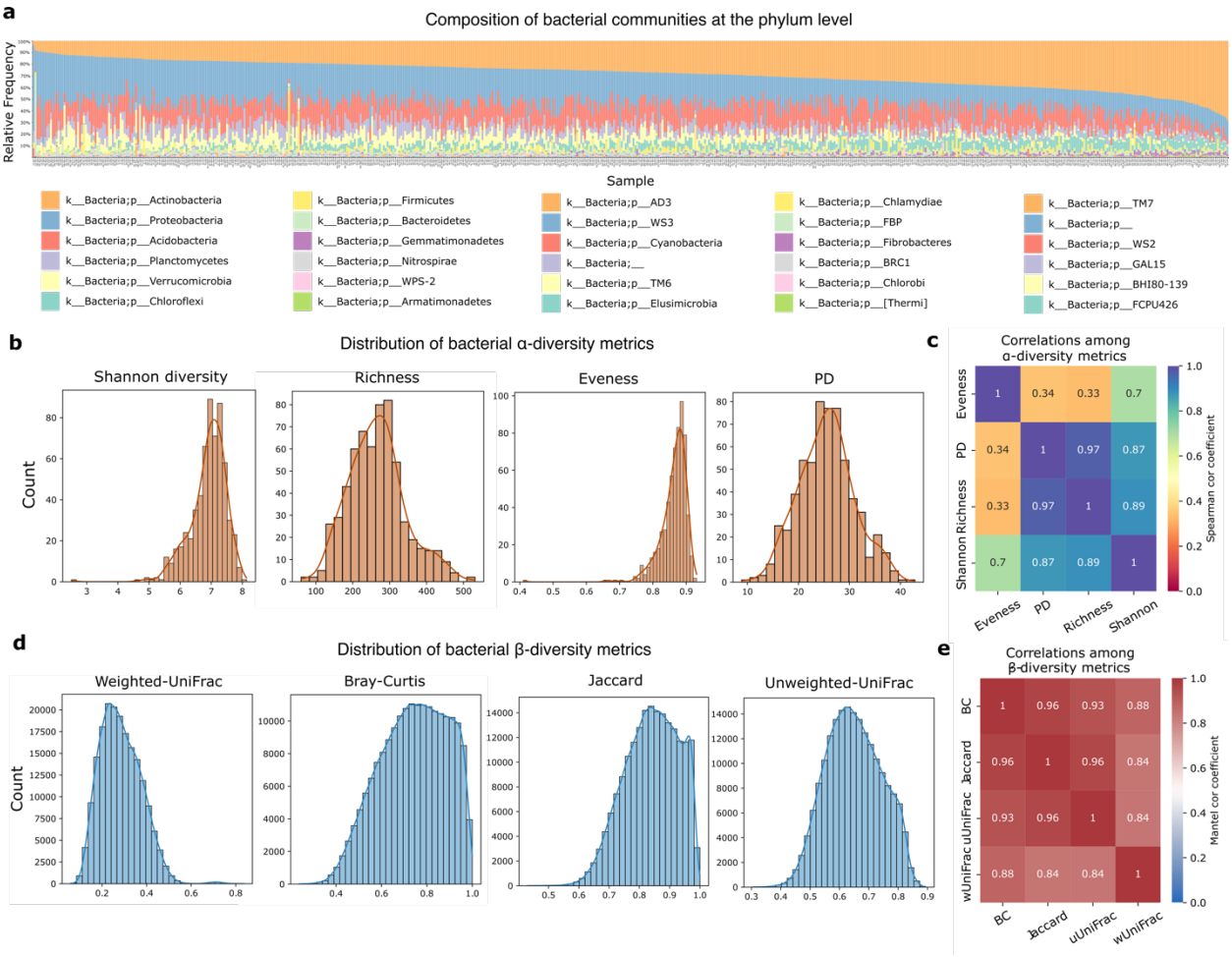
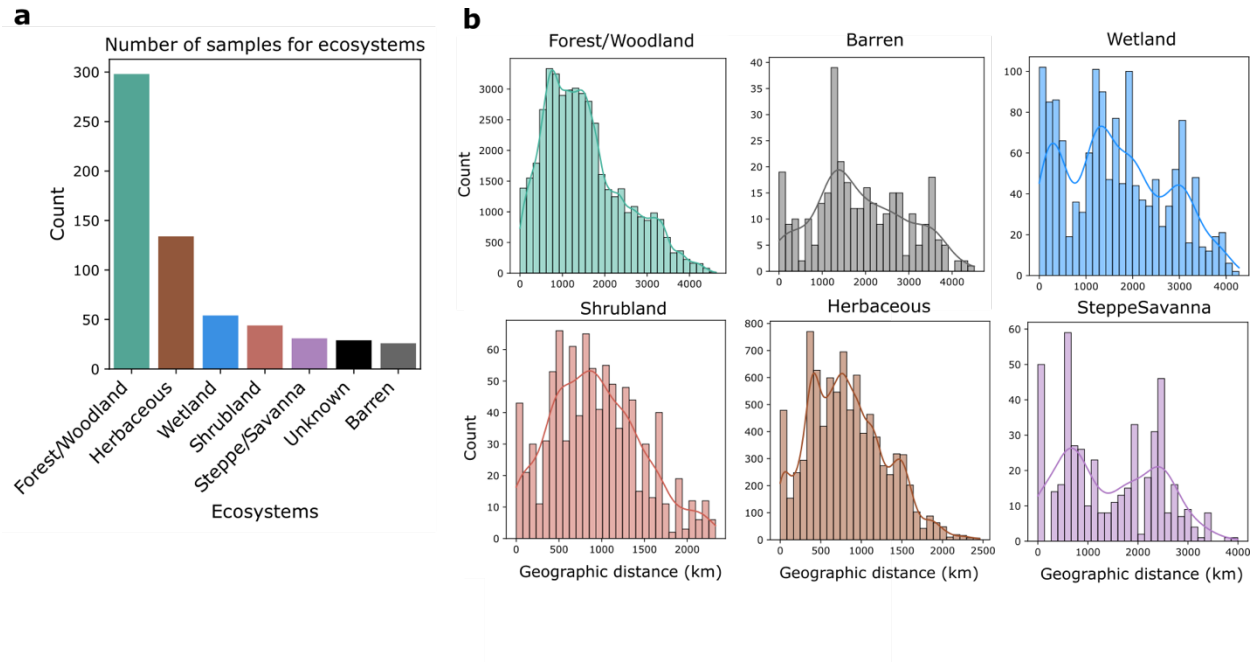
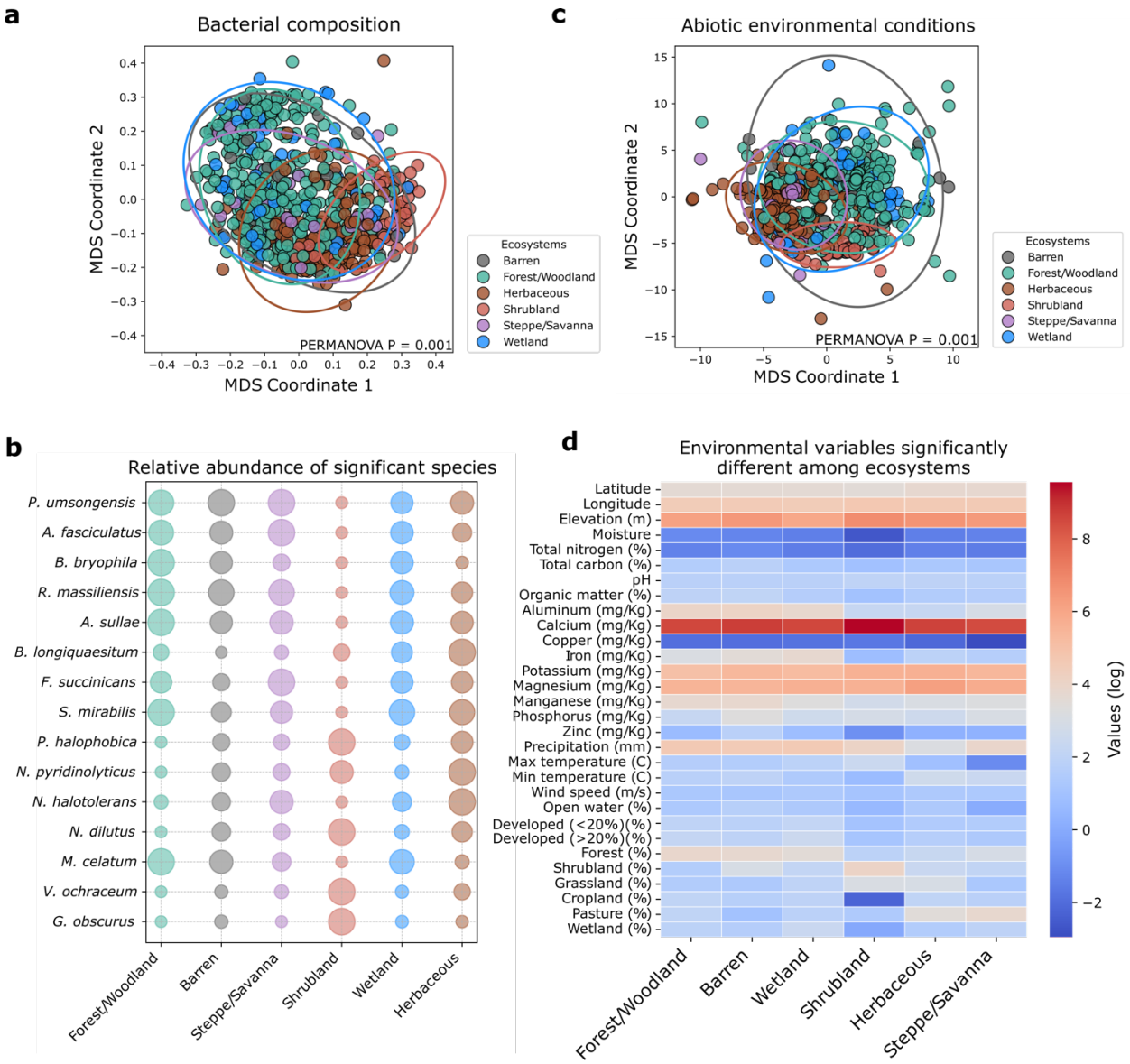
Supplementary figures



