Full length article

A trait-based ecological perspective on the soil microbial antibiotic-related genetic machinery

Olesya Dulya [a,b,*], Vladimir Mikryukov [a,b], Daniil V. Shchepkin [b], Mari Pent [a], Heidi Tamm [a],
Massimo Guazzini [c], Panos Panagos [d], Arwyn Jones [d], Alberto Orgiazzi [d,e], Fabio Marroni [c],
Mohammad Bahram [a,f,g], Leho Tedersoo [b]

[a] Institute of Ecology and Earth Sciences, University of Tartu, Tartu 50409, Estonia
[b] Center of Mycology and Microbiology, University of Tartu, Tartu 50409, Estonia
[c] Department of Agriculture, Food, Environmental and Animal Sciences, University of Udine, Udine 33100, Italy
[d] European Commission, Joint Research Centre (JRC), Ispra, Province of Varese 21027, Italy
[e] European Dynamics, Brussels B-1000, Belgium
[f] Department of Ecology, Swedish University of Agricultural Sciences, Uppsala 75007, Sweden
[g] Department of Agroecology, Aarhus University, Forsøgsvej 1 4200, Slagelse, Denmark

ARTICLE INFO

ABSTRACT

Antibiotic resistance crisis dictates the need for resistance monitoring and the search for new antibiotics. The development of monitoring protocols is hindered by the great diversity of resistance factors, while the "street-light effect" denies the possibility of discovering novel drugs based on existing databases. In this study, we address these challenges using high-throughput environmental screening viewed from a trait-based ecological perspective. Through an in-depth analysis of the metagenomes of 658 topsoil samples spanning Europe, we explored the distribution of 241 prokaryotic and fungal genes responsible for producing metabolites with antibiotic properties and 485 antibiotic resistance genes. We analyzed the diversity of these gene collections at different levels and modeled the distribution of each gene across environmental gradients. Our analyses revealed several nonparallel distribution patterns of the genes encoding sequential steps of enzymatic pathways synthesizing large antibiotic groups, pointing to gaps in existing databases and suggesting potential for discovering new analogues of known antibiotics. We show that agricultural activity caused a continental-scale homogenization of microbial antibiotic-related machinery, emphasizing the importance of maintaining indigenous ecosystems within the landscape mosaic. Based on the relationships between the proportion of the genes in the metagenomes with the main predictors (soil pH, land cover type, climate temperature and humidity), we illustrate how the properties of chemical structures dictate the distribution of the genes responsible for their synthesis across environments. With this understanding, we propose general principles to facilitate the discovery of antibiotics, including principally new ones, establish abundance baselines for antibiotic resistance genes, and predict their dissemination.

1. Introduction

Secondary metabolites with antibiotic activities serve as mediators of antagonistic interactions within biotic communities, and the perpetual evolution of antibiotic synthesis and resistance stands as a pictorial example of an arms race in nature (Allen et al., 2010; Larsson and Flach, 2022; Nesme and Simonet, 2015; Waglechner et al., 2021; Wright, 2007). Humans consciously joined this race in the 1940th when antibiotic resistance was discovered soon after the massive production of penicillin. The ubiquitous utilization of antimicrobials in medicine,

veterinary, and agriculture (Kümmerer, 2009) made antibiotic resistance one of the greatest global challenges to humanity (Murray et al., 2022) necessitating searching for new antimicrobials (Melander and Melander, 2017; Peek et al., 2018) and monitoring and predicting sources of resistance and routes of its dissemination (Access to Medicines Foundation, 2020; Bengtsson-Palme et al., 2023, 2017; Environment, 2017; Larsson and Flach, 2022; Velazquez-Meza et al., 2022). Meanwhile, the "streetlight effect" denies the possibility of discovering new antibiotics and biosynthetic pathways based on existing databases, necessitating rational optimization in determining where to "install the new light pole" and commence the search for principally novel drugs (Lewis, 2020). Creating the protocols for antibiotic resistance also highlights the necessity of cost-effective approaches, and understanding the baseline resistance levels in natural settings (Abramova et al., 2023; Bengtsson-Palme et al., 2023; Morar and Wright, 2010).

Tools and knowledge from various fields, including epidemiology and microbial ecology (e.g., Bengtsson-Palme et al., 2023; Larsson and Flach, 2022), −omics approaches (Zhuang et al., 2021), and research on horizontal gene transfer (Meng et al., 2022), have been proposed for the development of resistance monitoring. We propose that certain challenges can also be addressed through the lens of trait-based thinking, commonly used in ecological forecasting (Garnier et al., 2015; Green et al., 2022; McGill et al., 2006). In accordance with the central concept in trait ecology the capacities of microorganisms to synthesize and neutralize antibiotics are functional traits that are not only interrelated but also undergo concurrent pressure from multiple environmental factors such as sorption of an antibiotic to mineral particles, medium pH, concentration of interfering compounds (soil complexity at the nanoscale is explicitly illustrated in a review of Nesme and Simonet (2015)). For example, an extensive research of peptidases distribution across prokaryotes (Nguyen et al., 2019) provides evidence for the connection between the pH optimum of the metabolites and the pH preferences of taxa excreting them.

While multifactorial experiments on antibiotic resistance (e.g., (Schaffner et al., 2021)) are rare, numerous small-scale screenings of the environmental resistome (the collection of resistance genes, *cf.* (Wright, 2007)) vary too widely to derive consistent patterns in environment–resistance relationships, as demonstrated in the extensive meta-analysis (Abramova et al., 2023). A few existing large-scale studies targeting genes encoding antibiotic synthesis and resistance in natural microbial communities suggest that the potentials (i.e., proportions of the genes in a metagenome) for both functions are not only interrelated but also distributed across habitats and environmental gradients in a distinctive manner (Bahram et al., 2018; Delgado-Baquerizo et al., 2022; Kerfahi et al., 2023; Nesme et al., 2014). These results suggest that accounting for environmental properties may, at the very least, secure against the infeasible task of 'monitoring everything and everywhere' or may be crucial for determining baseline levels of resistance and predicting routes of its dissemination (Bengtsson-Palme et al., 2023). In other words, understanding the mechanisms driving the ability to synthesize antibiotics and resist them in natural microbial communities may streamline efforts for monitoring resistance and discovering new antimicrobials.

Here we sought to find large-scale environmental regulators of soil microbial community genetic machinery that are responsible for antibiotic synthesis and resistance. We focus on soil microbiome since it is a major contributor to indoor microbiomes (Banerjee and van der Heijden, 2023; Miletto and Lindow, 2015), holds the majority of known antibiotic synthesizers (Santos-Aberturas and Vior, 2022) and participates in the emergence and spread of antibiotic resistance (Forsberg et al., 2012). The work is based on the European Commission's LUCAS Soil survey presenting more than 650 soil samples collected across the European Union by a single protocol and accompanied by comprehensive metadata including land cover, edaphic, and climatic variables (Orgiazzi et al., 2022). We aim to answer the following questions: 1) How does agricultural activity (as the most geographically extensive

human activity) impact the diversity and composition of soil ABG and ARG? 2) To what extent are the compositions of antibiotic biosynthesis and resistance genes (ABG and ARG, respectively) interrelated and linked with community taxonomic composition? 3) What are the primary drivers behind the distributions of various ABGs and ARGs, and which ones are the most common? Based on the answers we conceptualize the patterns of gene distribution across environments in the frame of trait ecology and propose principles that may ease ABG and ARG monitoring.

2. Materials and methods

2.1. Soil sampling

The LUCAS soil samples were collected from April to December 2018 across 26 European countries. Sampling covered all major vegetation formations (boreal, mesophytic, Pannonian, thermophilic, and Mediterranean forests) and land cover types (coniferous and deciduous forests, grasslands and croplands) (fig. S1). In each sampling site, five 20-cm-deep soil cores within a circular plot (with 4 m diam.) were collected and bulked. A subsample destined for molecular-genetic analyses was frozen within a few hours, transported to the University of Tartu and kept frozen until DNA extraction, with no evidence of melting during the logistics and storage. Variables describing physico-chemical soil properties were obtained with analyses previously described (Orgiazzi et al., 2018), climatic data with spatial resolution 30 sec ($\sim$1 km^2) were taken from WorldClim database v2 (Fick and Hijmans, 2017). Environmental variables which entered statistical analyses ranged 3.76–9.62 for pH, 2.17–225 mS/m for electrical conductivity (denoted as EC in the figures), 0–795 g/kg for carbonate content (CaCO$_3$), 0–157.7 mg/kg for extractable phosphorus content (P$_E$), 2.3–536.9 g/kg for organic carbon content (C$_{ORG}$), 2.88–69.38 for C:N ratio, 0–66 % for clay content (Clay), 0–2.86 for water volumetric content (MCv), 341–2193 mm for mean annual precipitation (MAP), and −0.88–19.63 °C for mean annual temperature (MAT).

2.2. Molecular analysis

DNA was extracted from 2.0 g of frozen soil using the PowerMax Soil DNA Isolation kit (MoBio, Carlsbad, CA, USA), following the manufacturer's protocol, and further purified using the FavorPrepTM Genomic DNA Clean-Up Kit (Favorgen, Vienna, Austria). For metagenomic analysis, DNA samples were processed following the protocol of Bahram *et al.* (Bahram et al., 2018). Briefly, we utilized the Nextera XT DNA Library Prep Kit (Illumina Inc., San Diego, CA, USA) to create metagenomic libraries following the manufacturer's guidelines. We used the Nextera Index set to index samples in the library and employed 5 μl of DNA template with a concentration of 0.2 ng/μl. To measure DNA concentration, we used the Qubit 1X dsDNA High Sensitivity kit (Invitrogen, Thermo Fisher Scientific, Waltham, Massachusetts, USA). Sequencing was performed on an Illumina NovaSeq platform in 2x150 paired-end mode.

2.3. Bioinformatics processing

In total, metagenomics libraries comprised 15.7 billion paired-end reads. Sequences were quality-checked using MultiQC v.1.10 (Ewels et al., 2016). We used fastp v.0.20.1 (Chen et al., 2018) to perform quality filtering of reads, removal of the sequencing adapters and poly-G tails (length > 4 bp), and the correction of reads based on the overlap of paired ends (options '*–overlap_len_require 40 –overlap_diff_limit 5 –overlap_diff_percent_limit 15*'). Since Illumina NovaSeq uses binned quality scores, we excluded reads with more than 20 % of unqualified bases during quality filtering. Furthermore, we identified and removed low-quality reads with three or more expected errors (Edgar and Flyvbjerg, 2015) and three or more ambiguous nucleotides.

