2120.

Boolchandani M., D'Souza A.W., Dantas G. (2019). Sequencing-based methods and resources to study antimicrobial resistance. *Nat Rev Genet* **20:** 356-370.

Buchfink B., Xie C., Huson D.H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12:** 59-60.

Chen Q.-L., Hu H.-W., Yan Z.-Z., Zhu Y.-G., He J.-Z., Delgado-Baquerizo M. (2022). Cross-biome antibiotic resistance decays after millions of years of soil development. *ISME J* **16:** 1864-1867.

Edgar R.C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* **10:** 996-998.

Forsberg K.J., Reyes A., Wang B., Selleck E.M., Sommer M.O.A., Dantas G. (2012). The Shared Antibiotic Resistome of Soil Bacteria and Human Pathogens. *Science* **337:** 1107-1111.

Ghoul M., Mitri S. (2016). The ecology and evolution of microbial competition. *Trends Microbiol* **24:** 833-845.

Henderson P.J.F., Maher C., Elbourne L.D.H., Eijkelkamp B.A., Paulsen I.T., Hassan K.A. (2021). Physiological functions of bacterial “multidrug” efflux pumps. *Chem Rev* **121:** 5417-5478.

Hernando-Amado S., Coque T.M., Baquero F., Martínez J.L. (2019). Defining and combating antibiotic resistance from One Health and Global Health perspectives. *Nat Microbiol* **4:** 1432-1442.

Hu J., Zhao Y., Yao X., Wang J., Zheng P., Xi C. *et al.* (2021). Dominance of comammox Nitrospira in soil nitrification. *Sci Total Environ* **780:** 146558.

Jadeja N.B., Worrich A. (2022). From gut to mud: dissemination of antimicrobial resistance between animal and agricultural niches. *Environ Microbiol* DOI: doi.org/10.1111/1462-2920.15927.

Jakubczyk D., Dussart F. (2020). Selected Fungal Natural Products with Antimicrobial Properties. *Molecules* **25:** 911.

Jiang X., Ellabaan M.M.H., Charusanti P., Munck C., Blin K., Tong Y. *et al.* (2017). Dissemination of antibiotic resistance genes from antibiotic producers to pathogens. *Nat Commun* **8:** 15784.

Kang D.D., Li F., Kirton E., Thomas A., Egan R., An H. *et al.* (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7:** e7359.

Kim D.-W., Cha C.-J. (2021). Antibiotic resistome from the One-Health perspective: understanding and controlling antimicrobial resistance transmission. *Exp Mol Med* **53:** 301-309.

Kim H.-S., Nagore D., Nikaido H. (2010). Multidrug efflux pump MdtBC of *Escherichia coli* is active only as a B2C heterotrimer. *J Bacteriol* **192:** 1377-1386.

Klenotic P.A., Moseng M.A., Morgan C.E., Yu E.W. (2021). Structural and Functional Diversity of Resistance–Nodulation–Cell Division Transporters. *Chem Rev* **121:** 5378-5416.

Larsson D.G.J., Flach C.F. (2022). Antibiotic resistance in the environment. *Nat Rev Microbiol* **20:** 257-269.

Lewis K. (2020). The science of antibiotic discovery. *Cell* **181:** 29-45.

Li D., Liu C.-M., Luo R., Sadakane K., Lam T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31:** 1674-1676.

Li W., Liu Z.S., Hu B.L., Zhu L.Z. (2021). Co-occurrence of crAssphage and antibiotic resistance genes in agricultural soils of the Yangtze River Delta, China. *Environ Int* **156**.

Lomovskaya O., Zgurskaya H.I., Totrov M., Watkins W.J. (2007). Waltzing transporters and 'the dance macabre' between humans and bacteria. *Nat Rev Drug Discov* **6:** 56-65.

Murray C.J.L., Ikuta K.S., Sharara F., Swetschinski L., Robles Aguilar G., Gray A. *et al.* (2022). Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet* **399:** 629-655.