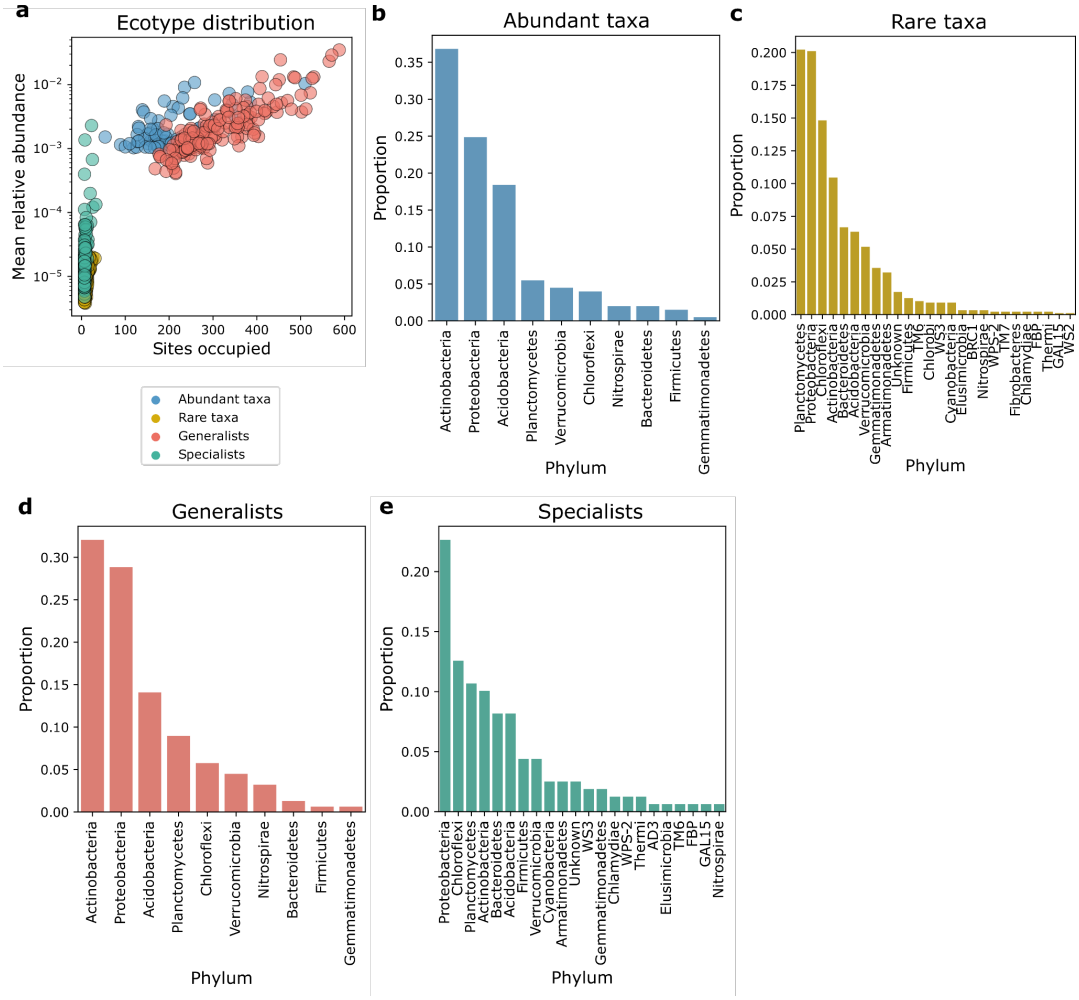
Supplementary Fig. 1 | Composition and diversity of bacterial communities across the US. a, Composition of bacterial communities for each sample at the phylum level. **b,** Histograms showing the distribution of bacterial α-diversity metrics, including Shannon-Wiener index, richness, evenness, and Faith's phylogenetic diversity (PD). **c,** Pairwise Spearman's rank correlation between α-diversity metrics. **d,** Histograms showing the distribution of bacterial β-diversity metrics, including weighte-UniFrac, Bray-Curtis, Jaccard, and unweighted-UniFrac distances. **e,** Pairwise Mantel correlation between β-diversity metrics. BC: Bray-Curtis distance; uUniFrac: unweighted UniFrac distance; wUniFrac: weighted UniFrac distance. Cells are annotated by correlation coefficient in (**c**) and (**e**).