For the removal of reads belonging to PhiX sequencing control and decontamination, we followed the DOE JGI Metagenome Workflow (Clum et al., 2021). The BBMap program from the BBTools software suite v.38.87 (Bushnell, 2024) was used to map reads to masked human, cat, dog, and mouse reference genomes at the 94 % identity threshold, and matching reads were excluded from further analysis. For correction of sequencing errors, we used clumpify ('*ecc passes=4*') and tadpole ('*mode=correct ecc k=60*') functions of the BBTools. After quality filtering, 14.8 billion reads, averaging 22.5 ± 17.8 (9.1–177.8) million per sample were included in subsequent analyses.

For the analysis of taxonomic composition of the communities, we employed Metaxa2 v.2.2.3 (Bengtsson-Palme et al., 2015) for taxonomic identification of the rRNA small (SSU) and large (LSU) subunit genes present in metagenomic data.

Six hundred seventy antibiotic biosynthesis genes (ABG, functional orthologs responsible for the biosynthesis of secondary metabolites with antibiotic activity, excluding terpenoids) of interest were selected based on KEGG (Kanehisa et al., 2023) and relevant literature sources (**table S1**). To search for them in the metagenomes, we performed functional annotation of reads using eggNOG-mapper v.2.1.2 (Cantalapiedra et al., 2021) based on eggNOG orthology data v.5.0.2 (Huerta-Cepas et al., 2019). For a high-throughput sequence search we followed (Bahram et al., 2018) and used DIAMOND v.2.0.10 (Buchfink et al., 2021) in blastx mode with '*–evalue 1e-4 –sensitive*' options. To select the best matches, we further filtered the aligned reads to keep only the matches with alignment percentage identity > 50 % and E-value < 1e-8. We determined the best match for a read pair by considering the highest bit score, longest alignment length, and highest percentage identity to the target sequence, for both the forward and reverse reads. In cases where a read pair had no target sequences in common, we assigned functional annotation independently to each read. For a higher-level taxonomic assignment of ABGs, we relied on the information about the taxonomy of a seed ortholog from the eggNOG database.

To identify antibiotic resistance genes (ARG, genes that if expressed provide antibiotic resistance to its host) in the metagenomes, we employed the short-sequence model of DeepARG v.1.0.2 (Arango-Argoty et al., 2018) with the following parameters: a minimum identity of 80 % for ARG alignments, a minimum probability of 0.8 for considering a read as ARG-like, and a minimum E-value of 1e-10 for ARG alignments. These stringent parameters were employed to address the considerable sequence similarity observed between certain ARGs and their non-resistant counterparts. It is important to note that in certain instances, the similarity between these genes can reach approximately 90 %, as exemplified by the ARG *rpoB2* and its non-resistant counterpart *rpoB* (Bengtsson-Palme et al., 2017). However, imposing even more stringent parameters carries the risk of excluding actual antimicrobial resistance genes from the results. In light of this finding, we acknowledge the potential challenges associated with distinguishing between specific ARGs based solely on sequence similarity. To overcome this challenge, future studies could incorporate the sequence of the non-resistant version into the database and implement a strategy to exclude reads that align better with the non-resistant version compared to the resistant version.

For each detected ARG, we determined the gene family, antibiotic class, and mechanism of resistance (**table S4**). Genes with the potential for mobility included those with the proportion of findings on mobile genetic elements > 0.3 (data on gene occurrences in chromosomes and mobile genetic elements was taken from CARD (Alcock et al., 2023) and literature).

The bioinformatics workflow was managed with Snakemake (Mölder et al., 2021).

2.4. Statistical analysis

All statistical analyses were performed using R v.4.2.2 (R Core Team, 2023). As a proxy relative abundance of each gene, with sequencing depth taken into account, we calculated the residuals of logarithmically-transformed gene read counts by performing robust linear regression against the logarithm of sequencing depth. To reduce the influence of potential outliers on the model fit, we used the MM-estimator; the analysis was performed with the *rlm* function of the MASS package v.7.3-58.3 (Venables and Ripley, 2002). We analyzed the distribution of individual genes across different land cover types and the influence of environmental predictors on them, focusing only on genes with a relative abundance cardinality (number of unique values) > 25. This threshold was set because models failed to converge or yielded unreliable estimates for genes with a lower frequency and variability. Meanwhile, genes with low abundance and frequency were included in the analyses of the relative abundance of gene groups (e.g., ABG involved in the same biosynthetic pathway, ARG belonging to the same gene family, conferring a similar resistance mechanism or resistance to similar antibiotics **tables S3, S4**).

For the assessment of the effects of land cover on gene abundances and composition, we considered four land cover types (coniferous forests, deciduous forests, grasslands, and croplands). Most sampled coniferous forests are used for wood production (Köninger et al., 2023), whereas one third of deciduous (broadleaved) woodlands, represent agricultural stripes, gardens, recreational or abandoned areas. We, therefore, considered deciduous forests as more intensely managed than coniferous ones. We also assumed management in forests to be less disturbing compared to agricultural lands, while management in grasslands less disturbing compared to croplands. The distribution of samples across land cover types was uneven (**fig. S1**) reflecting the distribution of corresponding land use practices (Kuemmerle et al., 2016) across climatic gradients, necessitating accounting for spatial effects when analyzing gene distribution. Also, there are land cover-related trends in soil physico-chemical properties (**fig. S1, S2**), with soil pH, carbonate and clay content increasing, while organic carbon, C: N ratio, and water content decreasing from coniferous to deciduous forests, to grasslands, and finally croplands. Notably, gravimetric water content was less strongly linked with weather conditions for the 31 days prior to sampling (Spearman correlation coefficient, r = –0.40 and r = 0.32 for air temperature and precipitation, respectively, n = 658) than with soil density (r = –0.68) and organic carbon content (r = 0.82). Due to the significant increase in soil density from forests to croplands, gravimetric water content is a biased characteristic of free water availability. To partly avoid this bias, we transformed gravimetric water content to volumetric one using soil density.

We utilized a generalized additive model (GAM) to analyze the relationship between the abundance of genes and environmental variables. To account for spatial autocorrelation, we included two-dimensional thin plate splines of the spatial locations in a GAM model (Miller and Wood, 2014), while to prevent overfitting and improve the predictive performance of a model, we used penalization (options '*select = TRUE, gamma = 1*') as implemented in the mgcv package v.1.8–42 (Wood, 2011). To identify the most important predictors, we estimated the ε^2 (Epsilon-squared) metric using the effectsize package v.0.8.3 (Ben-Shachar et al., 2020). The effect strength (slope) of environmental variables on gene abundances was estimated with partial derivatives of the regression equation with respect to a regressor of interest using the marginaleffects package v.0.11.0 (Arel-Bundock V., 2023); comparison of model predictions between land cover types were performed with '*comparisons*' function. To account for differences in sequencing depth in the analysis, we incorporated the logarithm of the number of reads that passed functional annotation as an offset in GAM models.

To estimate the richness and diversity of ABG and ARG genes, we constructed diversity profiles based on Hill numbers using the package hilldiv v.1.5.1 (Alberdi and Gilbert, 2019). These metrics incorporate a continuous parameter (q) which controls the diversity index's sensitivity to gene relative abundances: $q = 0$ equates to gene richness, where each gene contributes equally regardless of its relative abundance; $q = 1$ corresponds to the exponent of the Shannon diversity index, weighting

genes proportionally to their relative abundance; and $q > 1$ emphasizes abundant genes while penalizing rare ones. At $q = 2$, the measure reflects the inverse of Simpson's index and can be interpreted as the effective number of dominant genes in the assemblage. We used the same GAM for this analysis as were used for examining gene abundance differences across various land cover types.

To assess the similarity of samples in gene composition, we used principal coordinates analysis (PCoA), with Euclidean distance calculated based on the gene residuals. To determine the significance of differences in gene composition between land cover types, we used non-parametric permutational multivariate analysis of variance (PERMANOVA, 10,000 iterations (Anderson, 2001)) and distance-based tests for homogeneity of multivariate dispersions (PERMDISP (Anderson, 2006)), with false discovery rate (FDR) correction to adjust P-values for multiple testing. For multivariate analysis of gene distribution across environmental gradients, we used distance-based redundancy analysis (db-RDA (Legendre and Anderson, 1999)). Significance testing of the effect of environmental variables was conducted using a permutation test (5000 iterations) for each marginal term in the model. The analysis was performed using the vegan package v.2.6–4 (Oksanen et al., 2023). To evaluate the similarity between ABG, ARG, and taxonomic profiles, we used Procrustes analysis and determined its statistical significance using

a permutation test with 10,000 iterations.

We used structural equation modeling (SEM) to depict and quantify the association between composition of ABG and ARG, and environmental properties. The land cover type variable was incorporated as an ordinal predictor, ordered based on management intensity: coniferous forest < deciduous forest < grasslands < croplands. Within the SEM framework, the primary two PCoA axes derived from the ordination of ABG and ARG profiles served as latent variables, representing the composition of antibiotic synthesis and resistance genes within microbial communities. The SEM analysis was conducted using JMP v.17.1.0 (SAS Institute Inc., Cary, USA). We used the Wald Z-statistic for standardized parameter estimates. Since the sign (positive or negative) of this criterion is not relevant for ARG and ABG profile dissimilarities, we considered the absolute value of the Wald Z-statistic as a measure of effect size.

