

MICROBIOLOGY

Global biogeography and projection of soil antibiotic resistance genes

Dongsheng Zheng^{1,2,3†}, Guoyu Yin^{1,2,*}, Min Liu^{1,2*}, Lijun Hou^{4*}, Yi Yang^{1,2}, Thomas P. Van Boeckel^{5,6}, Yanling Zheng^{1,2}, Ye Li^{1,2}

Although edaphic antibiotic resistance genes (ARGs) pose serious threats to human well-being, their spatially explicit patterns and responses to environmental constraints at the global scale are not well understood. This knowledge gap is hindering the global action plan on antibiotic resistance launched by the World Health Organization. Here, a global analysis of 1088 soil metagenomic samples detected 558 ARGs in soils, where ARG abundance in agricultural habitats was higher than that in nonagricultural habitats. Soil ARGs were mostly carried by clinical pathogens and gut microbes that mediated the control of climatic and anthropogenic factors to ARGs. We generated a global map of soil ARG abundance, where the identified microbial hosts, agricultural activities, and anthropogenic factors explained ARG hot spots in India, East Asia, Western Europe, and the United States. Our results highlight health threats from soil clinical pathogens carrying ARGs and determine regions prioritized to control soil antibiotic resistance worldwide.

INTRODUCTION

Antibiotic resistance is a growing threat of huge concern to many countries and sectors, which enables microorganisms to grow when exposed to antibiotics (1, 2). Antibiotic resistance in wide range of pathogens is recognized as “One Health” problems (3, 4), threatening the Sustainable Development Goal 3, good health, and well-being (5, 6). A review report (7) has predicted that nearly 700,000 people per year lose their lives worldwide because of infections by antibiotic resistance bacteria, and accumulative economic loss was expected to reach 100 trillion U.S. dollars by 2050 if we do not take strong actions. A more recent study (8) has estimated that 4.95 million people lost their lives worldwide in 2019 because of infections by antibiotic resistance bacteria, and those directly attributable to antibiotic resistance infections reached 1.27 million. To address such a threat, the World Health Organization (WHO) launched the global action plan that aims at combating antibiotic resistance across varied environments associated with human and animal health at the global scale (9).

Edaphic environments, essential reservoirs for antibiotic resistance genes (ARGs) (10), serve as important habitats for many pathogens associated with clinical infection (11, 12) and plant disease outbreak (13, 14). One of the most serious concerns is that ARGs would transfer from soils to anthropogenic, animal, and plant settings, thus posing severe threats to human and livestock health and food security (10, 15, 16). Despite wide studies on soil ARGs, particularly their occurrence patterns and interactions with ecological factors (13, 17, 18), a quantitative high-resolution map of soil ARGs

and their responses to environmental constraints at the global scale remains largely lacking. This knowledge gap is not only restricting the WHO global action plan to identify the priority regions combating soil antibiotic resistance but also limiting the understanding of health risk posed by soil antimicrobial resistance. Generating a spatially explicit understanding of soil ARGs at the global scale is hindered by amplification-based approaches, common ARG detection techniques that cannot provide detailed ARG profiles because of their capacity limitation and nonspecific amplification problem (19, 20). Sequencing-based approaches, such as metagenomics, are strong alternatives that align all the sequenced DNA segments against reference databases to detect ARGs (11, 21). Another challenge is that we lack a mechanistic model to quantitatively simulate the transfer of ARGs in environments. This gap is likely bridged by machine learning algorithms that can provide predictive values and mechanistic understandings (22, 23).

Here, 1088 metagenomics samples are applied to profile soil antibiotic resistome (fig. S1A and table S1) and mobile gene elements (MGEs). We identify microbes carrying ARGs and MGEs to explain the microbial driver of ARGs. Afterward, we disentangle the response of ARGs to environmental constraints using structural equation model. Last, we generate the first global map of soil ARG abundance at a 0.083° of resolution using advanced machine learning algorithms and 169 spatial covariates (table S2). This map discloses the hot spots of antibiotic resistance and determines regions prioritized to control soil antibiotic resistance worldwide.

RESULTS

Spatial patterns of soil ARGs

Our 1088 soil observations resulted in a total of 23 ARG types (Fig. 1A and table S3) and 558 ARG subtypes (Fig. 1A and table S4). These ARGs represented on average 12.42 types and 49.24 subtypes of ARGs at a given sample, most of which were genes encoding resistance to multidrug (100% of detected ARG type), macrolide-lincosamide-streptogramin (MLS; 99.08% of detected ARG type), vancomycin (98.53% of detected ARG type), and

Copyright © 2022
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim to
original U.S. Government
Works. Distributed
under a Creative
Commons Attribution
NonCommercial
License 4.0 (CC BY-NC).

¹Key Laboratory of Geographic Information Science (Ministry of Education), East China Normal University, Shanghai 200241, China. ²School of Geographic Sciences, East China Normal University, Shanghai 200241, China. ³Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing 100084, China. ⁴State Key Laboratory of Estuarine and Coastal Research, East China Normal University, Shanghai 200241, China. ⁵Health Geography and Policy Group, ETH Zürich, Switzerland. ⁶Center for Disease Dynamics, Economics, and Policy, Washington DC, USA.

*Corresponding author. Email: gyyin@geo.ecnu.edu.cn (G.Y.); mliu@geo.ecnu.edu.cn (M.L.); ljhou@sklec.ecnu.edu.cn (L.H.)

†These authors contributed equally to this work.

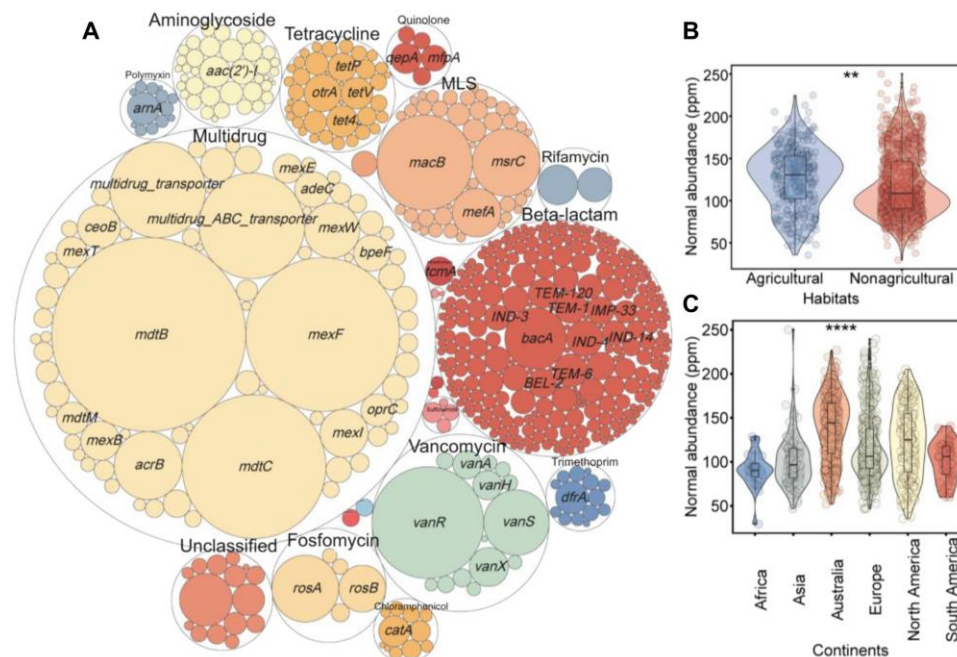


Fig. 1. Composition and distribution of soil ARGs. (A) ARG composition colored by ARG types. Outer and inner circles represent ARG types and subtypes, respectively. Circle size is proportional to the normalized abundance of ARGs. MLS refers to genes conferring resistance to macrolide-lincosamide-streptogramin. Normalized abundance of ARGs across habitats (B) (degree of freedom = 1, F value = 8.04) and continents (C) (degree of freedom = 5, F value = 23.45) were shown with error bars (SDs). Agricultural habitats include farmland and pasture, while nonagricultural habitats consist of desert, forest, peatland, and permafrost. The significance among different continents and habitats is examined using analysis of variance (ANOVA) (** $P < 0.01$ and **** $P < 0.0001$). The unit of ARGs is ppm (ARGs-like sequences per million sequencing reads).

fosmidomycin (annotated by 98.25% of samples) (Fig. 1A and table S3). The most frequently detected ARG subtype was *mexF* (annotated by 1085 samples), followed by *mdtB*, *multidrug_ABC_transporter*, *multidrug_transporter*, *mdtC*, and *macB* (Fig. 1A and table S4). The normalized abundance of ARGs in surveyed soil samples reached, on average, 121.20 parts per million (ppm; ARGs-like sequences per million sequencing reads), ranging from 29.34 to 250.02 ppm. Consistent with the detected number of ARGs, the normalized abundance of ARGs was dominated by genes conferring resistance to multidrug (Fig. 1A). The normalized abundance of multidrug resistance genes was, on average, 86.02 ppm, contributing to 70.97% of the total normalized abundance of ARGs, followed by genes encoding resistance to vancomycin (with the average at 12.93 ppm), MLS (with the average at 6.66 ppm), and fosmidomycin (with the average at 4.20 ppm) (Fig. 1A and fig. S1B).

A substantial ARG composition discrepancy was found in varied habitats using permutational multivariate analysis of variance [(PERMANOVA), coefficient of determination (R^2) = 0.11, $P < 0.001$; fig. S1C]. We observed higher ARG abundance in agricultural habitats (with the average at 126.46 ppm) compared with nonagricultural habitats (with the mean at 119.14 ppm) ($P < 0.01$, ANOVA; Fig. 1B). Similar to the difference among varied habitats, the observed ARG composition differed evidently from one continent to another (PERMANOVA, $R^2 = 0.10$, $P < 0.001$; fig. S1D). Specifically, Australia exhibited the most abundant ARGs with an average of 138.66 ppm, ranging from 52.08 to 226.75 ppm, followed by North America, Europe, South America, Asia, and Africa, harboring, on average, 122.25, 116.73, 101.03, 100.98, and 90.55 ppm of ARGs, respectively ($P < 0.0001$, ANOVA; Fig. 1C). We examined

whether ARGs followed a latitudinal gradient pattern, and a trend of increasing normalized abundance of ARGs from low latitudes to the poles was revealed ($R^2 = 0.19$, $P < 0.0001$; fig. S1E).