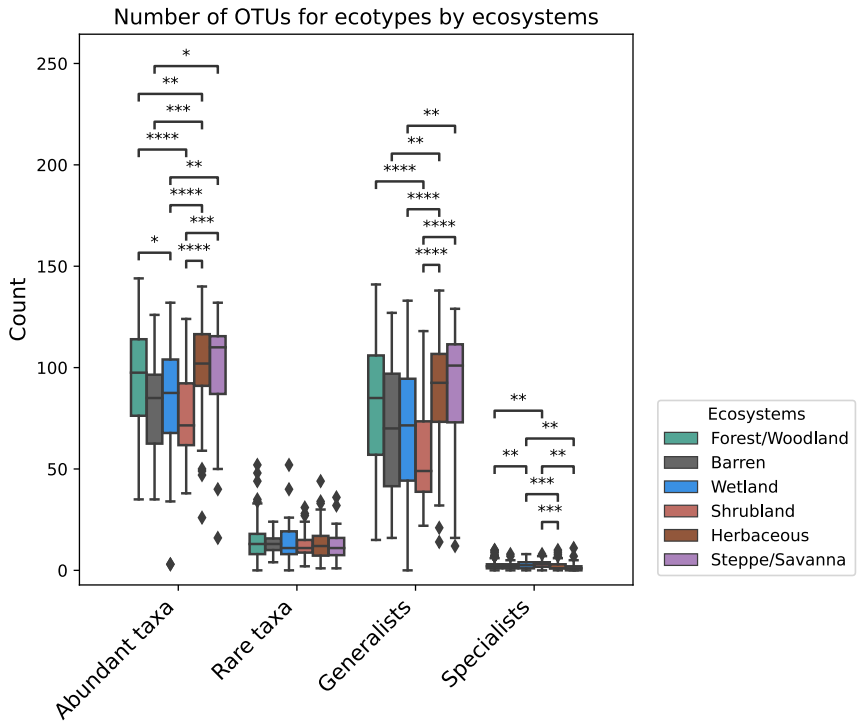
Supplementary Fig. 2 | Distribution of samples for ecosystems. a, Number of samples collected from each ecosystem. Bars are sorted in descending order by prevalence. **b,** Histograms showing the distribution of pairwise geographic distance (km) between samples for each ecosystem.