3. Results

3.1. Community taxonomic profiles

According to taxonomic profiling of 658 European soil metagenomes based on *SSU* and *LSU rRNA* genes, Actinobacteria generally dominated
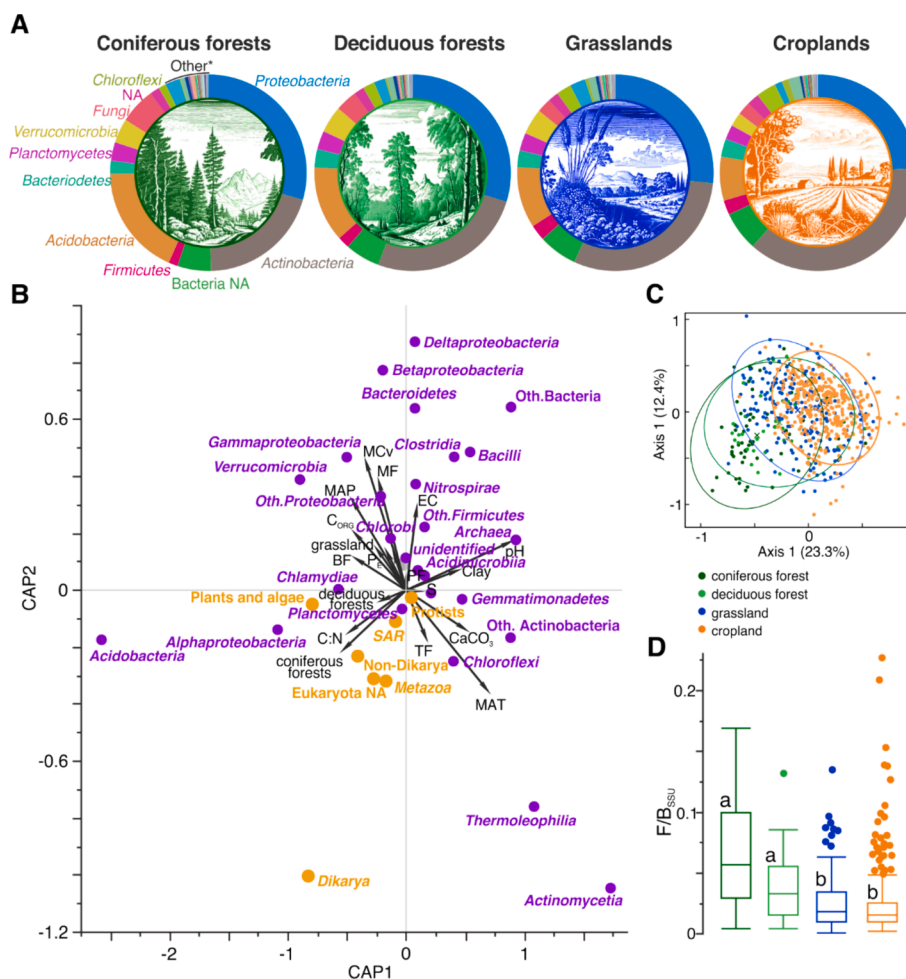
Fig. 1. Taxonomic profiling of soil biotic communities. The results of taxonomic profiling of soil metagenomes based on *SSU rRNA* gene. (**A**) Percentages of the gene reads belonging to different taxa in coniferous and deciduous forests, grasslands, and croplands. For the group "Other" see **table S1**. (**B**) Distance-based redundancy analysis (db-RDA) of taxa distribution in the gradients of environmental variables. The angle between an environmental vector and gene position reflects their correlation. BF, MF, PF, TF, and MtF denote boreal, mesophytic, Pannonian, thermophilic, and Mediterranean forests, respectively. (**C**) Taxonomic compositional dissimilarity of soil communities from different land cover types visualized with principal coordinates analysis. Cloud centroids and 90% confidence ellipses are shown for each land cover type. (**D**) Fungi-to-bacteria ratio of the gene read counts in different land cover types. Different letters denote significantly different mean values at the 5% level of statistical significance assessed with the Student's *t*-test.

soil microbial communities (26.8 % of reads, Fig. 1A; table S1) followed by *Proteobacteria* (24.5 %), with prevailing *Alphaproteobacteria* (10.5 %), and *Acidobacteria* (5.9 %). The proportions of archaeal and eukaryotic rRNA gene reads equaled 0.9 %, and 4.1 %, respectively, with fungi accounting for 1.6 % of reads, and animals for 0.3 %. Approximately 2.0 % of reads were not taxonomically annotated. The relative abundance of most taxa was significantly related to environmental parameters (Fig. 1B; fig. S3, S4), with soil pH, mean annual precipitation, soil

moisture content, land cover type, and clay content being the most influential for community composition (table S2). Correspondingly, samples from the four land cover types significantly differentiated in community composition and compositional variability (Fig. 1C; $F_{PER-MANOVA}(3; 653) = 46.2$, $P<0.001$ and $F_{BETADISP}(3; 653) = 20.5$, $P<0.001$ for *SSU* region data), with the most homogeneous communities in croplands (table S2). The fungi-to-bacteria (F/B) ratio based on *SSU* and *LSU rRNA* gene read counts was significantly higher in forests than
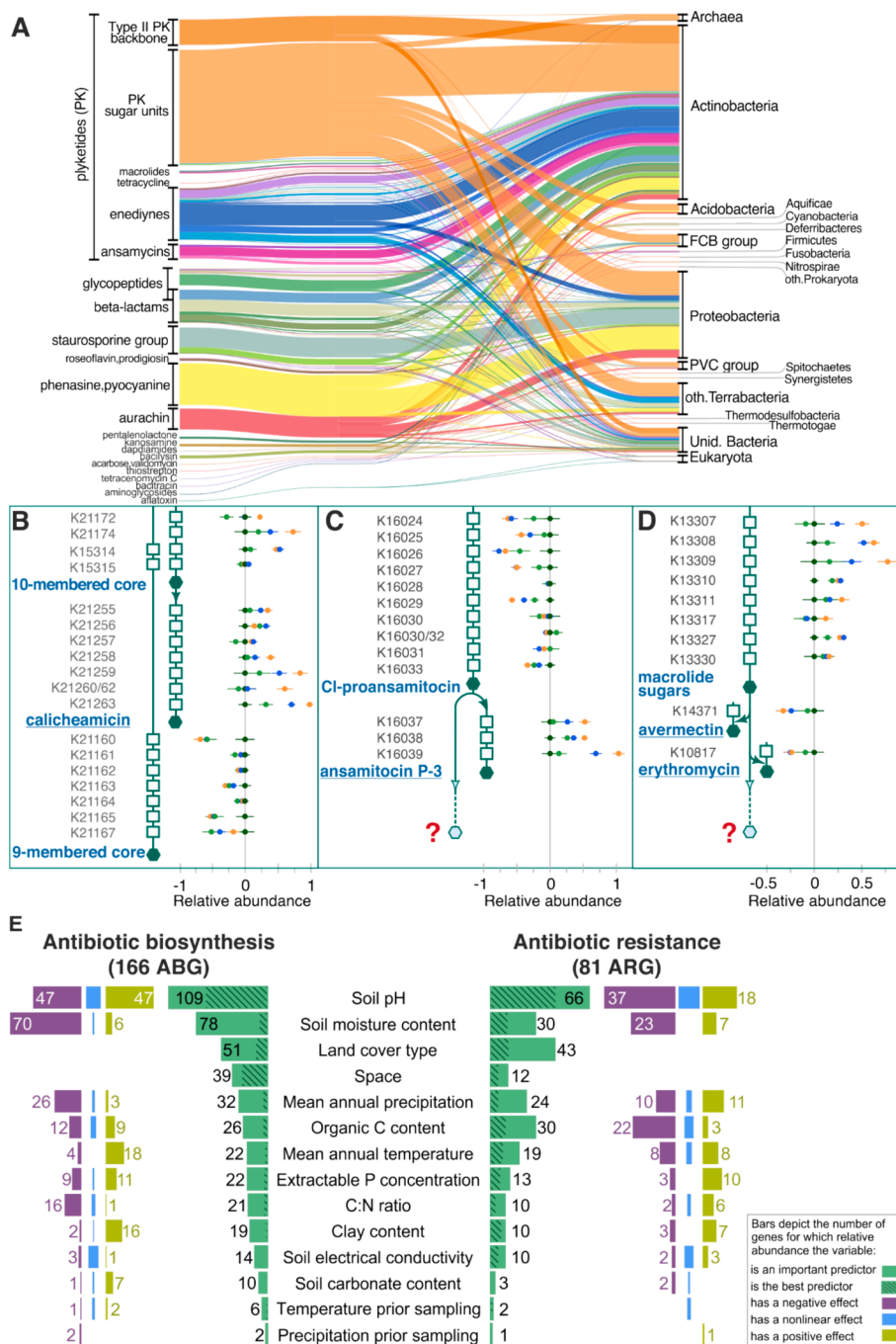
Fig. 2. Effect strength of land cover and environmental variables on gene distribution. (A) Distribution of ABG across biotic taxa (based on table S3). Line width is proportional to the relative abundance of a gene or gene group encoding a subpathway (denoted with the same color). Schematic examples of how genes encoding biosynthesis of **(B)** enediynes, **(C)** macrolides, and **(D)** ansamitocins differentiate in the distribution across land cover types. Mean value (point) and standard error (whiskers) of genes relative abundances (standardized for sequencing depth) predicted with the account for weather and space effects. Relative abundance of genes in coniferous forests is used as a reference value. Rectangles denote genes encoding the pathway; hexagons are the pathway products. **(E)** Importance and strength of the effect of environmental variables on the ABG and ARG relative abundances, as inferred with generalized additive models (for details, see tables S7–S10).

herbaceous communities (Fig. 1**D**; **Text S1**).

3.2. Community functional profiles

3.2.1. Antibiotic biosynthesis

Of 670 searched antibiotic biosynthesis genes (ABG), we revealed 241 genes (**table S3**), with omnipresent and the most relatively abundant K01710 (responsible for synthesizing the 6-deoxyhexose – a basic unit of most of the secondary metabolites) and those encoding enediyne, rifamycin, and vancomycin synthases, as well as enzymes involved in the biosynthesis of polyketide sugar units, phenazines, rebeccamycin, pyrrolnitrin, and aurachins. Ninety-nine percent of the ABG read counts belonged to Prokaryota (including 1.7 % of archaeal reads), with *Actinobacteria* accounting for 48.7 % of read counts, *Gamma-*, *Beta-*, and *Deltaproteobacteria* for 3.4–4.7 %, *Chloroflexi*, *Acidobacteria*, *Firmicutes*, *Cyanobacteria*, and *Bacteroidetes* for 2.5–3.2 %. Genes encoding enzymes involved in the biosynthesis of polyketide sugar units, enediynes, beta-lactams (except clavams), phenazines, rebeccamycin, pyrrolnitrin, and aurachins belonged to multiple taxa. Genes responsible for the synthesis of aminoglycosides and a large group of polyketides were assigned exclusively to Actinobacteria (Fig. 2**A; Table S3**), while genes encoding the production of certain peptides were restricted to *Firmicutes* (mostly *Bacilli*). *Fungi* possessed 0.14 % of all reads, accounting for 1–21 % of the abundance of three genes responsible for synthesizing enediynes and isopenicillin N, and for the total abundance of the genes encoding fumitremorgin, aflatoxin, meleagrin, and neoxaline production. Of the identified ABG, 166 reached the cardinality threshold for the analysis of their individual distribution.

3.2.2. Antibiotic resistance

In total, 485 antibiotic resistance genes (ARG) were identified (**table S4**). Most of the ARG, including those conferring resistance to fusidic acid, fosfomycin, free fatty acids, nucleosides, sulfonamides, and thiostrepton, were rare or present in low abundance. Among the 179 identified genes responsible for drug efflux, 54 genes, participating in generation of 50 (mainly multidrug) efflux systems, had enough frequency for analyzing their individual distribution drivers. Of the rest 306 genes conferring other mechanisms of resistance, 27 were frequent. Infrequent genes, whose individual distribution could not be analyzed, were included in the analysis of the distribution of gene groups (e.g., gene families).