Soil MGEs and ARG microbial hosts

We identified 9 MGE types and 157 MGE subtypes that would facilitate the prevalence and persistence of soil ARGs mediated via horizontal gene transfer (Fig. 2A and fig. S2A). The total normalized abundance of MGEs varied two orders of magnitude (7.31 to 391.05 ppm), with an average of 55.26 ppm (fig. S2B). The dominated MGE type was transposase (including 45 subtypes) with an average of 25.43 ppm, constituting 46.03% of the total MGE abundance (Fig. 2A and fig. S2C). The second abundant MGEs was insertion sequences with an average of 23.77 ppm, accounting for 43.03% of the total detected MGE abundance, followed by *ist* (with an average of 3.70 ppm) and insertion sequence common region (with an average of 1.38 ppm) (Fig. 2A and fig. S2C). The normalized abundance of ARGs was notably correlated with those of all MGEs ($R = 0.17$, $P < 0.001$; fig. S2D), transposase ($R = 0.17$, $P < 0.001$; fig. S2E), and integrase ($R = 0.12$, $P = 0.015$; fig. S2F). We also illustrated positive associations between the normalized abundance of ARGs and those of MGE subtypes, namely *istB* ($R = 0.12$, $P = 0.010$; fig. S2G) and *IS91* ($R = 0.13$, $P = 0.006$; fig. S2H).

Normalized abundance of soil microbes ranged from 52,627.96 to 488,083.97 ppm with an average of 232,179.53 ppm across all the soil samples (fig. S3A). The annotated microbiome represented 4 kingdoms, 69 phyla, 134 classes, 265 orders, 618 families, 2397 genera, and 15,071 species (Fig. 2B). Despite the similar composition of microbial communities among all the samples, a limited

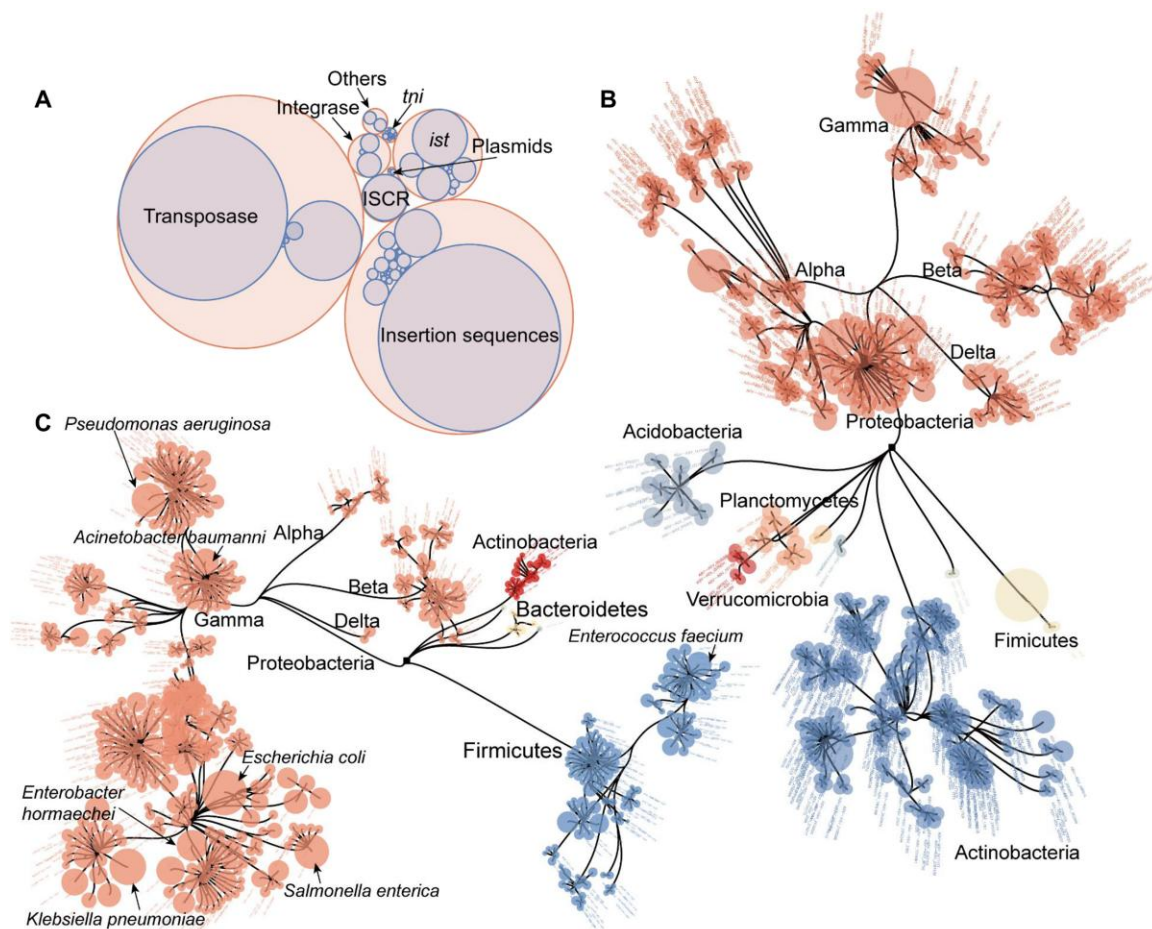


Fig. 2. Soil MGEs and microbial composition. (A) Composition and normalized abundance of MGEs. Red circles represent MGE types, and blue circles denote MGE subtypes. Circle size is proportional to the normalized abundance of MGEs. ISCR, insertion sequence common region. (B) Phylogenetic characterization of the 700 most abundant species in soils, which was dominated by the phylum of Proteobacteria and Actinobacteria. (C) Phylogenetic characterization of the most 700 abundant species carrying MGEs or ARGs in soils. Circle size is proportional to the normalized abundance of species. Circles are colored by their phyla. Lines connecting circles indicate phylogenetic relationship.

number of microbes dominated nearly across all the taxonomic levels (Fig. 2B). In kingdom level, bacteria were the dominant taxa with 230,408.27 ppm on average, constituting 98.22% of total microbial gene sequences (table S5). In phylum level, Proteobacteria dominated across all the samples, accounting for 53.81% of total microbial gene sequences (table S6). The most abundant classes in the phylum of Proteobacteria were Alphaproteobacteria, Gammaproteobacteria, Betaproteobacteria, and Deltaproteobacteria (Fig. 2B), which constituted 26.68, 12.44, 11.82, and 2.75% of total aligned microbial sequences, respectively (table S7). Actinobacteria was the second abundant phylum that occupied 31.58% of total sequences aligned with microbes (Fig. 2B and table S6), and this phylum was dominated by the class of Actinobacteria (30.09% of total microbial sequences, table S7). In order level, the most abundant one was Rhizobiales (belonged to Alphaproteobacteria) that occupied 18.66% of the total sequences with microbes, followed by Burkholderiales (belonged to Betaproteobacteria) with 10.33% of the total sequences with microbes, and Streptomycetales (one order of Actinobacteria) contributing 7.70% of the total microbial gene sequences (table S8).

We identified microbes that carried ARGs or MGEs (Fig. 2C), which represented 20.97% of the total observed species (fig. S3B), despite contributing merely, on average, 1.36% of the normalized abundance of microbiomes (fig. S3C). These identified microbial hosts composed of 19 phyla, 35 classes, 95 orders, 228 families, 682 genera, and 3160 species (Fig. 3), most of which (1490 species) harbored both MGEs and ARGs, with 886 species solely carrying ARGs and 784 species only with MGEs (Fig. 3). The composition of antibiotic resistance microbiomes (Fig. 2C) was largely different from that of soil microbial communities (Fig. 2B). In phylum level, the second abundant phylum of soil microbial communities was Firmicutes (Fig. 2B and table S6), while that of the antibiotic resistance bacteria belonged to Actinobacteria (Fig. 2C and table S9). In class level, Gammaproteobacteria was the most abundant and diverse class with ARGs and MGEs in the phylum of Proteobacteria (Fig. 2C and table S10), whereas the counterpart of soil microbial communities was Alphaproteobacteria (Fig. 2B and table S7). The identified antibiotic resistance bacteria in order and species levels were dominated by potential clinical pathogens and gut microbes that inhabited commonly in anthropogenic and animal