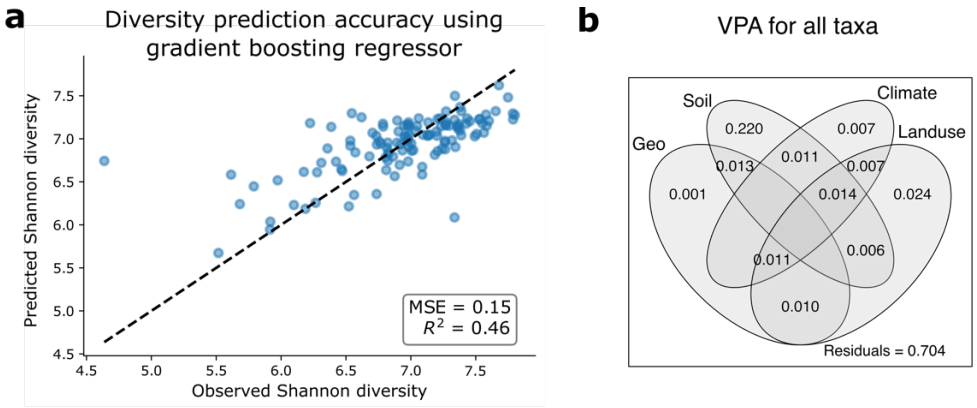
Supplementary Fig. 3 | Bacterial composition and environmental characteristics of ecosystems. a, Multidimensional scaling (MDS) analysis for bacterial composition of ecosystems. **b,** Bubble plot showing the mean relative abundance of bacterial species significantly different among ecosystems (adjusted Kruskal-Wallis (KW) $P < 0.05$). Mean relative abundance was standardized to enhance the comparison of each species among ecosystems. Thus, circle sizes do not represent true mean relative abundance. **c,** MDS analysis for abiotic environmental composition of ecosystems. PERMONOVA $P < 0.05$ indicates significant differences among ecosystems and the size of the ellipse is determined by two times the standard deviation from the mean in (**a**) and (**c**). **d,** Environmental variables significantly different among ecosystems (adjusted KW $P < 0.05$). Data is log-transformed.