Obermeier M.M., Wicaksono W.A., Taffner J., Bergna A., Poehlein A., Cernava T. *et al.* (2021). Plant resistome profiling in evolutionary old bog vegetation provides new clues to understand emergence of multi-resistance. *ISME J* **15:** 921-937.

Palmer J.D., Foster K.R. (2022). Bacterial species rarely work together. *Science* **376:** 581-582.

Parks D.H., Imelfort M., Skennerton C.T., Hugenholtz P., Tyson G.W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25:** 1043-1055.

Parks D.H., Chuvochina M., Waite D.W., Rinke C., Skarshewski A., Chaumeil P.-A. *et al.* (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* **36:** 996-1004.

Qian X., Gunturu S., Guo J., Chai B., Cole J.R., Gu J. *et al.* (2021). Metagenomic analysis reveals the shared and distinct features of the soil resistome across tundra, temperate prairie, and tropical ecosystems. *Microbiome* **9:** 108.

Sokol N.W., Slessarev E., Marschmann G.L., Nicolas A., Blazewicz S.J., Brodie E.L. *et al.* (2022). Life and death in the soil microbiome: how ecological processes influence biogeochemistry. *Nat Rev Microbiol* 10.1038/s41579-022-00695-z.

Staff S.S.D. (2017). Chapter 3: Examination and description of soil profiles. In: Ditzler C, Scheffe K, Monger H (eds). *Soil survey manual, USDA Handbook 18*. Government Printing Office: Washington, DC. pp 83-233.

Thomas IV J.C., Oladeinde A., Kieran T.J., Finger Jr. J.W., Bayona-Vásquez N.J., Cartee J.C. *et al.* (2020). 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* **13:** 1179-1200.

Wang F., Fu Y.-H., Sheng H.-J., Topp E., Jiang X., Zhu Y.-G. *et al.* (2021). Antibiotic resistance in the soil ecosystem: A One Health perspective. *Curr Opin Environ Sci Health* **20:** 100230.

Yang B., Cheng X., Zhang Y., Li W., Wang J., Guo H. (2021a). Insight into the role of binding interaction in the transformation of tetracycline and toxicity distribution. *Environ Sci Ecotechnol* **8:** 100127.

Yang B., Cheng X., Zhang Y., Li W., Wang J., Guo H. (2021b). Probing the roles of pH and ionic strength on electrostatic binding of tetracycline by dissolved organic matters: Reevaluation of modified fitting model. *Environ Sci Ecotechnol* **8:** 100133.

Yi X., Liang J.-L., Su J.-Q., Jia P., Lu J.-l., Zheng J. *et al.* (2022). Globally distributed mining-impacted environments are underexplored hotspots of multidrug resistance genes. *ISME J* 10.1038/s41396-022-01258-z.

Yin X., Jiang X.-T., Chai B., Li L., Yang Y., Cole J.R. *et al.* (2018). 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* **34:** 2263-2270.

Zhu D., Ma J., Li G., Rillig M.C., Zhu Y.-G. (2022). Soil plastispheres as hotspots of antibiotic resistance genes and potential pathogens. *ISME J* **16:** 521-532.

Zhu G., Wang X., Yang T., Su J., Qin Y., Wang S. *et al.* (2021). Air pollution could drive global dissemination of antibiotic resistance genes. *ISME J* **15:** 270-281.

Figures

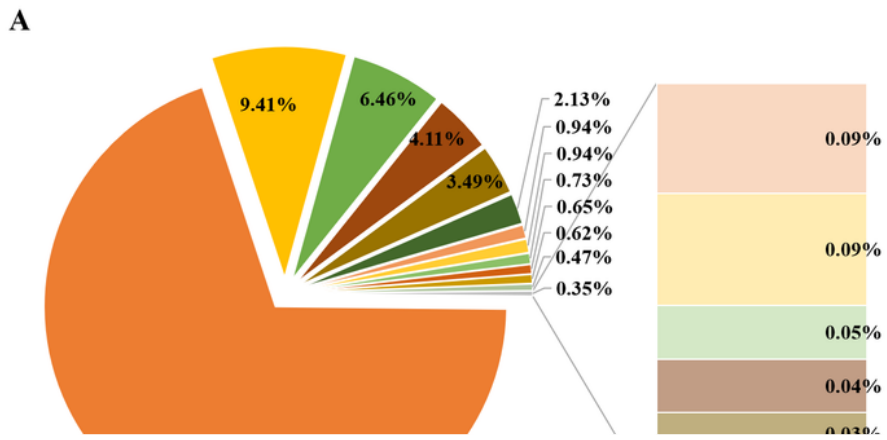

Figure 1

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).

