Supplementary Information

Life-history strategies of soil microbial communities in an arid ecosystem

Yongjian Chen[1,*], Julia W. Neilson[1], Priyanka Kushwaha[1], Raina M. Maier[1], Albert Barberán[1,*]

[1]Department of Environmental Science, University of Arizona, Tucson, AZ 85721, USA

*Corresponding author:

Yongjian Chen (chenyj@email.arizona.edu) or Albert Barberán (barberan@email.arizona.edu)

Supplementary Information includes:

Supplementary methods

Figures S1 to S7

Table S1

Reference

Supplementary methods

Analyses of taxonomic composition of microbial communities

To assess the taxonomic profiles of bacterial and archaeal communities, the V4 hypervariable region of the 16S rRNA gene was amplified by PCR using the 515-F (GTGCCAGCMGCCGCGGTAA) and 806-R (GGACTACHVGGGTWTCTAAT) primer pair. The primers included the appropriate Illumina adapters with the reverse primers (806-R) also having an error-correcting 12-bp barcode unique to each sample to permit multiplexing of samples. PCR was conducted in 40-µL triplicate reactions per sample, using 3 µl of extracted DNA, 3 µl of each primer, 20 µl of MyFi PCR Mix (Bioline), and 11 µl of water. Negative controls without DNA template were included in each batch of PCR reactions to check for possible contamination. The PCR consisted of an initial denaturing step at 95 °C for 1 min, 35 cycles of amplification (95 °C for 15 s, 60 °C for 15 s and 72 ° C for 15 s) and a final elongation step of 72 °C for 3 min. PCR products were cleaned using an UltraClean PCR Clean-Up Kit (MoBio) and quantified fluorescently with the Quant-It PicoGreen dsDNA Assay kit (ThermoFisher Scientific). The purified PCR products from all samples were pooled together in equimolar concentrations and sequenced on a 2 × 150 bp Illumina MiSeq platform.

Raw reads were first demultiplexed and then processed with DADA2 pipeline [1]. The reads were uniformly trimmed to the same length, and they were filtered by removing reads exceeding the maximum expected error of 2 and reads containing the N symbol. The quality-filtered reads were used to train the error model and dereplicated to obtain unique sequences, which were used to infer amplicon sequence variants (ASVs). Paired-end reads were merged and chimeric sequences were removed. Taxonomic identities were assigned to the ASVs using the RDP classifier [2] with a confidence threshold of 0.8 trained on the SILVA nr version 132 database [3]. Those ASVs identified as chloroplasts and mitochondria were removed.

Shotgun metagenomic sequencing

The library for shotgun metagenomics was constructed using QIASeq FX DNA Library Kit (QIAGEN) according to the manufacturer's protocol. Briefly, 50 ng of DNA from each sample was randomly fragmented with FX Enzyme Mix followed by the Adapter ligation step. Both i5 and i7 adapters contain unique 8 nucleotide barcodes. After removing free adapters from the reaction mixture with AMPure XP magnetic beads, the libraries were purified and the size selection 2-step purification (the negative followed by the positive selection) was performed with AMPure XP magnetic beads. The quality and quantity of all libraries was determined with Agilent 4150 TapeStation DNA bioanalyzer. All samples were shotgun-sequenced on the NextSeq 550 platform with 400M HighOutput 300 cycles sequencing chemistry.

The 35 single-copy genes

The 35 single-copy genes were identified in [4] and summarized in [5]. The COG numbers of the genes and the associated protein names are COG0012 (Predicted GTPase, probable translation factor), COG0016 (Phenylalanyl-tRNA synthetase alpha subunit), COG0048 (Ribosomal protein S12), COG0049 (Ribosomal protein S7), COG0052 (Ribosomal protein S2), COG0080 (Ribosomal protein L11), COG0081 (Ribosomal protein L1), COG0085 (DNA-directed RNA polymerase, beta subunit/140 kD subunit), COG0087 (Ribosomal protein L3), COG0088 (Ribosomal protein L4), COG0090 (Ribosomal protein L2), COG0091 (Ribosomal protein L22), COG0092 (Ribosomal protein S3), COG0093 (Ribosomal protein L14), COG0094 (Ribosomal protein L5), COG0096 (Ribosomal protein S8), COG0097 (Ribosomal protein L6P/L9E), COG0098 (Ribosomal protein S5), COG0099 (Ribosomal protein S13), COG0100 (Ribosomal protein S11), COG0102 (Ribosomal protein L13), COG0103 (Ribosomal protein S9), COG0124 (Histidyl-tRNA synthetase), COG0184 (Ribosomal protein