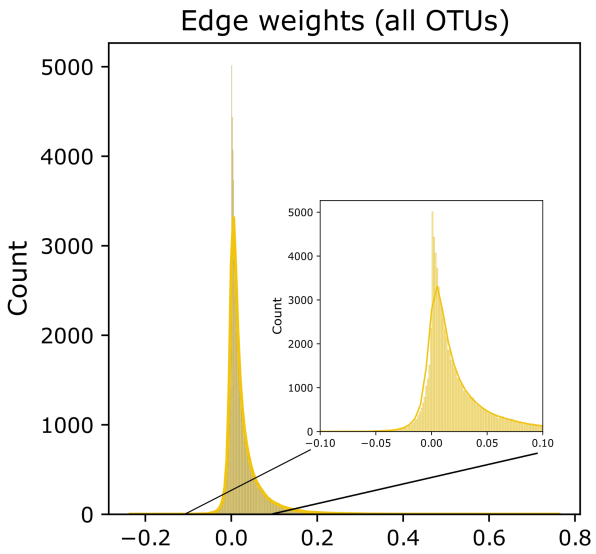
Supplementary Fig. 4 | Distribution and composition of ecotypes. a, Distribution of ecotypes based on the number of sites occupied and mean relative abundance. **b,c,d,e,** Proportion of (**b**) abundant taxa, (**c**) rare taxa, (**d**) generalists, and (**e**) specialists representing different phyla, sorted in descending order.