ARG with the potential for mobility were on average shorter than chromosomal ones ($F(1; 449) = 57.8$; $P \ll 0.001$), while considering this difference were similar in average relative abundance ($F(1; 449) = 21.9$; $P \ll 0.001$ for the factor "gene length" and $F(1; 449) = 0.07$; $P=0.79$ for "mobility potential" **table S4**).

In total, ca. 93 % of ARG sequences were taxonomically annotated with Kraken2, with 47 % of reads belonging to *Actinobacteria*, 39 % – to *Proteobacteria* (with *Alpha-*, *Beta-*, and *Gammaproteobacteria* accounting for 16 %, 9.5 % and 9.7 % of reads). *Acidobacteria* and *Firmicutes* each represented ca. 1.5 % of ARG reads (**table S4**).

3.3. Gene distribution across land cover types and environmental gradients

By accounting for spatial effects and weather conditions prior sampling, a significant impact of land cover type on gene distribution was revealed for 125 of 166 analyzed ABG (**table S5, S6**). Genes encoding alternative or consecutive branches of large biosynthetic pathways showed distinct distributions across different land cover types. For instance, genes encoding production of a 9-membered enediyne core were the most relatively abundant in forests, while those responsible for assembling a 10-membered enediyne core – in croplands and grasslands (Fig. 2**B**). Also, genes destined for generating proansamitocin prevailed in coniferous forests, while genes providing its finalizing to ansamitocin P-3 – in croplands (Fig. 2**C**). Similarly, ABG responsible for the synthesis

of macrolide sugar units were more abundant in croplands, while ABG encoding macrolide synthetases – in forests (Fig. 2**D**). The partitioning of the other genes among land cover types is shown in **figs. S5-S9**.

Of 81 analyzed ARG, 76 ARG significantly differed in relative abundance between land cover types. Members of the same gene families or encoding same type efflux systems were relatively similarly distributed across land cover types (**fig. S10; tables S7, S8**). A detailed description of the other ABG and ARG distributions across land cover types is in **Text S2**.

Of all explored environmental predictors, soil pH significantly affected the distribution of the largest portion of genes (66 % of ABG and 81 % of ARG), and it was the best predictor for 63 % of the affected genes (Fig. 2**E**). With some exceptions (**Text S3**), the majority of genes which had significantly higher relative abundance in forests and grasslands compared to croplands (see above) were negatively related to pH and vice versa (**tables S9, S10**). Soil water volumetric content negatively influenced the abundance of most affected ARG and ABG, while clay and phosphorus contents positively. C:N ratio was influential (mostly negatively) on the distribution of many ABG, while soil organic carbon content, on ARG. Of the 69 ABG and 36 ARG, whose relative abundance was significantly related to geographical location, mean annual precipitation and temperature, the abundance of 60 ABG and 17 ARG increased towards arid and warm regions, while 7 ABG and 17 ARG were more abundant in wetter and colder humid regions. Distribution of ABG and ARG in relation to environmental variables is depicted in Fig. 3 **and** Fig. 4.

3.4. Differentiation of land cover types in antibiotic synthesis and resistance potential

In the summary results of the analyses of individual gene distributions, 34–38 % of ABG and 30–43 % of ARG showed higher relative abundance in croplands compared to the other land cover types and in grasslands compared to coniferous forests (Fig. 5**A, B**). Comparable fractions of ARG, while only 10–20 % of ABG, increased in abundance towards forests. Accordingly, the Hill number (*D*) values at $q = 0$ (which corresponds to richness) calculated for ARG and ABG increased from forests to croplands (Fig. 5**C, D**). Meanwhile, with increasing q (that is with decreasing contribution of unabundant genes or gene group into *D*) the difference between forests and agricultural lands decreases. At $q \geq 1$, *D* calculated for ARG and ARG families is significantly smaller in agricultural lands compared to forests. Notably, the pattern is especially pronounced for chromosomal ARG (**fig. S11A**).

3.5. Congruence between community functional and taxonomic profiles

Land cover types significantly differentiated in the ABG and ARG composition (PERMANOVA: $F(3; 654) = 17.7P \ll 0.001$; $R^2 = 0.08$ for ABG and $F(3; 654) = 23.15$; $P \ll 0.001$; $R^2 = 0.10$ for ARG; Fig. 5A), and compositional variability (PERMDISP: $F(3; 654) = 14.6P \ll 0.001$ for ABG and $F(3; 654) = 5.3$; $P \ll 0.001$ for ARG). The most different gene compositions were between croplands and coniferous forests, while croplands also varied in the gene compositions significantly less than the other types of land cover (**table S11**). SEM analysis identifies soil properties, and particularly soil pH, as the primary drivers of ABG and ARG composition followed by climate variables (Fig. 6**B**). Soil properties also act as intermediaries in the influence of climate and land cover.

According to the results of Procrustes analysis, taxonomic, ABG, and ARG compositions were significantly correlated, with the best congruence of ABG composition with ARG (Procrustes congruence coefficient 0.60) and with taxonomic composition (0.60) (**tables S12** – Fig. 6**C**). The least congruence with community taxonomic and ABG structures was found for mobile ARG (0.44 and 0.46, respectively). The degree of congruence was positively affected by pH and clay content, while negatively by organic C content and annual mean temperature (**table S12**). Comparison of land cover types showed that with management
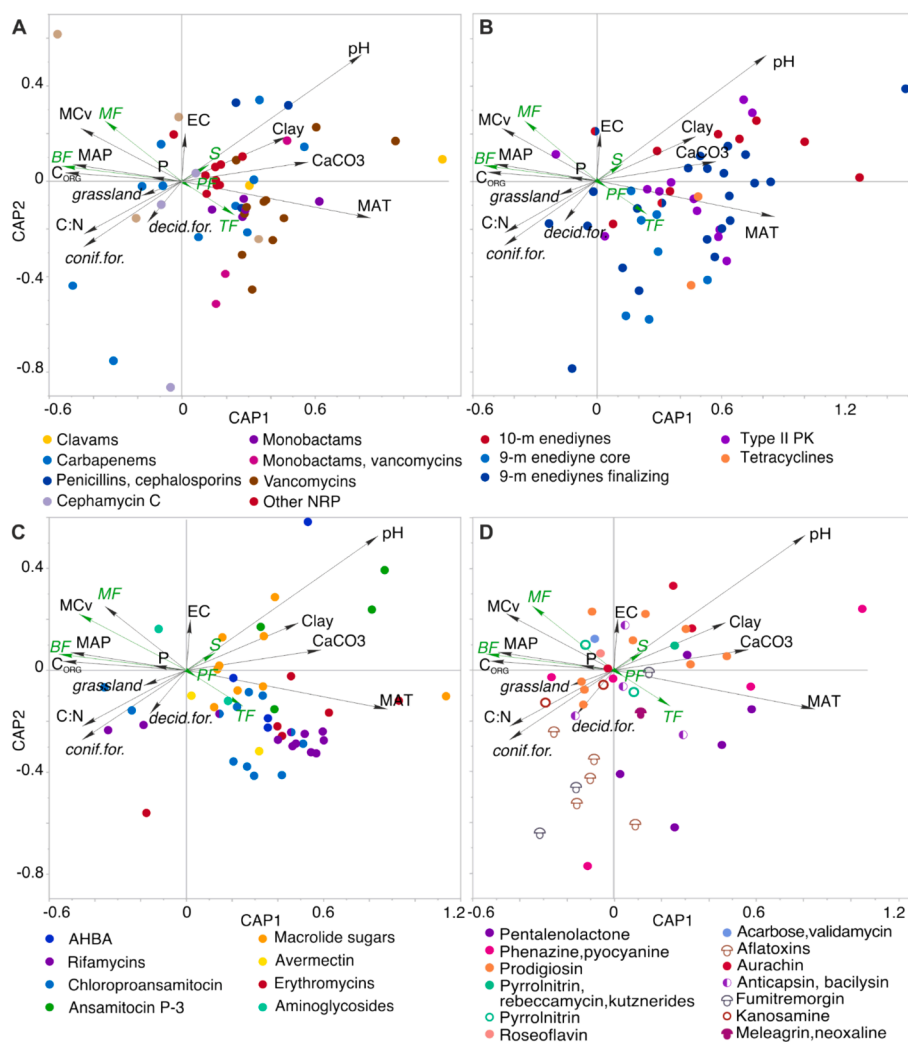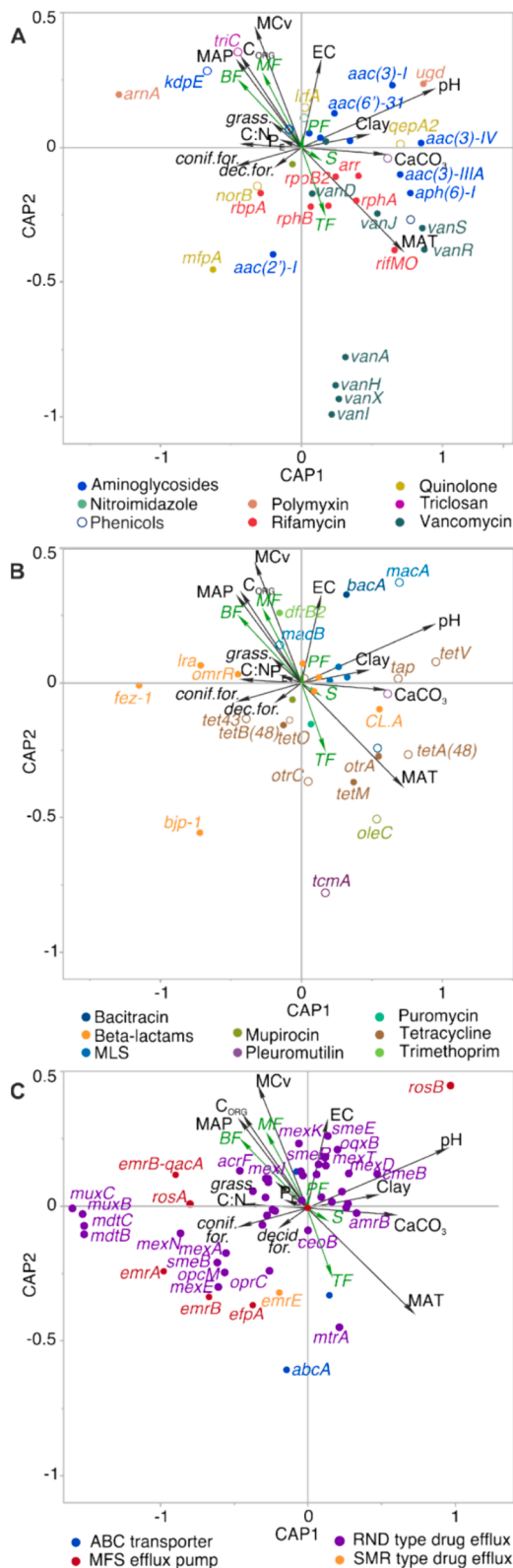
Fig. 3. ABG distributions in environmental gradients. Distance-based redundancy analysis (db-RDA) results showing the distributions of genes encoding synthesis of **(A)** nonribosomal peptides, **(B and C)** polyketides and aminoglycosides, **(D)** other antimicrobials in the gradients of environmental variables. The angle between an environmental vector and gene position reflects their correlation. Vegetation formations (green arrows) are denoted as in Fig. 1. Solid markers denote antibacterial or nondiscriminatory toxins, empty markers denote antifungal toxins; circles depict bacterial metabolites, mushroom-shaped markers depict mycotoxins. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

load, the congruence between the communities' taxonomic, ARG, and ABG compositions increases (Fig. 6C). Only the level of congruence between the composition of mobile ARG with chromosomal ARG and taxonomic composition did not differ between land cover types.