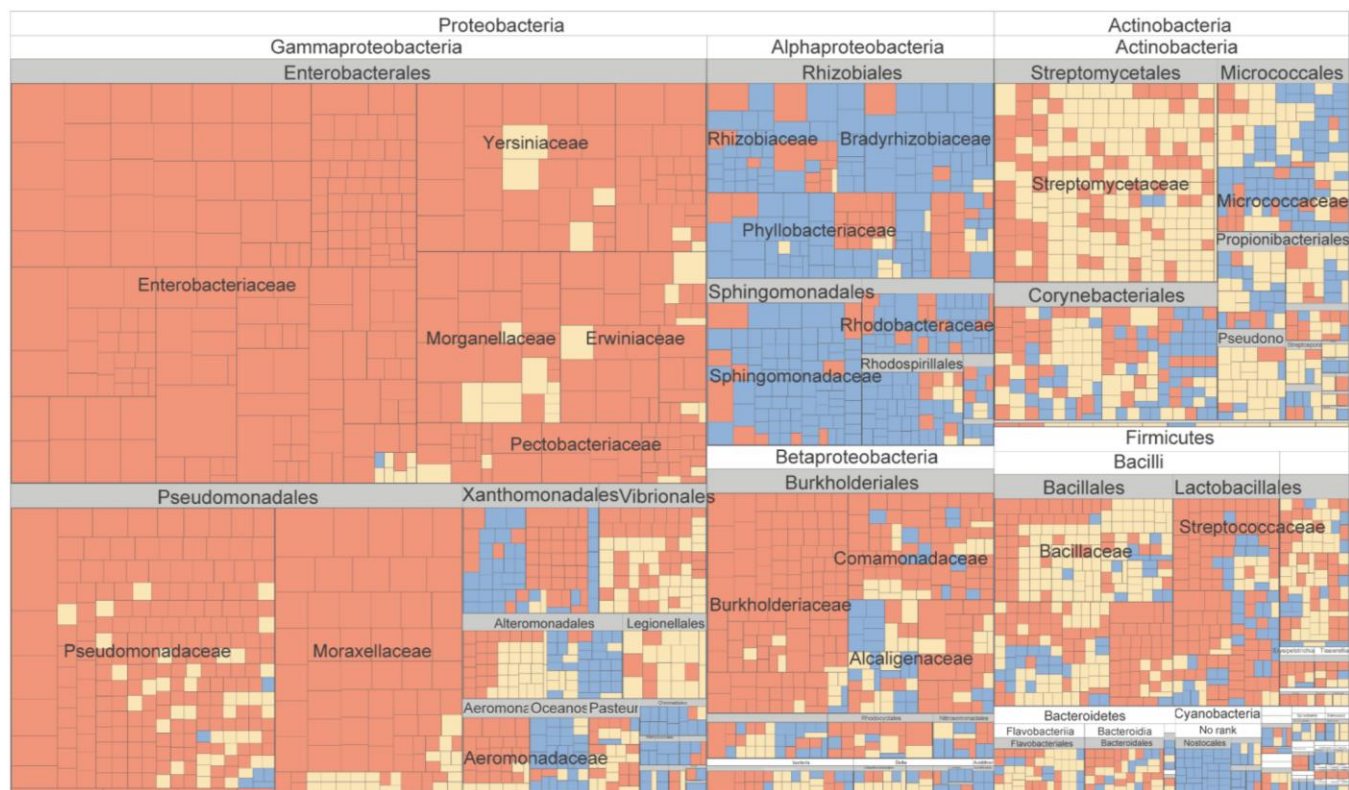


Fig. 3. Phylogenetic characterization of soil microbes carrying MGEs and ARGs Each block represents a species with ARGs or MGEs, which are colored by their carried genes (carrying ARGs for yellow, MGEs for blue, and both for red). Block size is proportional to the normalized abundance of antibiotic resistance microbes.

environments (tables S11 and 12). These species spanned across the order of Enterobacteriales (for example, *Escherichia coli* that was identified as the most abundant host species with 39.11 ppm on average; *Klebsiella pneumoniae* with the mean at 31.10 ppm) and Pseudomonadales (e.g., *Pseudomonas aeruginosa* with 32.22 ppm on average) (Figs. 2C and 3 and table S12). Notably, these orders were also the core hosts of multidrug resistance genes (the dominant ARG type; Fig. 1A and fig. S4A) and transposases (the most abundant MGE type; Fig. 2A and fig. S4A). The normalized abundance of ARGs and their microbial hosts were positively correlated ($R = 0.21$, $P < 0.001$; fig. S4B), and such a positive association was also revealed between MGEs and their microbial hosts ($R = 0.37$, $P < 0.001$; fig. S4C).

Geographical drivers of soil ARGs

To examine the geographical mechanisms that drive the spatial patterns of soil ARG abundance, we integrated the potential environmental constraints into 16 principal components (Fig. 4A and table S13). The variability of ARG abundance was largely controlled by principal components associated with anthropogenic activities (57.52%) that mostly represents the contribution of human inputs, animal husbandry, and agricultural contaminations to soil ARGs (Fig. 4A). In contrast, a relatively small part of ARG abundance variability was contributed by climate and vegetation (17.18%) and soil nutrients (7.07%) (Fig. 4A). The effects of climate and plant factors were not completely separated and were partly interconnected (table S13). Our analyses further illustrated positive associations between ARG abundance and the most

anthropogenic factors (Fig. 4, B and C). For example, livestock production ranked the most important driver increased with soil ARG abundance across its wide range of values ($P < 0.01$; Fig. 4, A and B). We also found that increasing soil nutrients resulted in elevated ARG abundance, particularly in its intermediated range ($P < 0.01$; Fig. 4, A and C). Climate parameters showed much more complicated patterns, where ARG abundance decreased with temperature ($P < 0.01$; Fig. 4, A and B) but increased with precipitation ($P < 0.05$; Fig. 4, A and C). We examined direct and indirect cause effects of geographical attributes on the normalized abundance of ARGs through structural equation model (Fig. 4D and fig. S5). Our model illustrated that the impacts of geographical drivers on soil ARGs were ultimately mediated via microbial factors. Anthropogenic drivers, such as livestock production, irrigation, and manure, would introduce microorganisms carrying ARGs and MGEs, indirectly raising the normalized abundance of ARGs ($R > 0.10$, $P < 0.05$; Fig. 4D). Similarly, the impacts of climatic variables on ARGs were also indirect and mediated via soil nutrients, ARG hosts, MGE hosts, and MGEs (Fig. 4D). For example, a lower temperature would allow a higher content of soil nutrients ($R = -0.54$, $P < 0.001$; Fig. 4D), which facilitated the proliferation of ARG microbial hosts ($R = 0.24$, $P < 0.05$; Fig. 4D), ultimately raising the normalized abundance of ARGs ($R = 0.17$, $P < 0.01$; Fig. 4D).

Mapping soil ARG normalized abundance

We used 169 spatial covariates to predict the normalized abundance of soil ARGs on the basis of four candidate machine learning

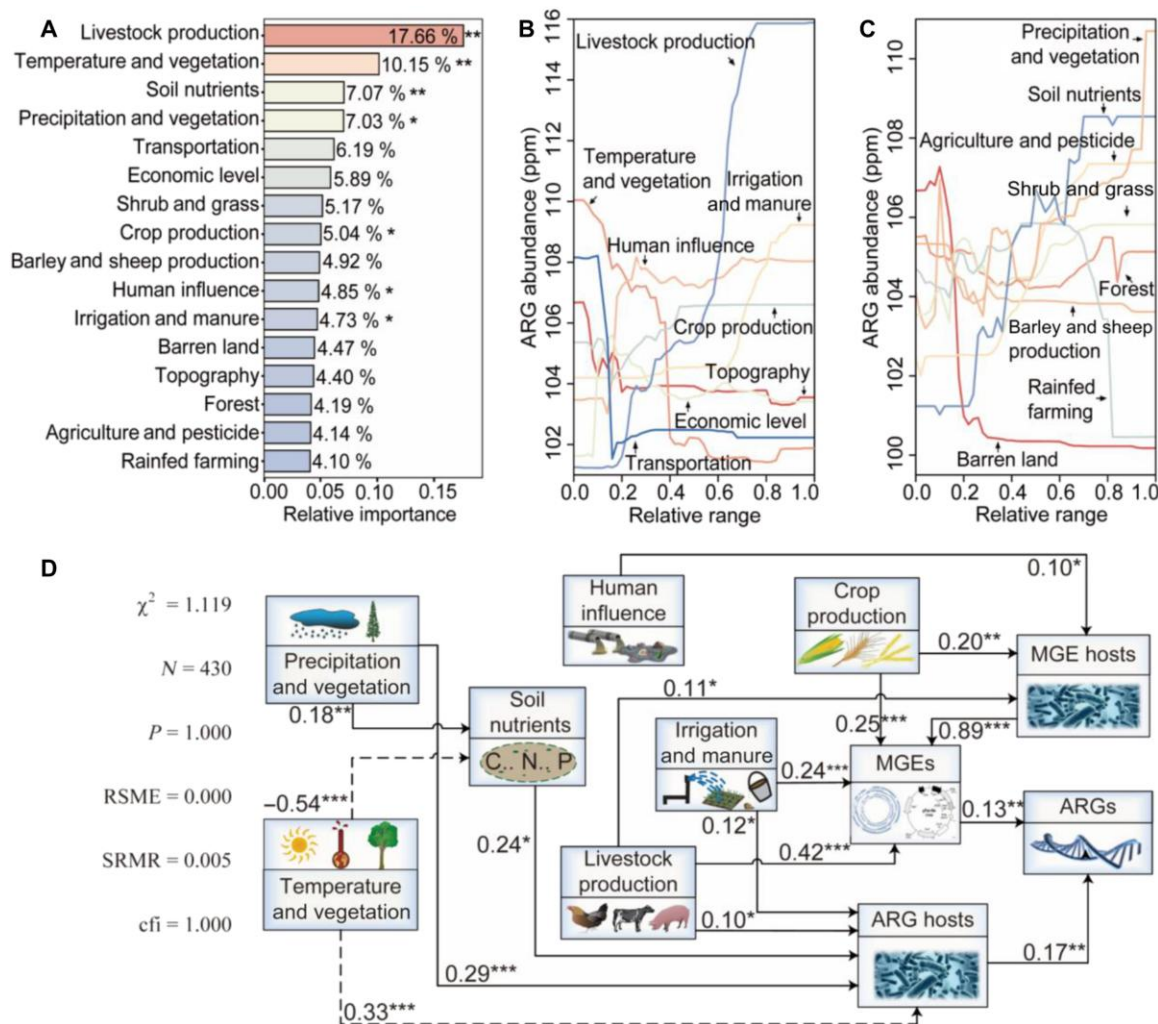


Fig. 4. Geographical variables affect soil ARGs through microbes. (A) Relative importance of the principal components of geographical variables reveals the dominance of anthropogenic activities on the soil ARG abundance. Partial dependence plots show the impacts of the first (B) and the second (C) eight principal components on ARG abundance. (D) Structural equation model differentiating the impact pathways of geographic principal components ARG abundance. Numbers adjacent to arrows indicate path coefficients, and asterisks are the significance of pathways. *, **, and *** represent $P < 0.05$, $P < 0.01$, and $P < 0.001$ significance level, respectively.

algorithms (support vector machine, k -nearest neighbor, gradient boosting tree, and random forest). These candidate algorithms were optimized through feature selection (fig. S6 and table S14) and parameter tuning (fig. S7 and table S15) with 10-fold cross-validation. Ultimately, random forest with 71 independent predictors outperformed other models with a relatively high confidence (table S15 and fig. S8). The final model, together with its optimal predictive covariates, enabled us to extend this relationship across the global scale to construct an atlas of the normalized abundance of ARGs at a 0.083° of resolution (Fig. 5A). This map disclosed the highest normalized abundance of ARGs in Western Europe (Fig. 5B), East Asia (Fig. 5C), South Asia (Fig. 5D), and eastern United States (Fig. 5E), characterized with highly dense population across the globe. Along with these population-driven ARG hot spots, our map also illustrated a high normalized abundance of ARGs in comparatively high latitudes, such as northern Europe and New Zealand, which were consistent with the latitudinal gradient showed by our observations (fig. S1E). We applied coefficient of

variation to quantify the uncertainty of our estimates, which resulted in a relatively high uncertainty in Siberia, Sahara, northern Canada, and Central Asia (fig. S9), despite the robust performance of our models (fig. S8 and table S16).