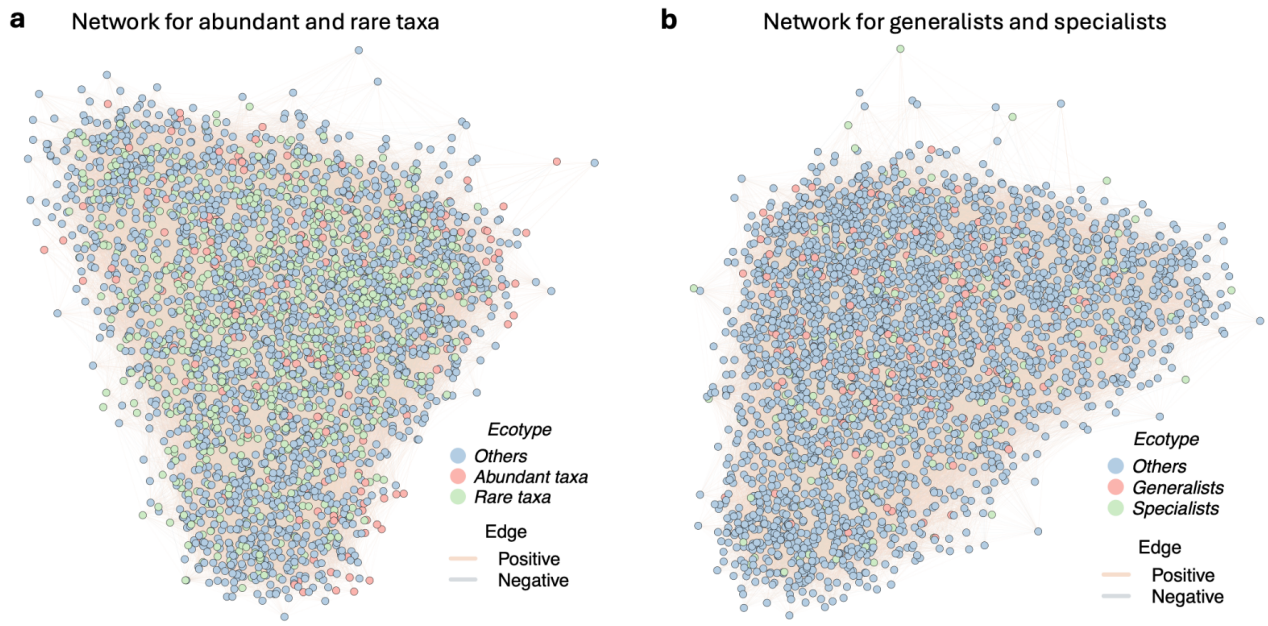
Supplementary Fig. 5 | Number of OTUs representing each ecotype compared among ecosystems. ****, ***, **, and * denotes adjusted two-sided Mann-Whitney *U* $P < 0.0001$, 0.001, 0.01, and 0.05, respectively. Only pairs with adjusted $P < 0.05$ are shown. Box plots show the interquartile range (IQR), with the line representing the median and whiskers extending to 1.5 times the IQR.



Supplementary Fig. 6 | Associations between environmental variables and bacterial diversity. a, Venn diagram of variation partitioning analysis (VPA) showing the variation of the diversity of all taxa explained by the environmental variable groups. Residuals indicate unexplained variation. **b,** Relationship between observed and predicted bacterial diversity from environmental variables using a gradient boosting regressor. R^2 = 0.46 and a mean squared error (MSE) = 0.15.