4. Discussion

Based on rRNA data and our previous metabarcoding-based research (Labouyrie et al., 2023), Actinobacteria constitute a significant fraction of soil microbial communities, particularly in croplands. Meanwhile, the fraction of ABGs and ARGs harbored by Actinobacteria in the metagenomes was disproportionately high compared to Proteobacteria and Acidobacteria, which are also dominant in soil microbiomes. This suggests that ABGs and ARGs associated with these and potentially other minor bacterial phyla may be underrepresented in our study and others focusing on soil antibiotic-related genetic machinery, possibly due to the overrepresentation of Actinobacteria in reference databases (Santos-Aberturas and Vior, 2022). Exploring whether this discrepancy is attributable to inherent competitive abilities of Actinobacteria or artificial biases in reference databases is a subject for future investigation.

4.1. Human-driven homogenization of soil biota

The continental-scale analysis of topsoil metagenomes reveals that land cover type is among the primary drivers of composition and diversity of antibiotic biosynthesis and resistance genes. Generally, the number of ABG and potentially synthesized antimicrobials (within the studied spectrum), and correspondingly the number of ARG, increases with agricultural land use intensity (see Fig. 5C, D). The analysis of diversity profiles suggests that the increase in gene richness in agricultural lands results from an increased number of rare and unabundant genes while a certain set of highly abundant genes primarily chromosomal (e.g. conferring rifamycin and vancomycin resistance or encoding RND type efflux pumps, fig. S11B) dominate in all samples.

The trends of increasing microbial alpha diversity and in particular the diversity of soil chemoheterotrophic bacteria, pathogenic fungi, and invertebrates in agricultural lands has been numerously shown (Köninger et al., 2023; Labouyrie et al., 2023; Trivedi et al., 2016). These results illustrate how a short-period high resources availability (during fertilization, irrigation, and crop yielding) alternating with microhabitat destruction (tillage) (Moon et al., 2019; Szoboszlay et al., 2017), together with soil homogenization, creates conditions favorable for species-rich communities of copiotrophic *r*-strategists, for whom

Fig. 4. ARG distributions in environmental gradients. Distance-based redundancy analysis (db-RDA) results showing ARG distributions in the gradients of environmental variables. **(A and B)** Genes conferring resistance to certain antibiotics; empty markers denote genes providing resistance with efflux, solid markers denote other mechanisms. **(C)** Genes encoding multidrug efflux pumps. The angle between an environmental vector and gene position reflects their correlation. Vegetation formations (green arrows) are denoted as in Fig. 1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

competitive ability is a vital characteristic. Instead, in extensive grasslands and forests, the horizontal heterogeneity of the organic horizon is maintained by the compositional variability of diverse plant roots and litter. More importantly, the succession of litter decomposition results in vertical fine-scale stratification of the undisturbed topsoil (Vorobeichik and Korkina, 2023). These horizontal and vertical components provide numerous niches for microorganisms, mitigating resource competition (Lindahl et al., 2007; Mikryukov, Dulya, 2018).

When analyzing taxonomic and functional beta-diversity, we registered a significantly lower diversity among cropland sites compared to what was found among other land cover types across Europe. The findings are in concordance with the results of recent metabarcoding-based studies of LUCAS samples, showing lower variability of croplands in the composition of eukaryotic communities compared to less intensively managed ecosystems (Köninger et al., 2023; Banerjee et al., 2024). In addition, we revealed that management intensification increases the coherence between community genetic machineries responsible for antibiotic synthesis and resistance as well as their congruence with taxonomic composition. Together these results imply that cropland soil microbiota, while taxonomically and functionally richer, is taxonomically and functionally uniformly structured, homogeneous and is more predictable than microbiota of forests and extensive grasslands.

Reasonably, agricultural regularization is the main cause of such large-scale homogenization: croplands are primarily oligo- and mono-cultures with a limited spectrum of crops (Raggi et al., 2022) and standardized edaphic characteristics (exemplified here by a greater edaphic homogeneity than forest samples; **fig. S1**). The observed continental-scale tendency for the homogenization of soil biota in agricultural lands emphasizes the importance of maintaining indigenous ecosystems within the landscape mosaic. Retaining regional biota specificity, in turn, facilitates the provision of essential ecosystem services (in this specific case, providing with new antibiotics and barriering against resistance proliferation and dissemination).

4.2. Shapers of community antibiotic biosynthetic machinery

With an in-depth analysis of community genetic machinery for antibiotic biosynthesis, we revealed that soil metagenomes contain higher proportions of genes involved in common steps (generating basic units or core structures destined for assembling the members from the same large groups of metabolites) compared to specific steps (maturation of particular compounds) of the same biosynthetic pathway. This is consistent with phylogenetic studies, showing that taxa sharing common biosynthetic steps show divergence in specific ones (e.g., Liu et al., 2018).

More intriguingly, genes responsible for different biosynthetic steps exhibited divergence in their distributions across land cover types and environmental gradients (**fig. S5-S9**, **Text S2**). Consequently, the ratio of gene abundances responsible for different steps in the biosynthesis of specific metabolite groups varies among habitats. In some habitats, this results in an overproportional abundance of genes responsible for the initial synthetic steps and, correspondingly, a "lack" of the genes known for encoding later steps (i.e., finalizing certain known metabolites). This phenomenon can be exemplified by the genes responsible for assembling ansamitocins (see Fig. 2**D**) and macrolides (see Fig. 2**C**) and points to the potential synthesis of as-yet-undiscovered members of the metabolite group in the given habitat (e.g., ansamitocins in forests and macrolides in croplands). The potential can be further explored in a targeted search, exemplified by the survey identifying new genes encoding enzymes tailoring rifamycin congeners active against common rifamycin-resistant strains (Peek et al., 2018). Such findings might not overcome the "streetlight effect" but point to the gaps in the extant databases highlighting omitted analogues of clinically used antimicrobials in the soil microbiome.

The search for new antibiotics can also be optimized by the
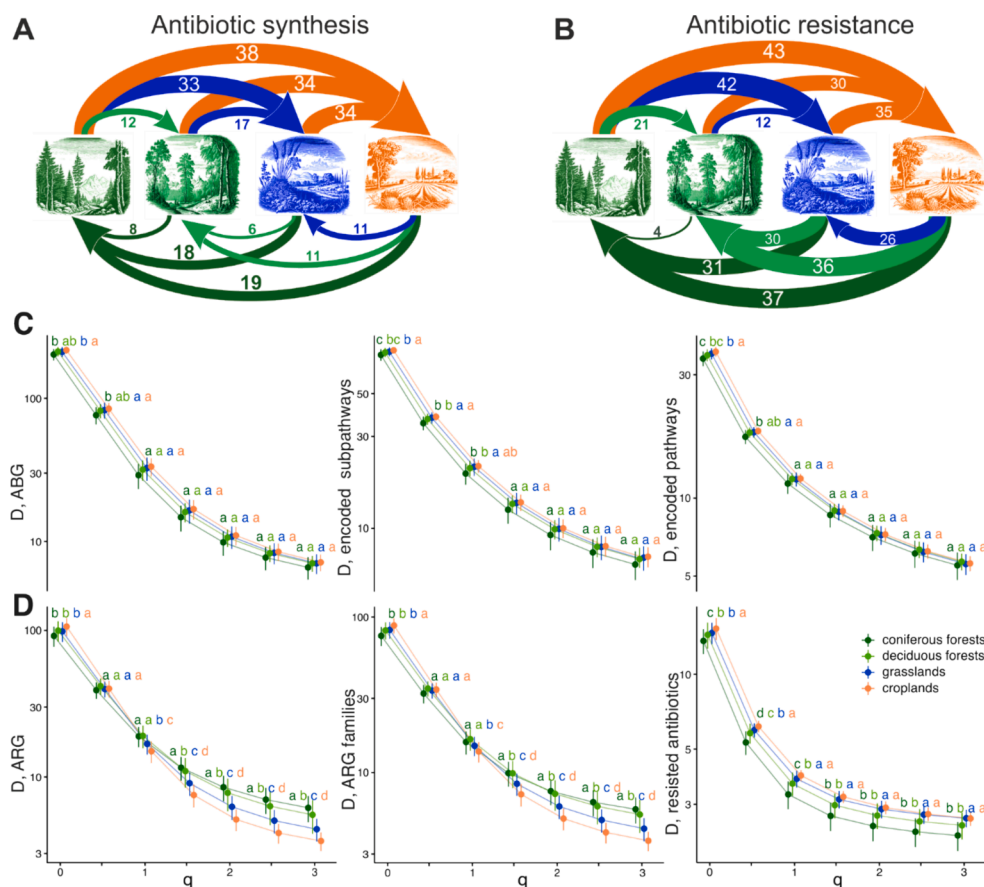
Fig. 5. ARG and ABG diversity across land cover types. Percentages of (**A**) ABG and (**B**) ARG with significantly different relative abundance between pairs of land cover types; lines connect the pairs of compared land cover types; arrows point to a land cover type with a higher abundance of genes. Hill numbers (*D*) calculated for antibiotic biosynthesis (**C**) and resistance (**D**) machineries in soil metagenomes. *D* was calculated at different levels of gene grouping (grouping principles are in **tables S2, S3**). Average and standard deviation are shown for each land cover, different letters denote significant differences between land cover types. Notably, Y axes are logarithmic.

knowledge about the factors driving the potential for their biosynthesis. Given a recently validated link between antibiotic structure and specificity (Wong et al., 2024), particular tight links revealed in this work between habitat characteristics and abundance of genes responsible for the biosynthesis of certain metabolites illustrate that the potential for producing an antimicrobial in a community can be driven by the presence (or abundance) of organisms targeted by the antimicrobial. For instance, gene families intrinsic to various bacterial phyla, but providing the production of primarily antifungal kanosamine and pyrrolnitrin (**table S1**) (Masschelein et al., 2017), peak in relative abundance in forest and/or acidic soils, where fungi prevail (Bahram et al., 2018; Hagh-Doust et al., 2023; He et al., 2020; Labouyrie et al., 2023; Siles et al., 2023; Smith et al., 2021; Szoboszlay et al., 2017). Basing the search on the presence of a targeted taxon is especially useful when looking for narrow-spectrum antimicrobials that are particularly valuable during the antibiotic resistance crisis (Chou et al., 2022; Melander and Melander, 2017).