DISCUSSION

Soil antibiotic resistance structured by multiple biotic and abiotic factors is highly complex. To shed light on some of these understandings, we have compiled 1088 soil metagenomic data to generate a high-resolution quantitative map of ARGs and to disentangle their driving mechanisms beyond it through integrating with 169 spatial covariates. These spatial constraints include anthropogenic activities, physicochemical properties, climatic variables, and land use, among which anthropogenic activities ranked the most important factors enriching soil ARGs (Fig. 4A). Our global map also revealed that the hot spots of soil ARGs are located across eastern United States, Western Europe, South Asia, and East China

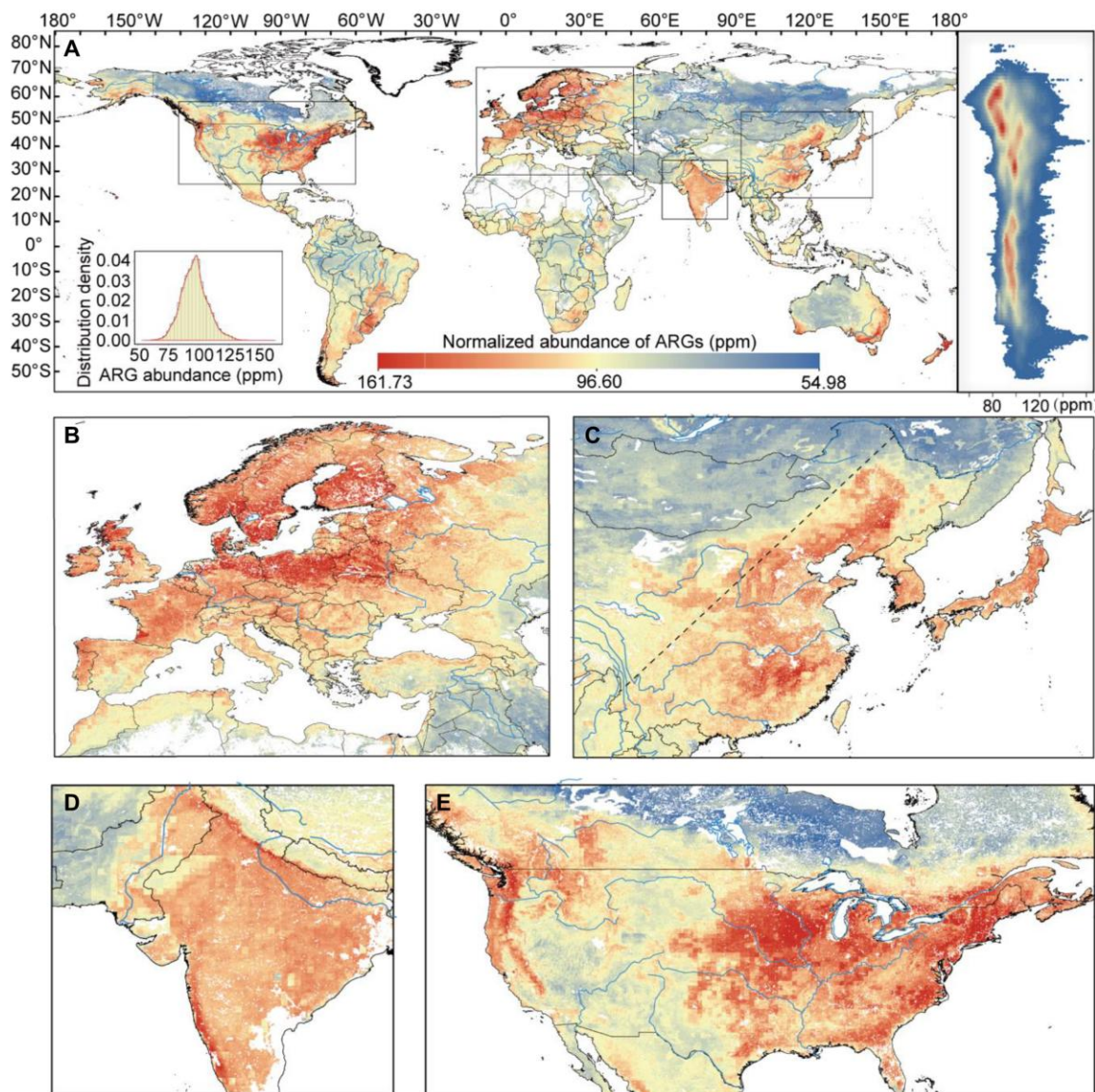


Fig. 5. Global map of the normalized abundance of soil ARGs. (A) Normalized abundance of soil ARGs across the world. The right subfigure depicts the latitudinal variation of ARG abundance across the world, and the left subfigure describes the distribution histogram of ARG abundance. (B to E) Normalized abundance of soil ARGs in Europe (B), East Asia (C), South Asia (D), and North America (E). The dash line in China is the well-known Chinese demographic “Hu Huanyong line” (25).

(Fig. 5), regions that are characterized with highly dense population (24, 25). Moreover, compared with nonagricultural habitats, a higher normalized abundance of ARGs was observed in agricultural habitats (Fig. 1B) that suffer from more extensive human activities. These results—together with noticeable positive connections of ARGs with livestock and crop production, irrigation and manure, agriculture and pesticide, as well as barley and sheep production (Fig. 4, B and C)—provide converging evidence that elevated soil antibiotic resistance is primarily attributed to anthropogenic activities.

One possible mechanism underlying the strong contribution of human activities to soil antibiotic resistance is the wide propagation of anthropogenic microbiomes, which is confirmed by the taxonomic composition of identified ARG hosts (Figs. 2C and 3). The ARG microbial hosts are dominated by potential pathogens (such as *E. coli*, *P. aeruginosa*, and *K. pneumoniae*; Fig. 3 and table S12) that

inhabit in clinical settings and animal gut environments (26, 27). A previous study (21), integrating 484 metagenomes across varied habitats, also demonstrated that ARG abundance in anthropogenically affected environments is largely explained by gut microbes from fecal pollution. Similar results have also been underpinned by regional field observations across the Yangtze River Delta (28), Amazon rainforest (17), Chinese croplands (18), and south-central Idaho (29), where soil ARGs have been pronouncedly enriched by wastewater irrigation and animal manure application. Apart from the mechanism driven by anthropogenic microbiomes carrying ARGs, another feasible explanation is suggested by associations between pesticides and the normalized abundance of ARGs (Fig. 4C). This relationship implies that soil ARGs are likely enriched by anthropogenic selective agents. The mechanism is confirmed by mounting control experiments and field monitoring,

where soil ARGs were enriched by antibiotics (30), arsenic (31), fungicide (32), mercury (20), cuprum (20), and polychlorinated biphenyls (33) due to wide coselection between antibiotic and metal resistance genes (20, 33). Such a coselection is partly because genes conferring resistance to antibiotics and other contaminants are carried by the same MGEs (20, 33). Besides, it would be driven by cross-resistance in which some efflux pump genes, such as *acrF* and *adeA*, can excrete both antibiotics and other pollutants (20, 33). Some contaminants are supplemented to livestock feed, thereby resulting in elevated residual concentrations in animal manure and soils (30), while others, such as pesticides preventing crops from invasive infections and disease outbreaks, are directly imported into farmland (34). A recent study (35) further demonstrated that Gammaproteobacteria, the dominant microbial class carrying ARGs we observed (Figs. 2C and 3), was potential responders to soil contaminants. Anthropogenic pollutants, such as oxytetracycline and azoxystrobin, would enrich Gammaproteobacteria, thus indirectly raising the normalized abundance of ARGs (35).

Our global map revealed a comparatively high normalized abundance of ARGs in New Zealand and northern Europe (Fig. 5). Soils in these regions are characterized by relatively abundant organic carbon, phosphorus, and nitrogen contents (36). It is demonstrated by our structural equation model that presented the controls of soil nutrients to ARG microbial hosts (Fig. 4D). These results, combined with the uptrend of ARGs with higher soil nutrients observed in our analyses (Fig. 4C), suggest that increased nutrients would provide carbon and energy sources for the growth and proliferation of microbes carrying ARGs. Although soil organic carbon, nitrogen, and phosphorus also provide carbon and energy sources for microbes without ARGs, these microbes are largely inhibited when exposed to antibiotics and other anthropogenic pollutants (13, 18). Therefore, nutrients would exclusively provide carbon and energy sources for microbes carrying ARGs, leading to elevated normalized abundance of soil ARGs. This suggestion is also supported by regional observations across the Yangtze River Delta (28), Chinese forest (37), and North China grassland (38), where the contents of organic carbon, nitrogen, phosphorus, and potassium markedly controlled either soil ARG abundance, richness, or diversity. These results suggest that soil ARGs and their carriers at the global scale are subjected to environmental filtering, an essential deterministic process that shapes the composition, abundance, diversity, and function of microbial communities (13, 23).