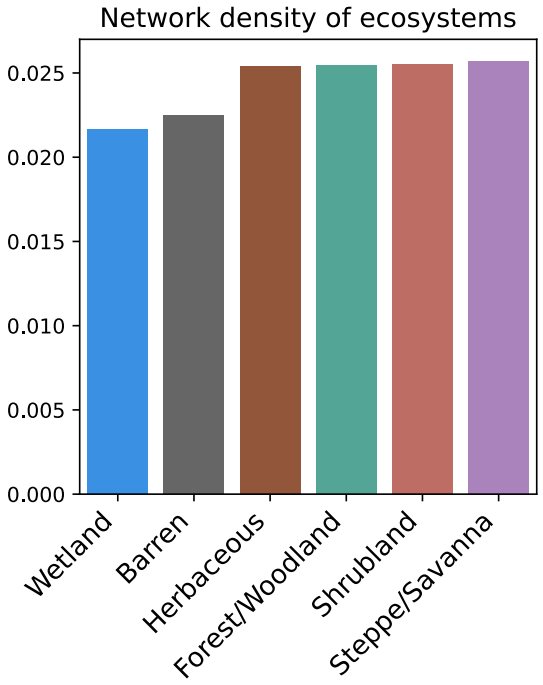


Supplementary Fig. 7 | Histogram showing the distribution of edge weights of the co-occurrence network constructed using all OTUs. The histogram is zoomed in for weights ranging between -0.1 and 0.1. The network is shown in **Fig. 3a**.

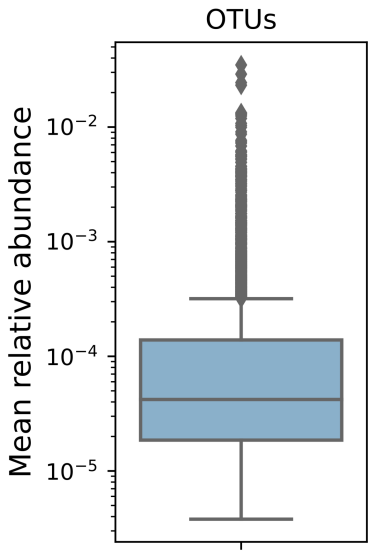


Supplementary Fig. 8 | Bacterial co-occurrence network for ecotypes. Nodes are color-coded by abundant, rare tax, and others in (**a**), and by generalists, specialists, and others in (**b**). The network was constructed using all OTUs. Positive and negative edges are shown in orange and gray, respectively.

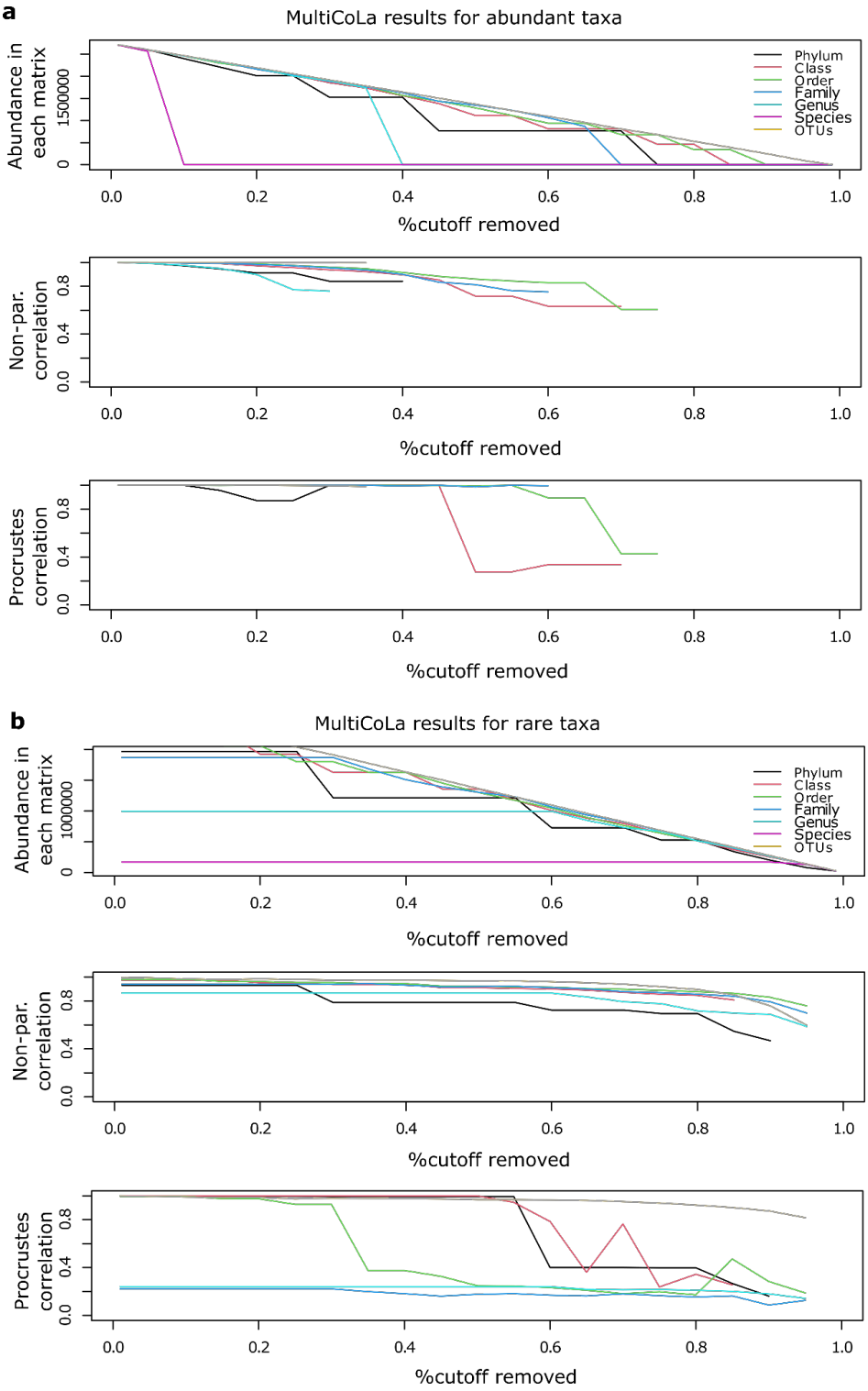
Supplementary Fig. 9 | Bacterial co-occurrence network for ecosystems. A network constructed using OTUs present in (**a**) forest/woodland, (**b**) shrubland, (**c**) wetland, (**d**), herbaceous, (**e**) steppe/savanna, and (**f**) barren ecosystems. Nodes are color-coded by phyla. Positive and negative edges are shown in orange and gray, respectively.