Considering the great structural diversity of antimicrobials, the potential for their production in a community should be regulated by other factors besides target group abundance. The effect of other factors can be illustrated with the genes encoding biosynthesis of toxins with nondiscriminatory activity, like extraordinary cytotoxic enediynes with high potency for DNA scission (promising candidates for anticancer agents). Possessing an unexpectedly high relative abundance and diversity of carriers (**Table S3**), genes involved in the biosynthesis of 9- and 10-membered enediynes differentiated in distribution patterns (see **Fig. 2B**). The revealed prevalence of 9-membered enediynes in forest (i.

e., acidic) soils is likely at least partly caused by the importance of strongly acidic conditions for their structural stability and stable association with its protein (Baker et al., 2007; Edo et al., 1988; Kandaswamy et al., 2008; Zein et al., 1995) by contrast with 10-membered enediynes.

Therefore, through the lens of a trait-based approach (Garnier et al., 2015; Green et al., 2022; McGill et al., 2006), the revealed large-scale patterns in the distribution of soil microbiome potential for synthesizing different metabolites across environmental gradients illustrate the well-known principle of "environment selects". We posit that guided by the principle it is possible to narrow down the search for new antimicrobials to habitats with a higher probability harboring microbiota that synthesizes a metabolite with desired properties.

4.3. Shapers of soil resistome

We show that the soil resistome is linked with community taxonomic composition and genetic machinery providing antibiotic biosynthesis and is also determined by the environment, corroborating the tight relationships between enzymes activity optima and ecological preferences of their carriers shown for secreted peptidases (Nguyen et al., 2019). Observed here, close relationships between the distribution of horizontally transferred ARG and habitat characteristics indicate relatively straightforward mechanisms for how the environment regulates a microbial community's potential for antibiotic resistance. This suggests that by considering the properties of resistance factors, one can anticipate the environmental conditions that would promote or impede their
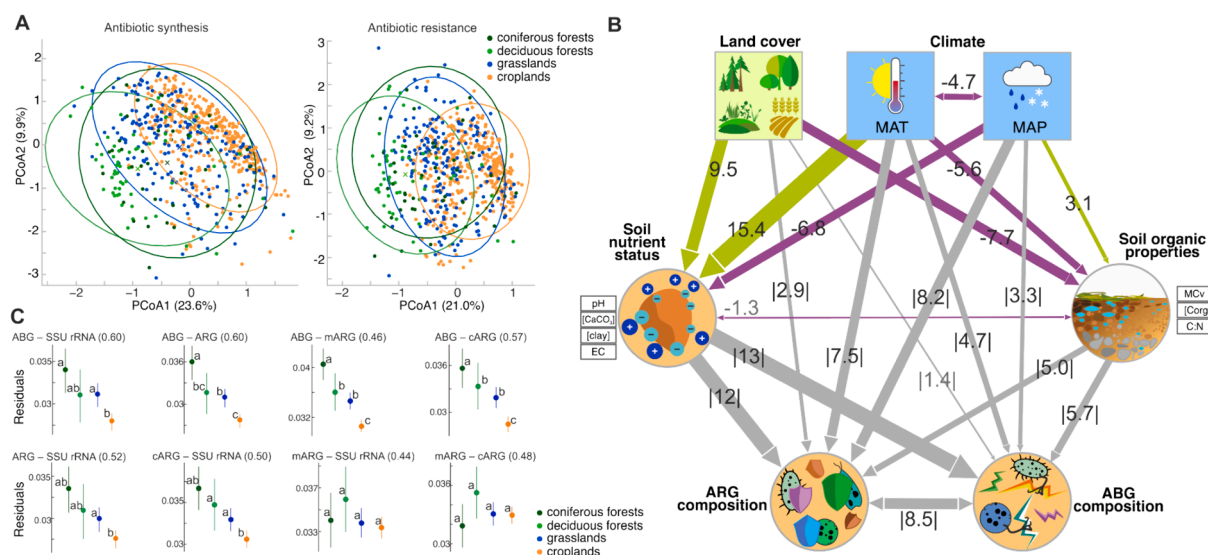
Fig. 6. Predicting ABG and ARG composition in soil metagenomes. (**A**) Differentiation of land cover types in the composition of ABG and ARG in the samples. Variance explained by the principal coordinates is shown in round brackets. Cloud centroids (crosses) and 90% confidence ellipses are shown. (**B**) Structural equation model of the relationships between ABG and ARG composition, and environmental properties. Numbers indicate the effect sizes for model parameters (Wald statistic); arrow width is proportional to the effect size; green denotes a positive effect, purple indicates a negative effect, and gray signifies the inapplicability of the effect sign. Land cover type is an ordinal variable reflecting the increase of management load with the least load in coniferous forests and the largest in croplands (see methods). (**C**) Mean predicted Procrustes distance between samples in the compositional ordinations based on SSU rRNA gene region, ABG, and mobile and chromosomal ARG (mARG, cARG, respectively) across land cover types (weather and spatial effects are taken into account). Procrustes congruence coefficients are shown in brackets. Different letters denote significantly different mean Procrustes distance. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

presence, thus improving the accuracy of predicting their natural levels and dissemination routes.

Specifically, the revealed increase of total abundance of genes encoding beta-lactamases towards acidic soils, corroborates global trends (Bahram et al., 2018). In a detailed view, this trend is due to metallo-beta-lactamase genes (class B, constituting the largest portion of found beta-lactam ARG in the soil, **table S4**), while the abundance of serine-beta-lactamase (specifically, class A) genes tends to increase towards neutral soils (**table S11**). Such divergence could be addressed from the perspective of taxonomic affiliation of the main gene carriers (Berglund et al., 2021; Philippon et al., 2016), though the *trans*-phylum gene transfers have already been documented (Galán, 2002). Meanwhile, as a functional trait, the ability to excrete an enzyme is apparently subject to environmental selection for the enzyme properties. Correspondingly, the revealed divergence in the distribution of genes encoding the two beta-lactamase classes between acidic and neutral soils reflects the difference in the optimum pH for their activity (higher for serine- than for metallo-beta-lactamases (Bounaga et al., 1998; Brannigan et al., 1991; Jacob et al., 1991; Rasia and Vila, 2002; Tsang et al., 2005; Waley, 1975). Similarly, the prevalence of *aac(2′)* in acidic soils, while genes encoding the other categories of aminoglycoside inactivating enzymes in neutral and alkaline soils, mirrors the dissimilarity of pH optima between encoded enzymes and the effect of external pH on their gene expression (Bistué et al., 2007; Franklin and Clarke, 2001; Jeong et al., 2020; Rather et al., 1997; Vetting et al., 2008). Considering the dependence of metallo-beta-lactamases functioning on Zn availability (Bounaga et al., 1998; Rasia and Vila, 2002; Tsang et al., 2005), factors influencing Zn mobility (and thus, bioavailability) such as soil pH, organic matter, and clay content, should be considered alongside Zn concentrations for a better prediction of the distribution of metallo-beta-lactamase-encoding ARGs.

Especially tight links with the environment were revealed for chromosomal ARG, among which are those encoding efflux pumps providing antimicrobial efflux, osmoregulation, pH homeostasis, etc. (Li et al., 2015). While a single genome may encode a wide variety of efflux systems, they are evolutionarily conserved (Górecki and McEvoy, 2020)

and their production and functioning is context-dependent (Alav et al., 2021; Li et al., 2015; Masi et al., 2007; Teelucksingh et al., 2020) illustrating a strong multifactorial environmental pressure on them. Studies based on experimental phenotypic regulation of MdtEF, AcrAB (Schaffner et al., 2021), EmrB, MdtB, and TolC (Deininger et al., 2011) well explain the mechanisms driving the distribution of corresponding genes across a soil pH gradient revealed in our work (see Fig. 4**C**, **table S10**). Thus, depending on the availability of information, either the ecological niche of the gene carriers or the encoded structural properties can serve as guidelines for chromosomally encoded ARGs.

5. Conclusion

Metagenomic analysis of geographically and environmentally extensive soil samples reveals consistent links between environmental characteristics and microbial community potential for synthesizing and resisting diverse antibiotics. The links illustrate that antibiotic synthesis and resistance can be viewed as functional traits subject to environmental selection. This concept offers specific guidelines for environmental monitoring, establishing baseline levels of resistance factors and predicting the dissemination of antibiotic resistance in response to the urgent need for measures to control it (Access to Medicines Foundation, 2020; Bengtsson-Palme et al., 2023; Environment, 2017; Larsson and Flach, 2022).

Here, following the central concept of trait-based ecology we propose general principles (reasonably simplified relative to the immense complexity of ecological relationships within natural microbial communities (Nesme and Simonet, 2015)) that may ease ABG and ARG monitoring.

1. Antibiotic biosynthesis is environmentally restricted. Consideration of environmental characteristics (such as abiotic properties and target organisms' abundance) aids in predicting the potential for antibiotic biosynthesis in specific environments, increasing the chance to find an antibiotic with desired properties.

2. Antibiotic resistance is linked to antibiotic biosynthesis. Predicting the abundance of ARGs can be guided by the microbial community's potential for the biosynthesis of corresponding antibiotics, which is especially important in setting the baseline levels of resistance.

3. The properties of structures encoded by ARGs (such as optimum conditions for stability, expression, and activity, fitness cost) influence the distribution of ARGs across different environments.

4. The distribution of taxonomically specific ARGs can be predicted based on the ecological niche of the carrier taxon (its geographical range or distribution across environmental gradients).

Importantly, among the various wildlife components, we highlight the role of the soil microbiome as a vast, continuous reservoir and proliferator of ARGs and ABGs, with which humans are in constant contact. It is likely that a trait-based ecological framework could also be applicable to other wildlife components, particularly animals, important vectors for pathogens and ARGs transmission (Larsson and Flach, 2022). The main limitation of the work includes scarcity and fragmentation of information regarding antibiotic and resistance factor properties. It means that collaborative efforts from specialists in biochemistry, physiology, microbial ecology, and evolution are required to develop the principles further and apply them in policymaking. In this regard, deep-learning approaches linking a drug molecule's structure with its properties are promising (Wong et al., 2024).

Declaration of generative AI and AI-assisted technologies in the writing process

The authors used DALL-E 3, a generative AI model developed by OpenAI, to create the background of the graphical abstract for this paper. Following the use of this tool, the authors carefully reviewed and edited the content as necessary, ensuring its relevance and suitability within the context of the manuscript. The authors take full responsibility for the content of the publication.