S15P/S13E), COG0185 (Ribosomal protein S19), COG0186 (Ribosomal protein S17), COG0197 (Ribosomal protein L16/L10E), COG0200 (Ribosomal protein L15), COG0201 (Preprotein translocase subunit SecY), COG0256 (Ribosomal protein L18), COG0495 (Leucyl-tRNA synthetase), COG0522 (Ribosomal protein S4 and related proteins), COG0525 (Valyl-tRNA synthetase), COG0533 (Metal-dependent proteases with possible chaperone activity) and COG0541 (Signal recognition particle GTPase).

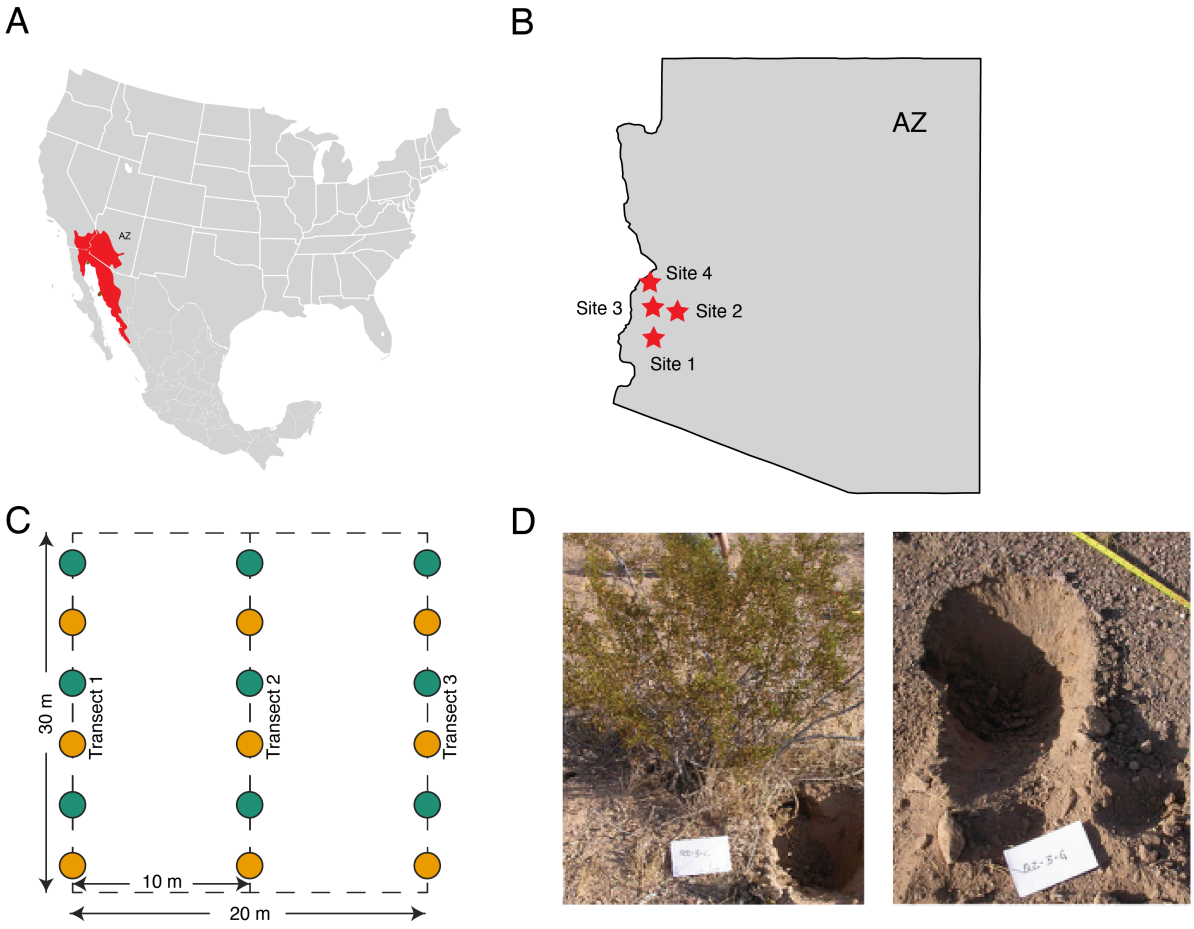

Figure S1. Sampling scheme in this study. (A) The Sonoran Desert. (B) Locations of the four sampling sites (Site 1: 33°22'44.3"N, 114°11'49.9"W; Site 2: 33°43'16.4"N, 113°55'18.6"W; Site 3: 33°45'02.6"N, 114°12'13.4"W; Site 4: 34°03'27.1"N 114°18'30.5"W). (C) Sampling scheme at each site. Green circles denote subsamples collected under vegetation patches, and yellow circles denote subsamples collected in bare ground devoid of vegetation. At each transect, the three subsamples in vegetated areas were pooled into one composite sample, and the three subsamples in bare ground were pooled into another composite sample. (D) Examples of two subsamples collected under vegetation patches (left) and in bare ground (right).