Observed soil ARGs were also strongly limited by climatic variables (Fig. 4). Temperature and precipitation appear to serve as contrasting effects on ARG abundance, that is, higher normalized abundance of ARGs resulting from lower temperature but increased precipitation (Fig. 4, B and C). One feasible mechanism underlying such a result is that low temperature and high soil moisture restrict decomposition rate, causing the accumulation of soil organic matter that are hospitable for the growth and proliferation of microbes carrying ARGs. Considering the wide occurrence of antibiotics and other anthropogenic pollutants that inhibits microbes susceptible to antibiotics (18), the effect of temperature and precipitation via soil carbon would exclusively affect antibiotic resistance microbes. Notably, our structural equation model also exhibited the indirect impacts of mean annual temperature and annual precipitation on ARGs via soil nutrients (Fig. 4D), which would validate our proposed hypothesis. Another possible explanation behind this relationship is that elevated temperature would enhance the

degradation and volatilization of selective agents, thereby resulting in lower selective pressure toward microbial communities and reduced ARG abundance. Previous studies (39, 40) also suggested that increasing temperature contributed to decreased chlorpyrifos, polycyclic aromatic hydrocarbons, perfluoroalkyl acids, and personal care products through facilitating their partition process and degradation. Besides, another important hypothesis is that soil ARGs in cold regions would be more sensitive to anthropogenic activities, compared with that in hot regions, owing to a lower microbial biomass at a low temperature. It means a small amount input of pathogens and gut microbes likely allow a high normalized abundance of ARGs. These possible mechanisms would be partly responsible for soil ARG hot spots in high latitudes, particularly those located in New Zealand and northern Europe (Fig. 5). Regional studies (41, 42) appeared to reveal more complex circumstances, where some indicated a negative correlation between temperature and ARG abundance, while others supported the opposite trend. It suggests that the relationship between climatic parameters and ARGs is very likely masked by local deviations, stressing the necessity of this study that investigated the biogeography of soil ARGs at the global scale. Together, however soil ARGs respond to climatic variables highlights that ARGs burden can be altered by soil carbon stocks and human-induced climate change. Further efforts should be devoted to performing long-term monitoring to disentangle the relationship between ARGs and climates and to further project soil ARGs into the future under different climate mitigation and socioeconomic scenarios.

Despite a relatively high confidence and robust prediction of the normalized abundance of soil ARGs in our study (fig. S8, A and B, and table S16), further efforts are needed to improve our estimates and to better characterize the response of ARGs to ecological variables at the global scale. Our models exhibited comparatively high predictive uncertainties of ARGs in some data-poor regions, including Sahara Desert, Siberia, central Australia, and Central Asia (fig. S9), mainly because regions with low sampling density (fig. S1A) cannot be as well trained as those with plentiful observations. Future regional observations in these data-poor regions are needed to optimize our surveillance, although the most existing global studies on microbiology at the global scale (13, 23, 43) presented limited samples in these regions, owing probably to poor transportations, harsh climates, and rugged terrains. Moreover, while this study investigates the global patterns of ARGs based on normalized abundance, further studies that combine absolute and normalized abundances might be more informative to disentangle the underlying mechanisms controlling soil ARGs. Last, although the current metagenomic sequencing data have provided some valuable insights as to the occurrence of soil ARGs, we will have a better understanding of how soil ARGs play a role in a situ environment based on metatranscriptome and culturomics analyses (44, 45).

In conclusion, our study generates spatially explicit understandings of soil antibiotic resistance and elucidates their potential microbial mechanisms by combining metagenomic sequence from public databases with environmental constraints. Our results not only provide baseline information of the soil ARGs at the global scale but also serve as a stepping stone to facilitate future modeling efforts under changing climate and anthropogenic scenarios. Furthermore, our machine learning approaches to predict the normalized abundance of ARGs shed light on mapping the large-scale

distribution of other essential functional genes, such as *mcrA*, *cbbM*, and *hcd*, that regulate biogeochemical cycles, greenhouse gases emission, and climate change. Moreover, we disentangle the response of ARGs to ecological constraints, particularly providing evidence that soil ARGs is driven by soil nutrients and bioclimatic variables via ARG microbial hosts, laying the foundation of developing mechanistic models that integrate physical, chemical, and biological processes. Last, our study suggests some constructive policy strategies to combat soil ARGs burden in a One Health approach, including controlling antibiotic misuse and overuse in animal husbandry, decoupling agricultural production and livestock, reducing untreated wastewater irrigation and pesticide overuse, as well as improving sanitation infrastructure, which would aid the achievement of Sustainable Development Goals.

MATERIALS AND METHODS

Metagenomic sample collection and quality control

We collected 1088 soil metagenomes in this study, all retrieved from European Nucleotide Archive (ENA; <https://ebi.ac.uk/ena/>) and National Center for Biotechnology Information (NCBI) Sequence Read Archive (<https://ncbi.nlm.nih.gov/sra/>). We applied the following metagenome selection criterion procedures to minimize possible bias: (i) Plant-associated sample, such as rhizosphere and rhizoplane soils, were excluded; (ii) soil samples were not collected from potentially heavily contaminated environments (such as pharmaceutical manufacturing parks, hospitals, livestock farms, and coal-fired power plants), where high levels of pharmaceuticals, heavy metals, and persistent organic pollutants would select potential antimicrobial strain to raise the uncertainty of our datasets; (iii) soil samples were not cultured and without additional experiment processes, such as nitrogen addition and warming after sampling; (iv) we only considered control group and excluded treatment group for those conducting controlled trials; (v) sequencing reads were generated by Illumina shotgun platforms, excluding those by Roche 454 and ABI SOLiD sequencing technologies; (vi) only paired-end sequencing reads with FASTQ format were included, and single-end sequences were excluded; (vii) the average read length of metagenomes exceeded 100 base pairs; and (viii) accurate coordinate information and habitats were available for spatial analyses. These quality control procedures were aimed to minimize possible uncertainty from root microenvironments, sampling locations, experimental processes, unexpected contaminations, sequence platforms, sequencing methods, and read lengths. After these filtering procedures, a final set of 1088 metagenomic sequencing datasets (430 locations) were retained for further analysis. We summarized detailed metagenome information on accession number, read length, base count, habitats, continents, country, and spatial coordinates in table S1. All the downloaded raw metagenomic sequences was processed for quality check and filtration to acquire clean reads using Trimmomatic-0.36 (46) (ILLUMINA-CLIP: adapters. fa: 2:30:10 SLIDINGWINDOW: 4:15 MINLEN:100).

ARG, MGE, and taxonomic annotation

ARG profiles in the collected metagenomic datasets were quantified using ARG annotation pipeline ARGs-OAP 2.3 (<http://smile.hku.hk/SARGs/>) through two-stage pipelines according to the protocol provided by previous publications (47, 48). In stage one, we put a

metadata file, and all the clean reads files into one directory on our local Linux system, where the stage one Perl script was executed to screen ARG-like and 16S ribosomal RNA (rRNA) gene sequences. In stage two, the candidate ARG sequences and a metadata online file were input to align against SARG reference databases by executing the stage two Perl script on our local Linux system. The parameters were set at 25 amino acids of alignment length, 80% of similarity, and 1×10^{-5} of *e*-value (47, 48). The stage two pipeline were expected to output a file that contained the abundances of 24 types and 1244 subtypes of ARGs for each metagenomic samples. ARG abundances are normalized to the number of cells (unit: ARG copies per cell), total 16S reads (unit: copies of ARG per copy of 16S rRNA), and ppm (reads carrying ARGs per million reads), a unit that is normalized by which total reads carrying ARGs divide per million of read counts. Following the same procedure and parameters with ARG annotation, MGEs were quantified through aligning against a comprehensive MGE database (49) that consists of 2706 genes, including transposase, insertion sequences, integrase, *istA*, *istB*, *qacEdelta*, *tniA*, *tniB*, and plasmids. Kraken (v2.1.1), along with the custom *k*-mers Bracken databases (<https://benlangmead.github.io/aws-indexes/k2>), was used to conduct taxonomic annotation. Furthermore, clean reads that carried MGEs and ARGs were selected to align the custom *k*-mers Bracken databases to identify the microbial hosts of MGEs and ARGs. For the sake of comparability, we used ppm as the unit to normalize ARGs, MGEs, and microbes.

Acquisition of gridded spatial covariates

To generate global predictive models of soil ARG abundance, we prepared the global maps of 169 spatial covariates that consisted of climatic variables, land use, physicochemical properties, and anthropogenic factors (table S2) from public databases or satellite observations. Climatic variables that control soil microbial communities were potential drivers for the abundance and diversity of ARGs (13, 50). Our climatic variable sets contained 19 core bioclimatic variables and 16 extended climatic variables from WorldClim (<https://worldclim.org/data/worldclim21.html>) and CliMond (<https://climond.org/BioclimRegistry.aspx#BioclimFAQ>) databases. Soil properties that would control the nutrient availability of antibiotic resistance microbes (13, 37) were derived from SoilGrids (<https://soilgrids.org/>) and EarthData (https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1223; https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1264) databases. Anthropogenic activity was a quite comprehensive factor that could not be reflected by any one single index. We collected multiple variables, including livestock, population density, human development index, travel time, crop yields, and pesticide use, to represent anthropogenic activities and to reduce potential uncertainty as possible. Livestock (buffalo, goat, cattle, horse, chicken, pig, ducks, and sheep) that likely represent manure and antibiotic pollution from domestic animals were attained from <http://fao.org/livestock-systems/global-distributions/en/>. Travel time to cities and ports were collected from CGIAR-CSI (<https://cgiarcsi.community/category/data/>). Population density was from Google Earth Engine (<https://explorer.earthengine.google.com/#detail/CIESIN%2FGPWv4%2Funwpp-adjusted.population-density>). Human development index was acquired from Dryad (<https://datadryad.org/stash/dataset/doi:10.5061/dryad.dk1j0>). Human influence index, development threat index, human modification of terrestrial systems, and pesticide

use (chlorpyrifos, glufosinate, and glyphosate used in soybean or corn) were available from EarthData (<https://beta.sedac.ciesin.columbia.edu/search/data?>). Crop yields (wheat, rice, maize, barley, sorghum, pearl, small, soybean, and tea yields) that possibly denote the impact of agricultural antibiotics and manure were collected from CGIAR-CSI (<https://cgiarcsi.community/2019/01/04/global-spatially-disaggregated-crop-production-statistics-data-for-2010/>). Antibiotic use in food animals was available in (51). Other variables, such as land use, vegetation, and topography, would shape antibiotic resistome through receiving different levels of antibiotic residues or affecting soil properties. Land use datasets stemmed from a recently released database that provided land use datasets across the world from 2015 to 2100 (52). Vegetative indicators were retrieved from EarthENV (<http://earthenv.org/texture>). Topographic variables were obtained from the U.S. Geological Survey (<https://pubs.usgs.gov/of/2011/1073/>). We resampled all the datasets to match a 0.083° (around 10×10 km at the equator) of resolution.