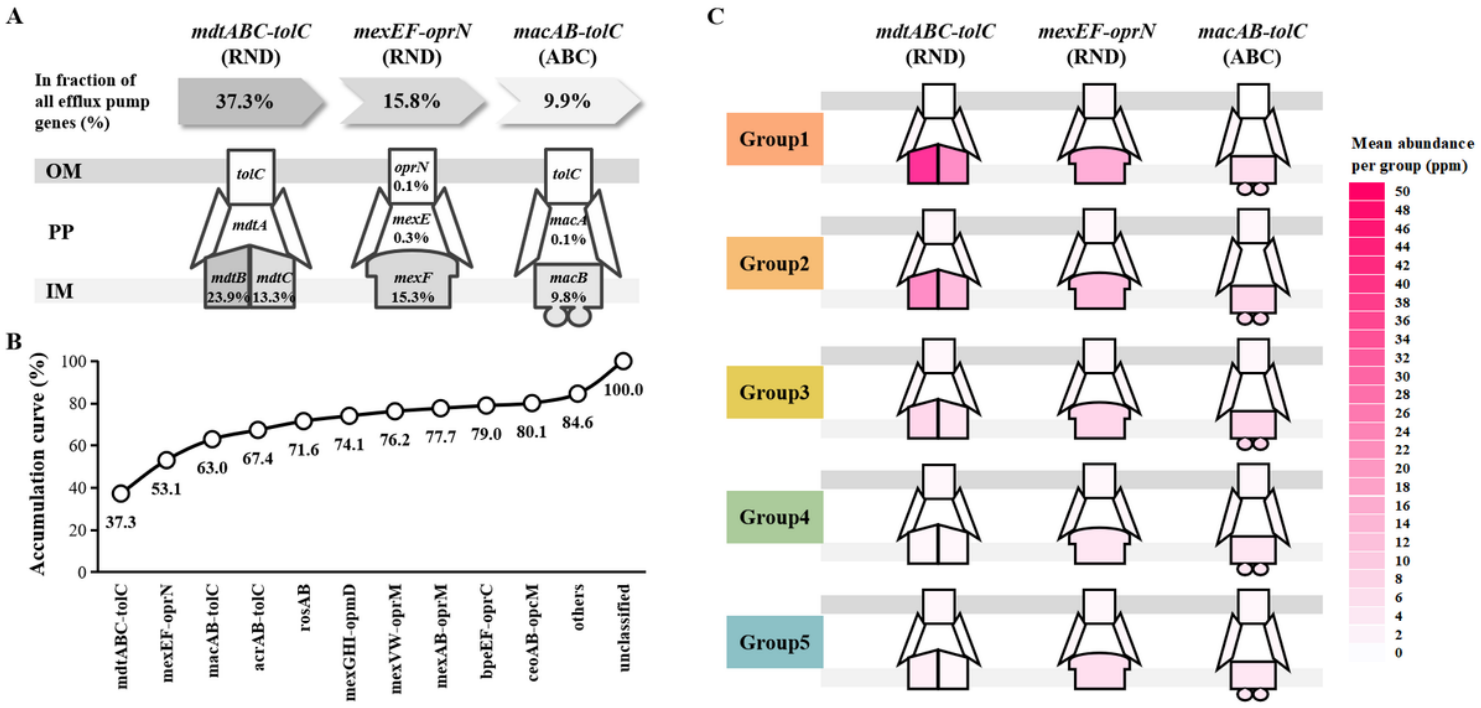

Figure 3

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

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.pdf
- Additionalfile2.xlsx