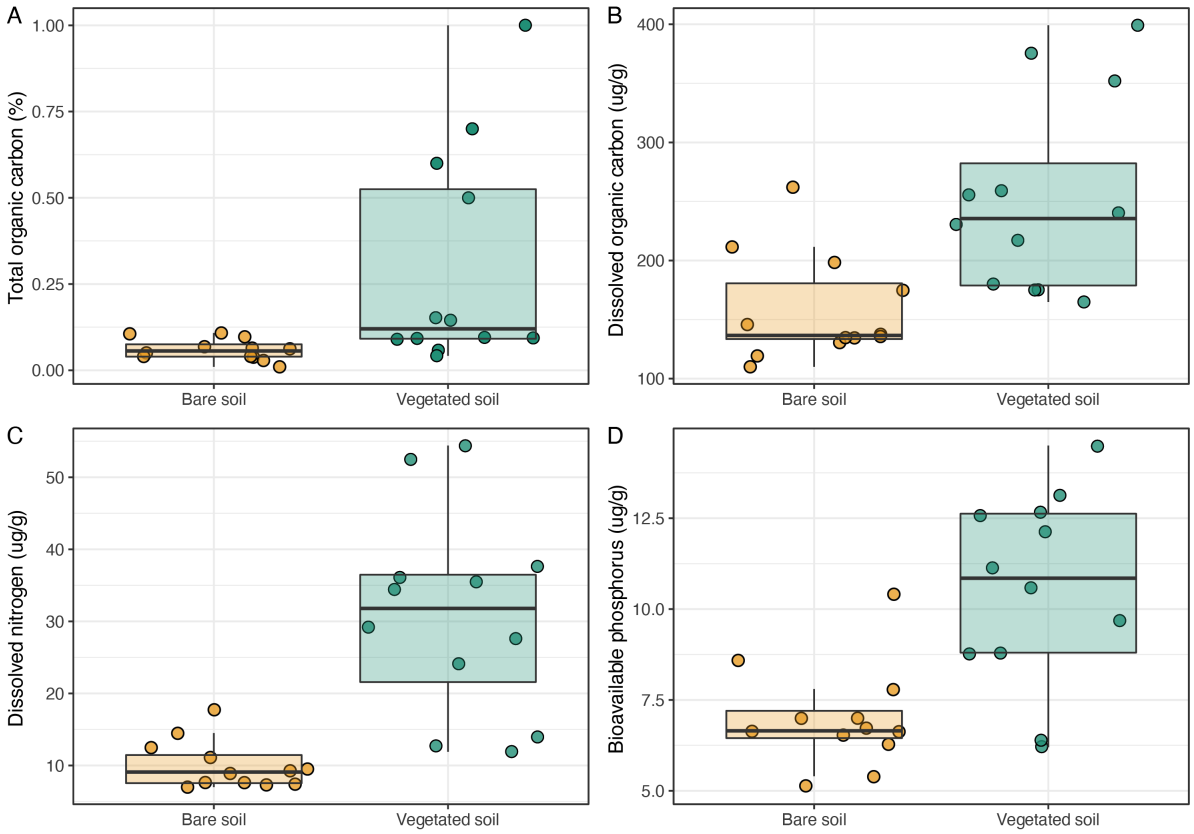

Figure S2. Comparison of soil nutrient availability in bare and vegetated soils. Results of linear mixed-effects models: total organic carbon ($F_{1,19} = 6.60$, $P = 0.02$, $R^2 = 0.22$), dissolved organic carbon ($F_{1,19} = 45.52$, $P = 1.9 \times 10^{-6}$, $R^2 = 0.31$), dissolved nitrogen ($F_{1,19} = 32.91$, $P = 1.6 \times 10^{-5}$, $R^2 = 0.51$), bioavailable phosphorus ($F_{1,19} = 17.06$, $P = 0.0006$, $R^2 = 0.42$). The medians in these box plots are as follows (bare soil vs vegetated soil): total organic carbon (0.06% vs 0.12%), dissolved