Machine learning algorithms

We input the prepared gridded spatial covariates into four candidate machine learning algorithms (support vector machine, k -nearest neighbor, gradient boosting decision tree, and random forest) to predict the normalized abundance of ARGs. Support vector machine (53) projects the input features into a feature space in a much higher dimension that enables linearly inseparable features separable. The objective of this algorithm is to find a hyperplane using kernel functions and get an optimum solution through iteratively adjusting the hyperplane. K -nearest neighbor algorithm (54) assumes that observations with similar predictors have a similar response. This model assigns an average response of the k -closest observations to another observation. Random forest (22) and gradient boosting tree (55) are ensemble learning models based on decision trees that compose of intermediate nodes and leaf nodes. The intermediate nodes with outgoing edges are labeled by conditions that are determined on the basis of information gain, information gain rate, or Gini coefficient. The leaf nodes without outgoing edges are labeled by decisions or actions. This algorithm assigns a mean value of samples within a trained leaf node to another observation. Bagged tree (54) randomly divides subsets from all the training datasets, and each subsets construct a decision tree. The average response of all decision trees is assigned to the predictive value of another observation. Random forest (22), which randomly selects both datasets and attributes could be regarded as a unique bagging tree algorithm. Gradient boosting tree construct an initial decision tree based on all the attributes and samples, and the negative gradient of the loss function is used as the residual approximation in the initial decision tree to fit into the next decision tree. This process works iteratively until the residual errors or the tree number reach a given value. Bagged tree and random forest are designed to minimize the variance of models, while gradient boosting tree aims to minimize the bias of models.

The applications of these models were preceded by recursive feature elimination algorithm to identify their optimal independent variables (fig. S6). Afterward, we conducted hyperparameter tuning for four algorithms with their optimal independent features using grid search, to determine the best hyperparameter combinations (fig. S7). The feature selection and parameter tuning procedures were performed on the basis of k -fold cross-validation (54), which

is an important data training tool to enable the test sets independent of training sets and to minimize model overfitting problem. A 10-fold cross-validation randomly splits the training sets into 10 equal-sized subsets. Nine of these subsets were combined into a dataset for model training, and the rest subset was regarded as test sets to estimate the model performances. The R^2 of each model prediction was stored, and the process was repeated 10 times using the other training and test subset combinations. The R^2 of model estimations was assigned to the final cross-validation score, and we selected the model with the highest R^2 as the best hyperparameter combination. Ultimately, random forest outperformed other algorithms in predicting the normalized abundance of ARGs, with the 10-fold cross-validation R^2 of 0.47 (fig. S7 and table S15). We trained 10 separate random forests by setting 10 different random seeds and averaged 10 predictive outputs as our final estimates. We calculated the coefficient of variation of 10 random forest predictions to estimate our model uncertainty.

To avoid our model projects far outside their training dataset, we excluded those pixels with potential large uncertainty (43, 56, 57), in which (i) land use was water bodies and urban; (ii) soil properties data collected from SoilGrids database was missing; and (iii) grid cells fell outside the value range of clay content, sand content, soil carbon density, total nitrogen, annual mean temperature, annual precipitation, population density, pig density, chicken density, and cattle density. This procedure eliminated 1,570,866 pixels, most of which are located in data-poor regions, such as Sahara Desert, Siberia, and North Canada. We also validated our model using an independent metagenomic dataset that resulted in a validated R^2 of 0.33 (table S17 and fig. S8B). We further summarized the prediction performances of previous microbial geography studies, showing that predictive model performance increased largely on sample size and/or the number of predictor variables (table S16). Our estimates outperformed most previous studies owing probably to a relatively large sample size and number of predictor variables (table S16).

To estimate the minimal sample size requirement, we examined whether our datasets follow normal distribution using a histogram and one-sample Kolmogorov-Smirnov test (K-S test) based on R stats package, where $P > 0.05$ indicates a significant result according to its null hypothesis. The resulting histogram and K-S test result showed that the ln-transformed ARG abundance of our training datasets significantly follows normal distribution ($P = 0.48$; fig. S8C). We further calculated the minimal sample size required to characterize soil resistome using the following equation (58)

$$n = \left(\frac{Z_{1-\alpha/2} \times \sigma}{\delta} \right)^2 \quad (1)$$

where n refers to the minimal sample size requirement; $Z_{1-\alpha/2}$ denotes the two-sided quantile of standard normal distribution, where 2.58 was set under 99% confidence level; δ represents distance from mean to limits that was set to 0.05; and σ indicates SD that was 0.29, calculated through

$$\sigma = \sqrt{\frac{\sum (X_i - \bar{X})^2}{N - 1}} \quad (2)$$

where N is the sample size of our datasets, \bar{X} indicates the average ln-transformed ARG abundance of all the samples, and X_i denotes

the ln-transformed ARG abundance of sample i . We estimated the minimal sample size requirement using PASS 15.0.5 software according to these equations, and the resulting minimal sample size requirement was 224 (table S18). Although additional 20% dropout rate was considered, the dropout-inflated minimal sample size requirement, 280 (table S18), was lower than our sample size (430 locations and 1088 samples), indicating that enough soil samples have been obtained.

Statistical analysis

Data analyses were mainly performed using R version 4.0.2 (R Foundation for Statistical Computing) with software packages. Recursive feature elimination algorithm, hyperparameter tuning, and uncertainty analysis were conducted using randomForest, mlbench, caret, e1071, and gbm packages. We mapped the global distribution of collected metagenomic datasets using ArcGIS 10.2 (ESRI, Berkeley, CA, USA), and the software was also used to visualize the normalized abundance of predicted soil ARGs. Bar chart with error bar was plotted using ggplot2 and ggpvr packages. Maptree (59) was used to characterize the taxonomic characteristics of soil microbes and ARG microbial hosts based on the ggraph, igrph, tidyverse, viridis, data.tree, phyloseq, ggtree packages, TreeMap (version 2019.8.1), and <https://github.com/1822318572/Note/tree/master/WenTao/191124Maptree>. ANOVA and PERMANOVA combined with principal components analysis (PCA) that differentiated the composition of ARGs among varied habitats and continents were conducted by vegan and ggplot2 packages. Histograms of the normalized abundance of ARGs, MGEs, and soil microbes were plotted on the basis of ggplot2 package. Box plot and violin plot with scatter were visualized by ggplot2 package. Scatter plots with fitting curves and red density shade were drawn by ggpvr and ggplot packages. LSD package was used to visualize the latitudinal pattern of the normalized abundance of ARGs based on kernel density estimation. Sankey diagram was used to show that the coexistence among ARGs, MGEs, and their microbial hosts through networkD3 package.

We performed rotated PCA (rPCA) (43) to increase the interpretability of environmental variables using IBM SPSS Statistics 25. Before the application of rPCA, all the variance inflation factor (VIF) of individual predictor variables was estimated. The variable with the maximal VIF was eliminated until the VIFs of all the predictor variables were lower than 10, which retained 57 independent variables (table 13). Variance maximizing rotation method was further used to minimize the potential multicollinearity. The optimal number of principal components was determined by Kaiser-Guttman rule, requiring that the eigenvalue of principal components should exceed 1. The resulting 16 principal components and their interpretations were listed in table S13. We assessed the relative importance of identified principal components through the variable importance tool in caret package. Briefly, this tool calculates the mean square error for every decision tree with out-of-bag estimates based on the random forest, which generates the relative importance for each predictor variables. The resulting relative importance was normalized on a scale of 0 to 100%, and the total relative importance was 100% (43). Partial dependence plots based on random forest to quantify the responses of ARGs to environmental constraints were performed through rfPermute, PDP, and vegan packages. We constructed a structural equation model to determine direct and indirect impacts of environmental principal components,

microbial communities, and MGEs on the normalized abundance of ARGs. A prior model (fig. S5) (13, 17, 19, 41) and maximum-likelihood estimation was used to estimate model parameters based on lavaan package. We applied modification index (13, 60) to optimize the structure of our model iteratively, where an ecologically sound path with ≥ 2 modification index is constructed in a stepwise way. We estimated path coefficients, significance, and fitting performance after excluding nonsignificant paths ($P > 0.05$). The fitting performance was examined using comparative fit index (cfi), root mean square error of estimation (RSME), standardized root square residual (SRMR), and chi-square test (χ^2), where $\text{cfi} \geq 0.90$, $\text{RSME} \leq 0.05$, $\text{SRMR} \leq 0.08$, $\chi^2 \leq 2$, and $P > 0.05$ indicate a good fit.