CRediT authorship contribution statement

Olesya Dulya: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Vladimir Mikryukov:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Daniil V. Shchepkin:** Writing – review & editing, Writing – original draft, Investigation. **Mari Pent:** Writing – review & editing, Methodology. **Heidi Tamm:** Writing – review & editing, Methodology. **Massimo Guazzini:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis. **Panos Panagos:** Writing – review & editing, Resources, Methodology, Funding acquisition. **Arwyn Jones:** Writing – review & editing, Resources, Project administration, Funding acquisition. **Alberto Orgiazzi:** Writing – review & editing, Resources, Project administration, Funding acquisition, Data curation. **Fabio Marroni:** Writing – review & editing, Supervision, Methodology. **Mohammad Bahram:** Writing – review & editing, Supervision, Methodology. **Leho Tedersoo:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All sampling protocols from the LUCAS survey are accessible at the European Soil Data Centre portal (https://ec.europa.eu/eurostat/web/lucas/data/primary-data/2018). The source code used for data analyses in this study is available for public access on GitHub https://github.com/Mycology-Microbiology-Center/LUCAS_ABRG (DOI: 10.5281/zenodo.10039980). DNA sequence data have been deposited in the NCBI Sequence Read Archive (SRA) with the primary accession code PRJNA1032917 (BioSamples SAMN38018670 – SAMN38019327).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envint.2024.108917.

References

Abramova, A., Berendonk, T.U., Bengtsson-Palme, J., 2023. A global baseline for qPCR-determined antimicrobial resistance gene prevalence across environments. Environ. Int. 178, 108084 https://doi.org/10.1016/j.envint.2023.108084.

Access to Medicines Foundation, 2020. 2020 Antimicrobial Resistance Benchmark [WWW Document]. URL https://accesstomedicinefoundation.org/resource/2020-antimicrobial-resistance-benchmark (accessed 8.18.23).

Alav, I., Kobylka, J., Kuth, M.S., Pos, K.M., Picard, M., Blair, J.M.A., Bavro, V.N., 2021. Structure, assembly, and function of tripartite efflux and type 1 secretion systems in gram-negative bacteria. Chem. Rev. 121, 5479–5596. https://doi.org/10.1021/acs.chemrev.1c00055.

Alberdi, A., Gilbert, M.T.P., 2019. A guide to the application of Hill numbers to DNA-based diversity analyses. Mol. Ecol. Resour. 19, 804–817. https://doi.org/10.1111/1755-0998.13014.

Alcock, B.P., Huynh, W., Chalil, R., Smith, K.W., Raphenya, A.R., Wlodarski, M.A., Edalatmand, A., Petkau, A., Syed, S.A., Tsang, K.K., Baker, S.J.C., Dave, M., McCarthy, M.C., Mukiri, K.M., Nasir, J.A., Golbon, B., Imtiaz, H., Jiang, X., Kaur, K., Kwong, M., Liang, Z.C., Niu, K.C., Shan, P., Yang, J.Y.J., Gray, K.L., Hoad, G.R., Jia, B., Bhando, T., Carfrae, L.A., Farha, M.A., French, S., Gordzevich, R., Rachwalski, K., Tu, M.M., Bordeleau, E., Dooley, D., Griffiths, E., Zubyk, H.L., Brown, E.D., Maguire, F., Beiko, R.G., Hsiao, W.W.L., Brinkman, F.S.L., Van Domselaar, G., McArthur, A.G., 2023. CARD 2023: Expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. Nucleic Acids Res. 51, D690–D699. https://doi.org/10.1093/nar/gkac920.

Allen, H.K., Donato, J., Wang, H.H., Cloud-Hansen, K.A., Davies, J., Handelsman, J., 2010. Call of the wild: Antibiotic resistance genes in natural environments. Nat. Rev. Microbiol. 8, 251–259. https://doi.org/10.1038/nrmicro2312.

Anderson, M.J., 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecol. 26, 32–46. https://doi.org/10.1111/j.1442-9993.2001.01070. pp. x.

Anderson, M.J., 2006. Distance-based tests for homogeneity of multivariate dispersions. Biometrics 62, 245–253. https://doi.org/10.1111/j.1541-0420.2005.00440.x.

Arango-Argoty, G., Garner, E., Pruden, A., Heath, L.S., Vikesland, P., Zhang, L., 2018. DeepARG: A deep learning approach for predicting antibiotic resistance genes from metagenomic data. Microbiome 6, 23. https://doi.org/10.1186/s40168-018-0401-z.

Arel-Bundock V., 2023. Predictions, Comparisons, Slopes, Marginal Means, and Hypothesis Tests [WWW Document]. URL https://vincentarelbundock.github.io/marginaleffects/ (accessed 4.13.23).

Bahram, M., Hildebrand, F., Forslund, S.K., Anderson, J.L., Soudzilovskaia, N.A., Bodegom, P.M., Bengtsson-Palme, J., Anslan, S., Coelho, L.P., Harend, H., Huerta-Cepas, J., Medema, M.H., Maltz, M.R., Mundra, S., Olsson, P.A., Pent, M., Põlme, S., Sunagawa, S., Ryberg, M., Tedersoo, L., Bork, P., 2018. Structure and function of the global topsoil microbiome. Nature 560, 233–237. https://doi.org/10.1038/s41586-018-0386-6.

Baker, J.R., Woolfson, D.N., Muskett, F.W., Stoneman, R.G., Urbaniak, M.D., Caddick, S., 2007. Protein-small molecule interactions in neocarzinostatin, the prototypical enediyne chromoprotein antibiotic. Chembiochem 8, 704–717. https://doi.org/10.1002/cbic.200600534.

McGill, B.J., Enquist, B.J., Weiher, E., Westoby, M., 2006. Rebuilding community ecology from functional traits. Trends Ecol. Evol. 21, 178–185. https://doi.org/10.1016/j.tree.2006.02.002.

Melander, R.J., Melander, C., 2017. The challenge of overcoming antibiotic resistance: An adjuvant approach? ACS Infect. Dis. 3, 559–563. https://doi.org/10.1021/acsinfecdis.7b00071.

Meng, M., Li, Y., Yao, H., 2022. Plasmid-mediated transfer of antibiotic resistance genes in soil. Antibiotics 11, 525. https://doi.org/10.3390/antibiotics11040525.

Mikryukov, V.S., Dulya, O.V., 2018. Fungal communities in organic and mineral soil horizons in an industrially polluted boreal forest. Biodiversity. https://doi.org/10.1080/14888386.2018.1500307.

Miletto, M., Lindow, S.E., 2015. Relative and contextual contribution of different sources to the composition and abundance of indoor air bacteria in residences. Microbiome 3, 61. https://doi.org/10.1186/s40168-015-0128-z.

Miller, D.L., Wood, S.N., 2014. Finite area smoothing with generalized distance splines. Environ. Ecol. Stat. 21, 715–731. https://doi.org/10.1007/s10651-014-0277-4.

Mölder, F., Jablonski, K.P., Letcher, B., Hall, M.B., Tomkins-Tinch, C.H., Sochat, V., Forster, J., Lee, S., Twardziok, S.O., Kanitz, A., Wilm, A., Holtgrewe, M., Rahmann, S., Nahnsen, S., Köster, J., 2021. Sustainable data analysis with Snakemake. F1000Res 10, 33. https://doi.org/10.12688/f1000research.29032.2.

Moon, J.B., Wardrop, D.H., Smithwick, E.A.H., Naithani, K.J., 2019. Fine-scale spatial homogenization of microbial habitats: a multivariate index of headwater wetland complex condition. Ecol. Appl. 29, e01816.

Morar, M., Wright, G.D., 2010. The genomic enzymology of antibiotic resistance. Annu. Rev. Genet. 44, 25–51. https://doi.org/10.1146/annurev-genet-102209-163517.

Murray, C.J.L., Ikuta, K.S., Sharara, F., Swetschinski, L., Robles Aguilar, G., Gray, A., Han, C., Bisignano, C., Rao, P., Wool, E., Johnson, S.C., Browne, A.J., Chipeta, M.G., Fell, F., Hackett, S., Haines-Woodhouse, G., Kashef Hamadani, B.H., Kumaran, E.A. P., McManigal, B., Achalapong, S., Agarwal, R., Akech, S., Albertson, S., Amuasi, J., Andrews, J., Aravkin, A., Ashley, E., Babin, F.-X., Bailey, F., Baker, S., Basnyat, B., Bekker, A., Bender, R., Berkley, J.A., Bethou, A., Bielicki, J., Boonkasidecha, S., Bukosia, J., Carvalheiro, C., Castañeda-Orjuela, C., Chansamouth, V., Chaurasia, S., Chiurchiù, S., Chowdhury, F., Clotaire Donatien, R., Cook, A.J., Cooper, B., Cressey, T.R., Criollo-Mora, E., Cunningham, M., Darboe, S., Day, N.P.J., De Luca, M., Dokova, K., Dramowski, A., Dunachie, S.J., Duong Bich, T., Eckmanns, T., Eibach, D., Emami, A., Feasey, N., Fisher-Pearson, N., Forrest, K., Garcia, C., Garrett, D., Gastmeier, P., Giref, A.Z., Greer, R.C., Gupta, V., Haller, S., Haselbeck, A., Hay, S.I., Holm, M., Hopkins, S., Hsia, Y., Iregbu, K.C., Jacobs, J., Jarovsky, D., Javanmardi, F., Jenney, A.W.J., Khorana, M., Khusuwan, S., Kissoon, N., Kobeissi, E., Kostyanev, T., Krapp, F., Krumkamp, R., Kumar, A., Kyu, H. H., Lim, C., Lim, K., Limmathurotsakul, D., Loftus, M.J., Lunn, M., Ma, J., Manoharan, A., Marks, F., May, J., Mayxay, M., Mturi, N., Munera-Huertas, T., Musicha, P., Musila, L.A., Mussi-Pinhata, M.M., Naidu, R.N., Nakamura, T., Nanavati, R., Nangia, S., Newton, P., Ngoun, C., Novotney, A., Nwakanma, D., Obiero, C.W., Ochoa, T.J., Olivas-Martinez, A., Olliaro, P., Ooko, E., Ortiz-Brizuela, E., Ounchanum, P., Pak, G.D., Paredes, J.L., Peleg, A.Y., Perrone, C., Phe, T., Phommasone, K., Plakkal, N., Ponce-de-Leon, A., Raad, M., Ramdin, T., Rattanavong, S., Riddell, A., Roberts, T., Robotham, J.V., Roca, A., Rosenthal, V.D., Rudd, K.E., Russell, N., Sader, H.S., Saengchan, W., Schnall, J., Scott, J.A.G., Seekaew, S., Sharland, M., Shivamallappa, M., Sifuentes-Osornio, J., Simpson, A.J., Steenkeste, N., Stewardson, A.J., Stoeva, T., Tasak, N., Thaiprakong, A., Thwaites, G., Tigoi, C., Turner, C., Turner, P., van Doorn, H.R., Velaphi, S., Vongpradith, A., Vongsouvath, M., Vu, H., Walsh, T., Walson, J.L., Waner, S., Wangrangsimakul, T., Wannapinij, P., Wozniak, T., Young Sharma, T.E.M.W., Yu, K. C., Zheng, P., Sartorius, B., Lopez, A.D., Stergachis, A., Moore, C., Dolecek, C., Naghavi, M., 2022. Global burden of bacterial antimicrobial resistance in 2019: A systematic analysis. Lancet 399, 629–655. https://doi.org/10.1016/S0140-6736(21)02724-0.