organic carbon (136.6 ug/g vs 235.5 ug/g), dissolved nitrogen (9.1 ug/g vs 31.8 ug/g), bioavailable phosphorus (6.65 ug/g vs 10.85 ug/g).

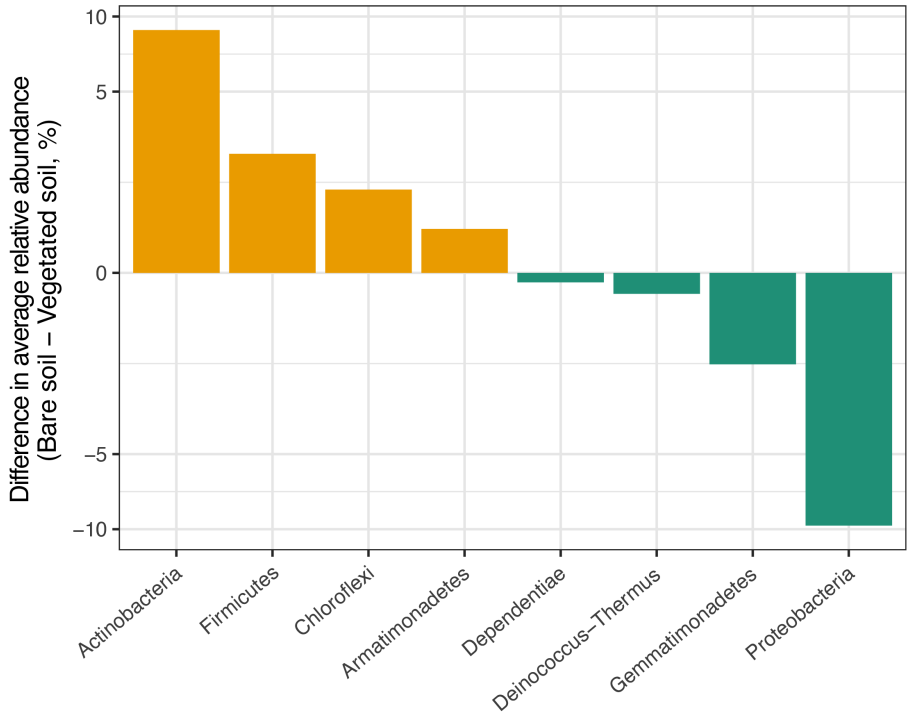

Figure S3. Compositional difference (phylum level) between microbial communities in bare and vegetated soils. Bar plots represent the differences in average relative abundances of phyla between bare and vegetated soils. Only those phyla that were significantly more abundant in bare soils (yellow) or vegetation patches (green) according to *DESeq2* analysis ($P < 0.05$ after FDR correction) are shown.