Supplementary Materials

This PDF file includes:

Figs. S1 to S9

Other Supplementary Material for this

manuscript includes the following:

Tables S1 to S18

[View/request a protocol for this paper from Bio-protocol.](#)

REFERENCES AND NOTES

1. A. H. Holmes, L. S. P. Moore, A. Sundsfjord, M. Steinbakk, S. Regmi, A. Karkey, P. J. Guerin, L. J. V. Piddock, Understanding the mechanisms and drivers of antimicrobial resistance. *Lancet* **387**, 176–187 (2016).
2. S. Hernando-Amado, T. M. Coque, F. Baquero, J. L. Martínez, Defining and combating antibiotic resistance from One Health and global health perspectives. *Nat. Microbiol.* **4**, 1432–1442 (2019).
3. H.-W. Hu, J.-T. Wang, B. K. Singh, Y.-R. Liu, Y.-L. Chen, Y.-J. Zhang, J.-Z. He, Diversity of herbaceous plants and bacterial communities regulates soil resistome across forest biomes. *Environ. Microbiol.* **20**, 3186–3200 (2018).
4. T. P. Robinson, D. P. Bu, J. Carrique-Mas, E. M. Fèvre, M. Gilbert, D. Grace, S. I. Hay, J. Jiwakanon, M. Kakkar, S. Kariuki, R. Laxminarayan, J. Lubroth, U. Magnusson, P. T. Ngoc, T. P. Van Boeckel, M. E. J. Woolhouse, Antibiotic resistance is the quintessential One Health issue. *Trans. R. Soc. Trop. Med. Hyg.* **110**, 377–380 (2016).
5. D. S. Perlin, R. Rautema-Richardson, A. Alastruey-Izquierdo, The global problem of antifungal resistance: Prevalence, mechanisms, and management. *Lancet Infect. Dis.* **17**, e383–e392 (2017).
6. F. Prestinaci, P. Pezzotti, A. Pantosti, Antimicrobial resistance: A global multifaceted phenomenon. *Pathog. Glob. Health.* **109**, 309–318 (2015).
7. J. O'Neill, Tackling drug-resistant infections globally: Final report and recommendations. *Rev. Antimicrob. Resist.*, 1–80 (2016).
8. Antimicrobial Resistance Collaborators, Global burden of bacterial antimicrobial resistance in 2019: A systematic analysis. *Lancet* **399**, 629–655 (2022).
9. WHO, Global Action Plan on Antimicrobial Resistance (World Health Organization 2015), pp. 1–19.
10. Y. G. Zhu, Y. Zhao, D. Zhu, M. Gillings, J. Penuelas, Y. S. Ok, A. Capon, S. Banwart, Soil biota, antimicrobial resistance and planetary health. *Environ. Int.* **131**, 105059 (2019).
11. J. Zeng, Y. Pan, J. Yang, M. Hou, Z. Zeng, W. Xiong, Metagenomic insights into the distribution of antibiotic resistome between the gut-associated environments and the pristine environments. *Environ. Int.* **126**, 346–354 (2019).
12. K. J. Forsberg, A. Reyes, B. Wang, E. M. Selleck, M. O. A. Sommer, G. Dantas, The shared antibiotic resistome of soil bacteria and human pathogens. *Science* **337**, 1107–1111 (2012).
13. M. Bahram, F. Hildebrand, S. K. Forslund, J. L. Anderson, N. A. Soudzilovskaia, P. M. Bodegom, J. Bengtsson-Palme, S. Anslan, L. P. Coelho, H. Harend, J. Huerta-Cepas, M. H. Medema, M. R. Maltz, S. Munda, P. A. Olsson, M. Pent, S. Pölme, S. Sunagawa, M. Ryberg, L. Tedersoo, P. Bork, Structure and function of the global topsoil microbiome. *Nature* **560**, 233–237 (2018).
14. M. Delgado-Baquerizo, C. A. Guerra, C. Cano-Díaz, E. Egidi, J. T. Wang, N. Eisenhauer, B. K. Singh, F. T. Maestre, The proportion of soil-borne pathogens increases with warming at the global scale. *Nat. Clim. Chang.* **10**, 550–554 (2020).