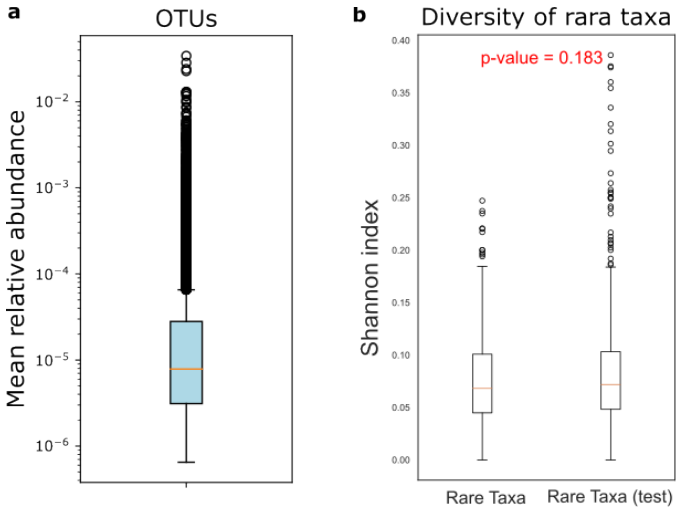
Supplementary Fig. 10 | Density of co-occurrence network constructed using OTUs present in each ecosystem. Bars are sorted in ascending order. The co-occurrence network of each ecosystem is shown in **Supplementary Fig. 9a-f**.



Supplementary Fig. 11 | Mean relative abundance of OTUs across samples. Box plots show the interquartile range (IQR), with the line representing the median and whiskers extending to 1.5 times the IQR.



Supplementary Fig. 12 | MultiCoLa results show the robustness of the mean relative abundance cutoff selection for ecotypes. a, b, Changes of the metrics with %cutoff removed for (**a**) abundant taxa and (**b**) rare taxa at different taxonomic levels.



Supplementary Fig. 13 | Detection of potential bias in classifying rare taxa caused by the method used for removing OTUs. a, Mean relative abundance of OTUs across samples (test dataset). **b,** Diversity of rare taxa compared between the dataset used in this study and the test dataset**.** The dataset used in this study: OTUs present in < 1% of samples and with total reads < 10 were removed. Test dataset: singletons (i.e., OTUs present in one sample with one read sequenced) were removed. Box plots show the interquartile range (IQR), with the line representing the median and whiskers extending to 1.5 times the IQR. *P*, two-sided Mann-Witney *U* test.