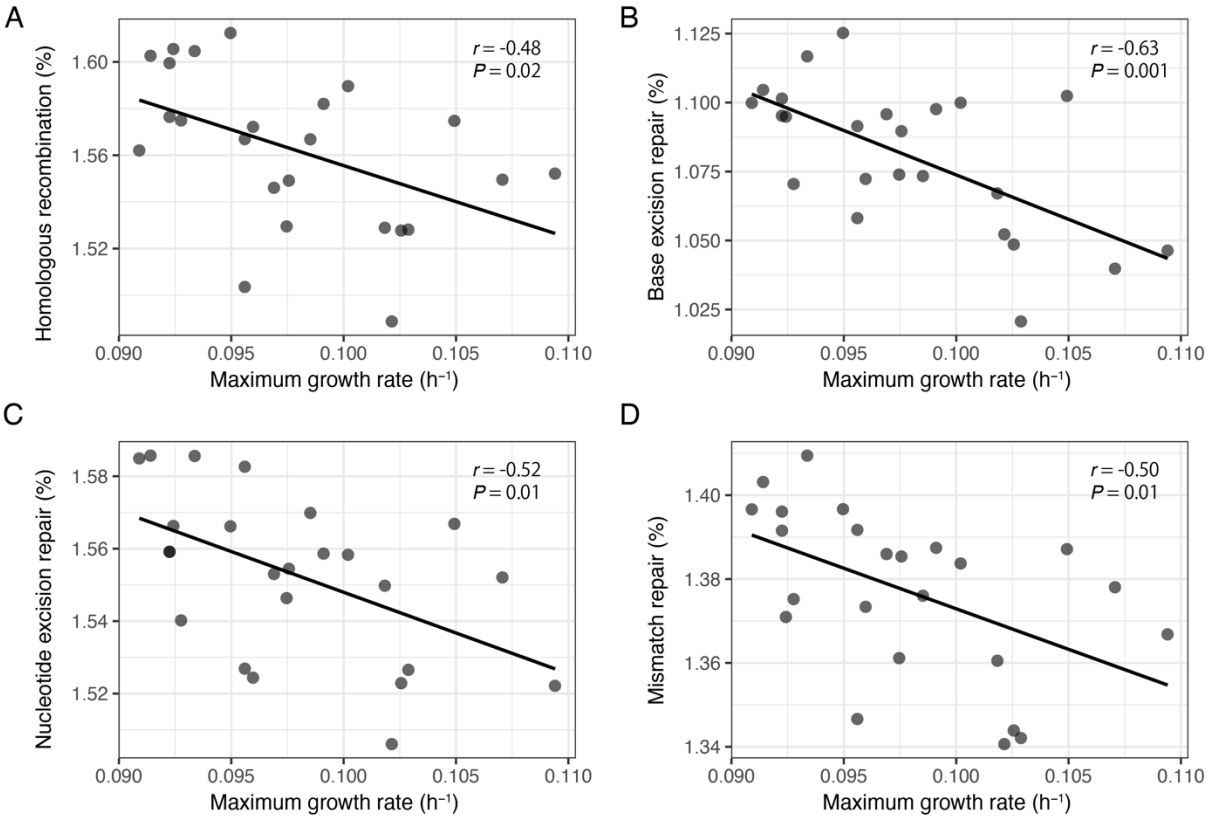

Figure S4. Correlations between maximum growth rate and relative abundances of genes associated with DNA repairing. Results of Pearson correlations are shown.