15. J. R. Rohr, C. B. Barrett, D. J. Civitello, M. E. Craft, B. Delius, G. A. DeLeo, P. J. Hudson, N. Jouanard, K. H. Nguyen, R. S. Ostfeld, J. V. Remais, G. Riveau, S. H. Sokolow, D. Tilman, Emerging human infectious diseases and the links to global food production. *Nat. Sustain.* **2**, 445–456 (2019).
16. Y.-J. Zhang, H. W. Hu, Q. L. Chen, B. K. Singh, H. Yan, D. Chen, J.-Z. He, Transfer of antibiotic resistance from manure-amended soils to vegetable microbiomes. *Environ. Int.* **130**, 104912 (2019).
17. L. N. Lemos, A. Pedrinho, A. T. R. de Vasconcelos, S. M. Tsai, L. W. Mendes, Amazon deforestation enriches antibiotic resistance genes. *Soil Biol. Biochem.* **153**, 108110 (2021).
18. S. Du, J. P. Shen, H. W. Hu, J. T. Wang, L. L. Han, R. Sheng, W. X. Wei, Y. T. Fang, Y. G. Zhu, L. M. Zhang, J. Z. He, Large-scale patterns of soil antibiotic resistome in Chinese croplands. *Sci. Total Environ.* **712**, 136418 (2020).
19. J. Zhang, Q. Sui, J. Tong, H. Zhong, Y. Wang, M. Chen, Y. Wei, Soil types influence the fate of antibiotic-resistant bacteria and antibiotic resistance genes following the land application of sludge composts. *Environ. Int.* **118**, 34–43 (2018).
20. Y. Zhao, T. Cocerva, S. Cox, S. Tardif, J. Q. Su, Y. G. Zhu, K. K. Brandt, Evidence for co-selection of antibiotic resistance genes and mobile genetic elements in metal polluted urban soils. *Sci. Total Environ.* **656**, 512–520 (2019).
21. A. Karkman, K. Pärnänen, D. G. J. Larsson, Fecal pollution can explain antibiotic resistance gene abundances in anthropogenically impacted environments. *Nat. Commun.* **10**, 80 (2019).
22. J. van den Hoogen, S. Geisen, D. Routh, H. Ferris, W. Trautspurger, D. A. Wardle, R. G. M. de Goede, B. J. Adams, W. Ahmad, W. S. Andriuzzi, R. D. Bardgett, M. Bonkowski, R. Campos-Herrera, J. E. Cares, T. Caruso, L. de Brito Caixeta, X. Chen, S. R. Costa, R. Creamer, J. M. da Cunha Castro, M. Dam, D. Djigal, M. Escuer, B. S. Griffiths, C. Gutiérrez, K. Hoberg, D. Kalinkina, P. Kardol, A. Kergunteuil, G. Korthals, V. Krashevska, A. A. Kudrin, Q. Li, W. Liang, M. Magilton, M. Marais, J. A. R. Martin, E. Matveeva, E. H. Mayad, C. Mulder, P. Mullin, R. Neilson, T. A. D. Nguyen, U. N. Nielsen, H. Okada, J. E. P. Rius, K. Pan, V. Peneva, L. Pellissier, J. C. P. da Silva, C. Pitteloud, T. O. Powers, K. Powers, C. W. Quist, S. Rasmann, S. S. Moreno, S. Scheu, H. Setälä, A. Sushchuk, A. V. Tiunov, J. Trap, W. van der Putten, M. Vestergård, C. Villenave, L. Waeyenberge, D. H. Wall, R. Wilschut, D. G. Wright, J. in Yang, T. W. Crowther, Soil nematode abundance and functional group composition at a global scale. *Nature* **572**, 194–198 (2019).
23. M. Delgado-Baquerizo, A. M. Oliverio, T. E. Brewer, A. Benavent-gonzález, D. J. Eldridge, R. D. Bardgett, F. T. Maestre, B. K. Singh, N. Fierer, A global atlas of the dominant bacteria found in soil. *Science* **359**, 320–325 (2018).
24. E. Doherty-Whitfield, K. MacManus, S. B. Adamo, L. Pistolesi, J. Squires, O. Borkovska, S. R. Baptista, Taking advantage of the improved availability of census data: A first look at the gridded population of the world, version 4. *Pap. Appl. Geogr.* **1**, 226–234 (2015).
25. J. Q. Su, X. L. An, B. Li, Q. L. Chen, M. R. Gillings, H. Chen, T. Zhang, Y. G. Zhu, Metagenomics of urban sewage identifies an extensively shared antibiotic resistome in China. *Microbiome* **5**, 84 (2017).
26. K. R. Chng, T. S. Ghosh, Y. H. Tan, T. Nandi, I. R. Lee, A. Hui, Q. Ng, C. Li, A. Ravikrishnan, K. M. Lim, D. Lye, T. Barkham, K. Raman, S. L. Chen, L. Chai, B. Young, Y. Gan, N. Nagarajan, Metagenome-wide association analysis identifies microbial determinants of post-antibiotic ecological recovery in the gut. *Nat. Ecol. Evol.* **4**, 1256–1267 (2020).
27. M. Subbiah, M. A. Caudell, C. Mair, M. A. Davis, L. Matthews, R. J. Quinlan, M. B. Quinlan, B. Lyimo, J. Buza, J. Keyyu, D. R. Call, Antimicrobial resistant enteric bacteria are widely distributed amongst people, animals and the environment in Tanzania. *Nat. Commun.* **11**, 228 (2020).
28. J. Sun, L. Jin, T. He, Z. Wei, X. Liu, L. Zhu, X. Li, Antibiotic resistance genes (ARGs) in agricultural soils from the Yangtze River delta, China. *Sci. Total Environ.* **740**, 140001 (2020).
29. R. S. Dungan, C. A. Strausbaugh, A. B. Leytem, Survey of selected antibiotic resistance genes in agricultural and non-agricultural soils in south-central Idaho. *FEMS Microbiol. Ecol.* **95**, f1z071 (2019).
30. X. Tang, C. Lou, S. Wang, Y. Lu, M. Liu, M. Z. Hashmi, X. Liang, Z. Li, Y. Liao, W. Qin, F. Fan, J. Xu, P. C. Brookes, Effects of long-term manure applications on the occurrence of antibiotics and antibiotic resistance genes (ARGs) in paddy soils: Evidence from four field experiments in south of China. *Soil Biol. Biochem.* **90**, 179–187 (2015).
31. X. Zhao, J. P. Shen, L. M. Zhang, S. Du, H. W. Hu, J. Z. He, Arsenic and cadmium as predominant factors shaping the distribution patterns of antibiotic resistance genes in polluted paddy soils. *J. Hazard. Mater.* **389**, 121838 (2020).
32. H. Zhang, S. Chen, Q. Zhang, Z. Long, Y. Yu, H. Fang, Fungicides enhanced the abundance of antibiotic resistance genes in greenhouse soil. *Environ. Pollut.* **259**, 113877 (2020).
33. A. V. Gorovtsov, I. S. Sazykin, M. A. Sazykina, The influence of heavy metals, polycyclic aromatic hydrocarbons, and polychlorinated biphenyls pollution on the development of antibiotic resistance in soils. *Environ. Sci. Pollut. Res.* **25**, 9283–9292 (2018).
34. F. H. M. Tang, M. Lenzen, A. McBratney, F. Maggi, Risk of pesticide pollution at the global scale. *Nat. Geosci.* **14**, 206–210 (2021).
35. Q. Zhang, Z. Zhang, T. Lu, Y. Yu, J. Penuelas, Y. G. Zhu, H. Qian, Gammaproteobacteria, a core taxon in the guts of soil fauna, are potential responders to environmental concentrations of soil pollutants. *Microbiome* **9**, 196 (2021).
36. T. Hengl, J. M. De Jesus, G. B. M. Heuvelink, M. R. Gonzalez, M. Kilibarda, A. Blagotić, W. Shangquan, M. N. Wright, X. Geng, B. Bauer-Marschallinger, M. A. Guevara, R. Vargas, R. A. MacMillan, N. H. Batjes, J. G. B. Leenaars, E. Ribeiro, I. Wheeler, S. Mantel, B. Kempen, SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE* **12**, e0169748 (2017).
37. M. Song, D. Song, L. Jiang, D. Zhang, Y. Sun, G. Chen, H. Xu, W. Mei, Y. Li, C. Luo, G. Zhang, Large-scale biogeographical patterns of antibiotic resistome in the forest soils across China. *J. Hazard. Mater.* **403**, 123990 (2021).
38. Z. Zheng, L. Li, T. P. Makhalanyane, C. Xu, K. Li, K. Xue, C. Xu, R. Qian, B. Zhang, J. Du, H. Yu, X. Cui, Y. Wang, Y. Hao, The composition of antibiotic resistance genes is not affected by grazing but is determined by microorganisms in grassland soils. *Sci. Total Environ.* **761**, 143205 (2021).
39. Y. Lu, P. Wang, C. Wang, M. Zhang, X. Cao, C. Chen, C. Wang, C. Xiu, D. Du, H. Cui, X. Li, W. Qin, Y. Zhang, Y. Wang, A. Zhang, M. Yu, R. Mao, S. Song, A. C. Johnson, X. Shao, X. Zhou, T. Wang, R. Liang, C. Su, X. Zheng, S. Zhang, X. Lu, Y. Chen, Y. Zhang, Q. Li, K. Ono, N. C. Stenseth, M. Visbeck, V. Ittekkot, Multiple pollutants stress the coastal ecosystem with climate and anthropogenic drivers. *J. Hazard. Mater.* **424**, 127570 (2022).
40. V. Delnat, J. Verborgt, L. Janssens, R. Stoks, Daily temperature variation lowers the lethal and sublethal impact of a pesticide pulse due to a higher degradation rate. *Chemosphere* **263**, 128114 (2021).
41. T. K. Dunivin, A. Shade, Community structure explains antibiotic resistance gene dynamics over a temperature gradient in soil. *FEMS Microbiol. Ecol.* **94**, f1y016 (2018).
42. Y. Zhou, L. Niu, S. Zhu, H. Lu, W. Liu, Occurrence, abundance, and distribution of sulfonamide and tetracycline resistance genes in agricultural soils across China. *Sci. Total Environ.* **599–600**, 1977–1983 (2017).
43. D. Haaf, J. Six, S. Doetterl, Global patterns of geo-ecological controls on the response of soil respiration to warming. *Nat. Clim. Chang.* **11**, 623–627 (2021).
44. J. C. Lagier, G. Dubourg, M. Million, F. Cadoret, M. Bilen, F. Fenollar, A. Levasseur, J. M. Rolain, P. E. Fournier, D. Raoult, Culturing the human microbiota and culturomics. *Nat. Rev. Microbiol.* **16**, 540–550 (2018).
45. V. Lázár, A. Martins, R. Spohn, L. Daruka, G. Grézal, G. Fekete, M. Számel, P. K. Jangir, B. Kintsés, B. Csörgő, Á. Nyerges, Á. Györkei, A. Kincses, A. Dér, F. R. Walter, M. A. Deli, E. Urbán, Z. Hegedus, G. Olajos, O. Méhi, B. Bálint, I. Nagy, T. A. Martinek, B. Papp, C. Pál, Antibiotic-resistant bacteria show widespread collateral sensitivity to antimicrobial peptides. *Nat. Microbiol.* **3**, 718–731 (2018).
46. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
47. X. Yin, X. Jiang, B. Chai, L. Li, Y. Yang, J. R. Cole, J. M. Tiedje, T. Zhang, 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 (2018).
48. X. Yin, Y. Deng, L. Ma, Y. Wang, L. Y. L. Chan, T. Zhang, Exploration of the antibiotic resistome in a wastewater treatment plant by a nine-year longitudinal metagenomic study. *Environ. Int.* **133**, 105270 (2019).
49. K. Pärnänen, A. Karkman, J. Hultman, C. Lyra, J. Bengtsson-Palme, D. G. J. Larsson, S. Rautava, E. Isolauri, S. Salminen, H. Kumar, R. Satokari, M. Virta, Maternal gut and breast milk microbiota affect infant gut antibiotic resistome and mobile genetic elements. *Nat. Commun.* **9**, 3891 (2018).
50. Y. Yang, G. Liu, C. Ye, W. Liu, Bacterial community and climate change implication affected the diversity and abundance of antibiotic resistance genes in wetlands on the Qinghai-Tibetan plateau. *J. Hazard. Mater.* **361**, 283–293 (2019).
51. T. P. Van Boeckel, C. Brower, M. Gilbert, B. T. Grenfell, S. A. Levin, T. P. Robinson, A. Teillant, R. Laxminarayan, Global trends in antimicrobial use in food animals. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 5649–5654 (2015).
52. M. Chen, C. R. Vernon, N. T. Graham, M. Hejazi, M. Huang, Y. Cheng, K. Calvin, Global land use for 2015–2100 at 0.05° resolution under diverse socioeconomic and climate scenarios. *Sci. Data.* **7**, 320 (2020).
53. L. Wang, F. Long, W. Liao, H. Liu, Prediction of anaerobic digestion performance and identification of critical operational parameters using machine learning algorithms. *Bio-resour. Technol.* **298**, 122495 (2020).
54. C. E. Reid, M. Jerrett, M. L. Petersen, G. G. Pfister, P. E. Morefield, I. B. Tager, S. M. Raffuse, J. R. Balmes, Spatiotemporal prediction of fine particulate matter during the 2008 Northern California wildfires using machine learning. *Environ. Sci. Technol.* **49**, 3887–3896 (2015).
55. Y. Zhang, B. Pan, S. K. Lam, E. Bai, P. Hou, D. Chen, Predicting the ratio of nitrification to immobilization to reflect the potential risk of nitrogen loss worldwide. *Environ. Sci. Technol.* **55**, 7721–7730 (2021).

56. N. Huang, L. Wang, X. P. Song, T. Andrew Black, R. S. Jassal, R. B. Myneni, C. Wu, L. Wang, W. Song, D. Ji, S. Yu, Z. Niu, Spatial and temporal variations in global soil respiration and their relationships with climate and land cover. *Sci. Adv.* **6**, eabb8508 (2020).
57. B. S. Steidinger, T. W. Crowther, J. Liang, M. E. Van Nuland, G. D. A. Werner, P. B. Reich, G. Nabuurs, S. de-Miguel, M. Zhou, N. Picard, B. Herault, X. Zhao, C. Zhang, D. Routh, K. G. Peay; GFBI consortium, Climatic controls of decomposition drive the global biogeography of forest-tree symbioses. *Nature* **569**, 404–408 (2019).
58. T. P. Ryan, *Sample Size Determination and Power* (John Wiley & Sons, 2013), pp. 58–64.
59. V. J. Carrión, J. Perez-Jaramillo, V. Cordovez, V. Tracanna, M. De Hollander, D. Ruiz-Buck, L. W. Mendes, W. F. J. van Ijcken, R. Gomez-Exposito, S. S. Elsayed, P. Mohanraju, A. Arifah, J. van der Oost, J. N. Paulson, R. Mendes, G. P. van Wezel, M. H. Medema, J. M. Raaijmakers, Pathogen-induced activation of disease-suppressive functions in the endophytic root microbiome. *Science* **366**, 606–612 (2019).
60. Y. Rosseel, lavaan: An R package for structural equation modeling and more version 0.5-12 (BETA). *J. Stat. Softw.* **48**, 1–36 (2012).

Acknowledgments: We appreciate anonymous reviewers for giving us valuable suggestions to improve this study. We are grateful to all those who provided raw data for this paper. **Funding:**

This research was funded by the National Natural Science Foundation of China (41730646, 2016YFE0133700, 42030411, 41725002, and 41501524). **Author contributions:** D.Z., G.Y., and M.L. designed this research. D.Z. contributed to investigation, data curation, and original draft writing. G.Y. performed the conceptualization, manuscript reviewing, and language modification. M.L. and L.H. provided the supervision and funding acquisition. T.P.V.B. contributed to data resource. Y.Y., Y.Z., and Y.L. supported the computation resource and software. All authors edited this study. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Raw metagenomics data are publicly available in the ENA (<https://ebi.ac.uk/ena/>) and NCBI (<https://ncbi.nlm.nih.gov/sra/>) databases. All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 2 May 2022

Accepted 20 October 2022

Published 16 November 2022

10.1126/sciadv.abq8015