Nesme, J., Cécillon, S., Delmont, T.O., Monier, J.-M., Vogel, T.M., Simonet, P., 2014. Large-scale metagenomic-based study of antibiotic resistance in the environment. Curr. Biol. 24, 1096–1100. https://doi.org/10.1016/j.cub.2014.03.036.

Nesme, J., Simonet, P., 2015. The soil resistome: a critical review on antibiotic resistance origins, ecology and dissemination potential in telluric bacteria. Environ. Microbiol. 17, 913–930. https://doi.org/10.1111/1462-2920.12631.

Nguyen, T.T.H., Myrold, D.D., Mueller, R.S., 2019. Distributions of extracellular peptidases across prokaryotic genomes reflect phylogeny and habitat. Front. Microbiol. 10.

Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., Caceres, M.D., Durand, S., Evangelista, H.B.A., FitzJohn, R., Friendly, M., Furneaux, B., Hannigan, G., Hill, M.O., Lahti, L., McGlinn, D., Ouellette, M.-H., Cunha, E.R., Smith, T., Stier, A., Braak, C.J.F.T., Weedon, J., 2023. vegan: Community Ecology Package.

Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., Fernández-Ugalde, O., 2018. LUCAS soil, the largest expandable soil dataset for Europe: A review. Eur. J. Soil Sci. 69, 140–153. https://doi.org/10.1111/ejss.12499.

Orgiazzi, A., Panagos, P., Fernández-Ugalde, O., Wojda, P., Labouyrie, M., Ballabio, C., Franco, A., Pistocchi, A., Montanarella, L., Jones, A., 2022. LUCAS Soil Biodiversity and LUCAS Soil Pesticides, new tools for research and policy development. Eur. J. Soil Sci. 73, e13299.

Peek, J., Lilic, M., Montiel, D., Milshteyn, A., Woodworth, I., Biggins, J.B., Ternei, M.A., Calle, P.Y., Danziger, M., Warrier, T., Saito, K., Braffman, N., Fay, A., Glickman, M. S., Darst, S.A., Campbell, E.A., Brady, S.F., 2018. Rifamycin congeners kanglemycins are active against rifampicin-resistant bacteria via a distinct mechanism. Nat. Commun. 9, 4147. https://doi.org/10.1038/s41467-018-06587-2.

Philippon, A., Slama, P., Dény, P., Labia, R., 2016. A structure-based classification of class A β-lactamases, a broadly diverse family of enzymes. Clin. Microbiol. Rev. 29, 29. https://doi.org/10.1128/CMR.00019-15.

R Core Team, 2023. R: A Language and Environment for Statistical Computing.

Raggi, L., Pacicco, L.C., Caproni, L., Álvarez-Muñiz, C., Annamaa, K., Barata, A.M., Batir-Rusu, D., Díez, M.J., Heinonen, M., Holubec, V., Kell, S., Kutnjak, H., Maierhofer, H., Poulsen, G., Prohens, J., Ralli, P., Rocha, F., Rubio Teso, M.L., Sandru, D., Santamaria, P., Sensen, S., Shoemark, O., Soler, S., Sträjeru, S., Thormann, I., Weibull, J., Maxted, N., Negri, V., 2022. Analysis of landrace cultivation in Europe: A means to support in situ conservation of crop diversity. Biol. Conserv. 267, 109460 https://doi.org/10.1016/j.biocon.2022.109460.

Rasia, R.M., Vila, A.J., 2002. Exploring the role and the binding affinity of a second zinc equivalent in B. cereus metallo-beta-lactamase. Biochemistry 41, 1853–1860. https://doi.org/10.1021/bi010933n.

Rather, P.N., Parojcic, M.M., Paradise, M.R., 1997. An extracellular factor regulating expression of the chromosomal aminoglycoside 2'-N-acetyltransferase of Providencia stuartii. Antimicrob. Agents Chemother. 41, 1749–1754. https://doi.org/10.1128/AAC.41.8.1749.

Santos-Aberturas, J., Vior, N.M., 2022. Beyond soil-dwelling actinobacteria: Fantastic antibiotics and where to find them. Antibiotics (Basel) 11, 195. https://doi.org/10.3390/antibiotics11020195.

Schaffner, S.H., Lee, A.V., Pham, M.T.N., Kassaye, B.B., Li, H., Tallada, S., Lis, C., Lang, M., Liu, Y., Ahmed, N., Galbraith, L.G., Moore, J.P., Bischof, K.M., Menke, C.C., Slonczewski, J.L., 2021. Extreme acid modulates fitness trade-offs of multidrug efflux pumps MdtEF-TolC and AcrAB-TolC in Escherichia coli K-12. Appl. Environ. Microbiol. 87, e00724–e00821. https://doi.org/10.1128/AEM.00724-21.

Siles, J.A., Vera, A., Díaz-López, M., García, C., van den Hoogen, J., Crowther, T.W., Eisenhauer, N., Guerra, C., Jones, A., Orgiazzi, A., Delgado-Baquerizo, M., Bastida, F., 2023. Land-use- and climate-mediated variations in soil bacterial and fungal biomass across Europe and their driving factors. Geoderma 434, 116474. https://doi.org/10.1016/j.geoderma.2023.116474.

Smith, L.C., Orgiazzi, A., Eisenhauer, N., Cesarz, S., Lochner, A., Jones, A., Bastida, F., Patoine, G., Reitz, T., Buscot, F., Rillig, M.C., Heintz-Buschart, A., Lehmann, A., Guerra, C.A., 2021. Large-scale drivers of relationships between soil microbial properties and organic carbon across Europe. Glob. Ecol. Biogeogr. 30, 2070–2083. https://doi.org/10.1111/geb.13371.

Szoboszlay, M., Dohrmann, A.B., Poeplau, C., Don, A., Tebbe, C.C., 2017. Impact of land-use change and soil organic carbon quality on microbial diversity in soils across Europe. FEMS Microbiol. Ecol. 93, fix146. https://doi.org/10.1093/femsec/fix146.

Teelucksingh, T., Thompson, L.K., Cox, G., 2020. The evolutionary conservation of Escherichia coli drug efflux pumps supports physiological functions. J. Bacteriol. 202, e00367–e00420. https://doi.org/10.1128/JB.00367-20.

Trivedi, P., Delgado-Baquerizo, M., Anderson, I.C., Singh, B.K., 2016. Response of soil properties and microbial communities to agriculture: Implications for primary productivity and soil health indicators. Front. Plant Sci. 7, 990. https://doi.org/10.3389/fpls.2016.00990.

Tsang, W.Y., Ahmed, N., Hinchliffe, P.S., Wood, J.M., Harding, L.P., Laws, A.P., Page, M. I., 2005. Different transition-state structures for the reactions of beta-lactams and analogous beta-sultams with serine beta-lactamases. J. Am. Chem. Soc. 127, 17556–17564. https://doi.org/10.1021/ja056124z.

Velazquez-Meza, M.E., Galarde-López, M., Carrillo-Quiróz, B., Alpuche-Aranda, C.M., 2022. Antimicrobial resistance: One Health approach. Vet. World 15, 743–749. https://doi.org/10.14202/vetworld.2022.743-749.

Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, Statistics and Computing. Springer, New York, NY. Doi: 10.1007/978-0-387-21706-2.

Vetting, M.W., Park, C.H., Hegde, S.S., Jacoby, G.A., Hooper, D.C., Blanchard, J.S., 2008. Mechanistic and structural analysis of aminoglycoside N-acetyltransferase AAC(6')-Ib and its bifunctional, fluoroquinolone-active AAC(6')-Ib-cr variant. Biochemistry 47, 9825–9835. https://doi.org/10.1021/bi800664x.

Vorobeichik, E.L., Korkina, I.N., 2023. A bizarre layer cake: Why soil animals recolonizing polluted areas shape atypical humus forms. Sci. Total Environ. 904, 166810 https://doi.org/10.1016/j.scitotenv.2023.166810.

Waglechner, N., Culp, E.J., Wright, G.D., 2021. Ancient Antibiotics, Ancient Resistance. EcoSal Plus 9, eESP-0027-2020. Doi: 10.1128/ecosalplus.ESP-0027-2020.

Waley, S.G., 1975. The pH-dependence and group modification of beta-lactamase I. Biochem. J. 149, 547–551.

Wong, F., Zheng, E.J., Valeri, J.A., Donghia, N.M., Anahtar, M.N., Omori, S., Li, A., Cubillos-Ruiz, A., Krishnan, A., Jin, W., Manson, A.L., Friedrichs, J., Helbig, R., Hajian, B., Fiejtek, D.K., Wagner, F.F., Soutter, H.H., Earl, A.M., Stokes, J.M., Renner, L.D., Collins, J.J., 2024. Discovery of a structural class of antibiotics with explainable deep learning. Nature 626, 177–185. https://doi.org/10.1038/s41586-023-06887-8.

Wood, S.N., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J. R. Stat. Soc. Ser. B (Stat. Methodol.) 73, 3–36. https://doi.org/10.1111/j.1467-9868.2010.00749.x.

Wright, G.D., 2007. The antibiotic resistome: the nexus of chemical and genetic diversity. Nat. Rev. Microbiol. 5, 175–186. https://doi.org/10.1038/nrmicro1614.

Zein, N., Reiss, P., Bernatowicz, M., Bolgar, M., 1995. The proteolytic specificity of the natural enediyne-containing chromoproteins is unique to each chromoprotein. Chem. Biol. 2, 451–455. https://doi.org/10.1016/1074-5521(95)90262-7.

Zhuang, M., Achmon, Y., Cao, Y., Liang, X., Chen, L., Wang, H., Siame, B.A., Leung, K.Y., 2021. Distribution of antibiotic resistance genes in the environment. Environ. Pollut. 285, 117402 https://doi.org/10.1016/j.envpol.2021.117402.