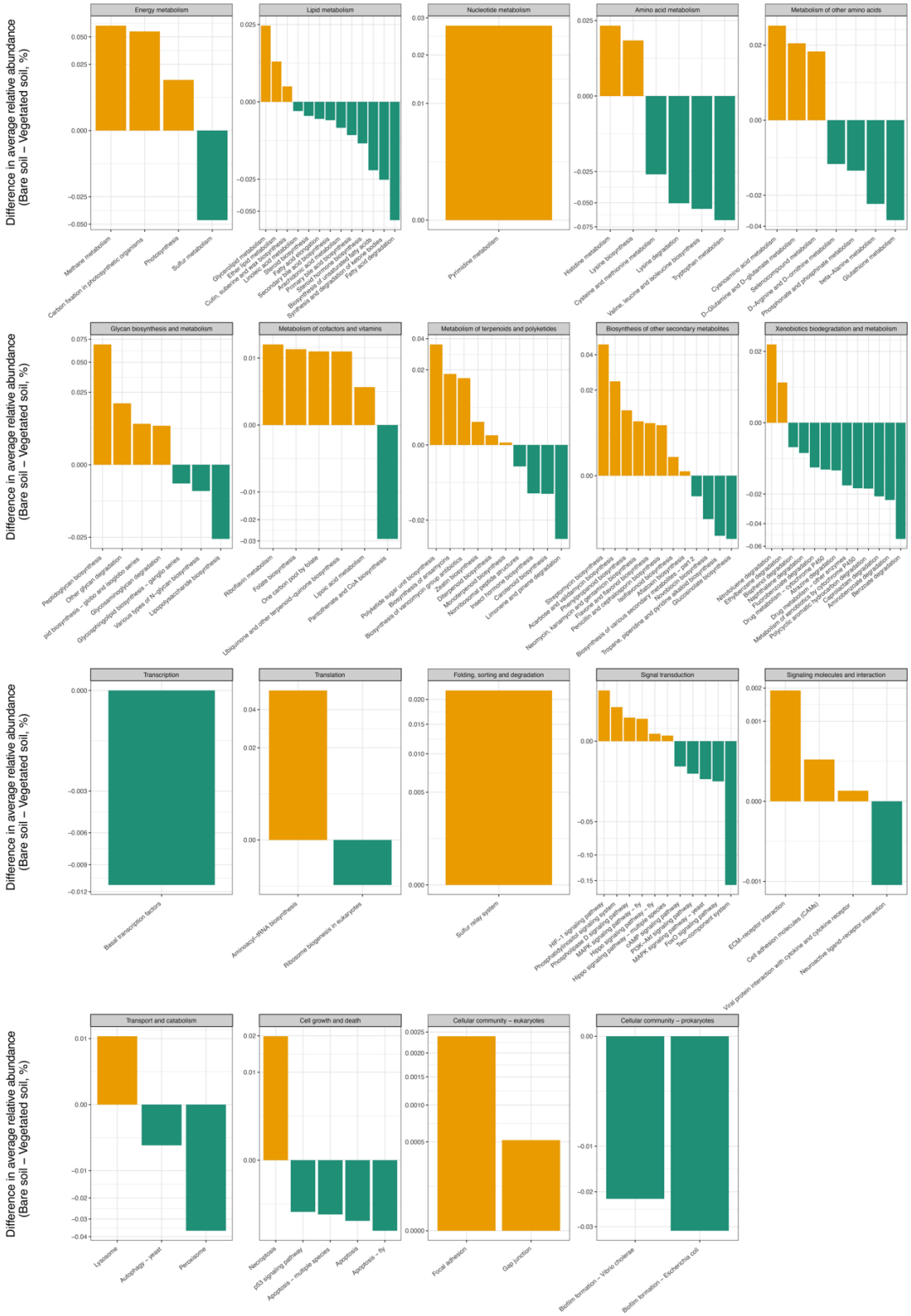

Figure S5. Microbial functional difference (KEGG level3) between bare and vegetated soils. Only those classes that were significantly more abundant in bare soils (yellow) or vegetation patches (green) according to *DESeq2* analysis ($P < 0.05$ after FDR correction) are shown.

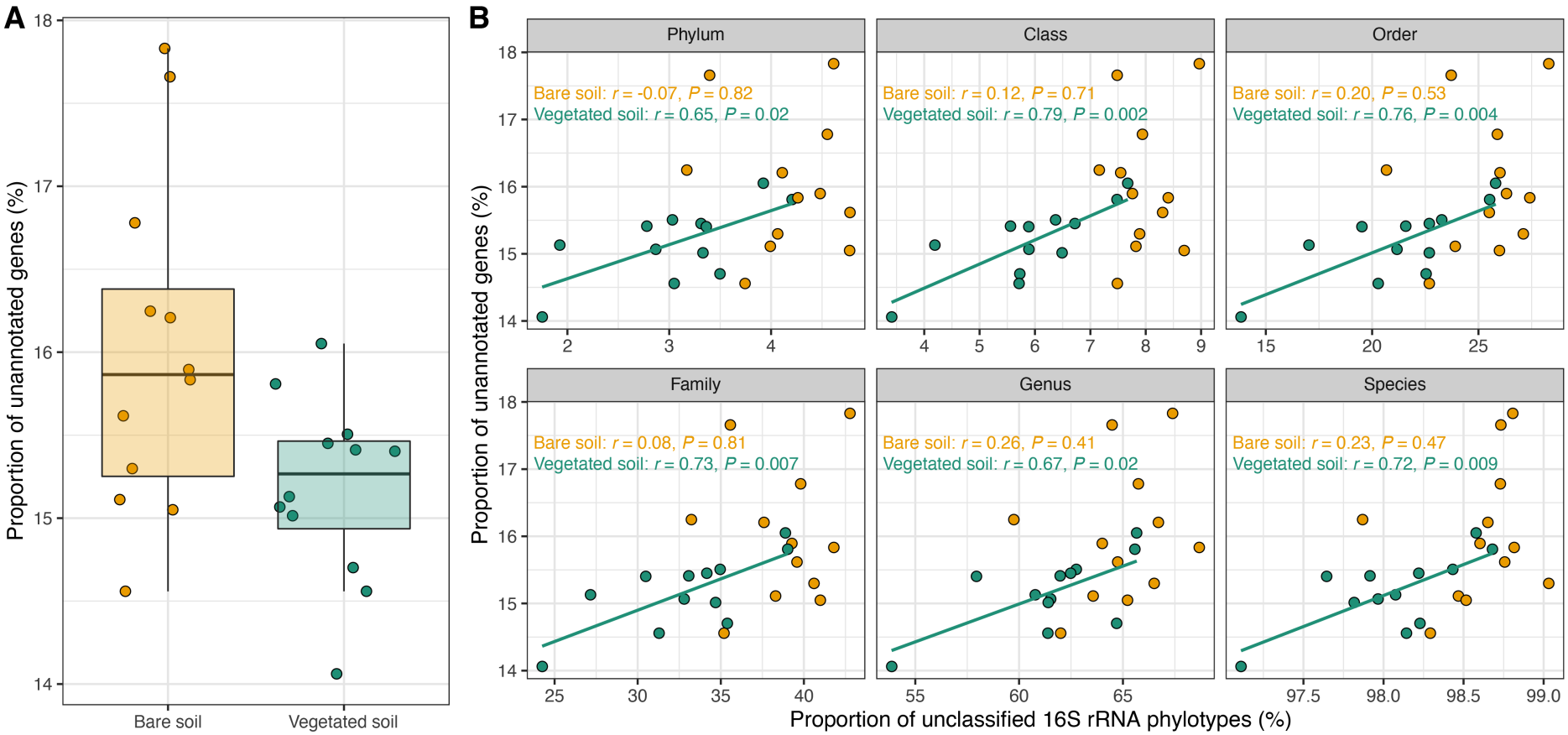

Figure S6. Potential sources of genes without eggNOG annotation. (A) Comparison of the proportion of genes without eggNOG annotation in bare (median = 15.87%) and vegetated soils (median = 15.27%). Results of linear mixed-effects models: $F_{1,19} = 5.91$, $P = 0.02$, $R^2 = 0.21$. (B) Relationship between the proportion of unannotated and the proportion of unclassified 16S rRNA phylotypes. Results of Pearson correlations are shown.

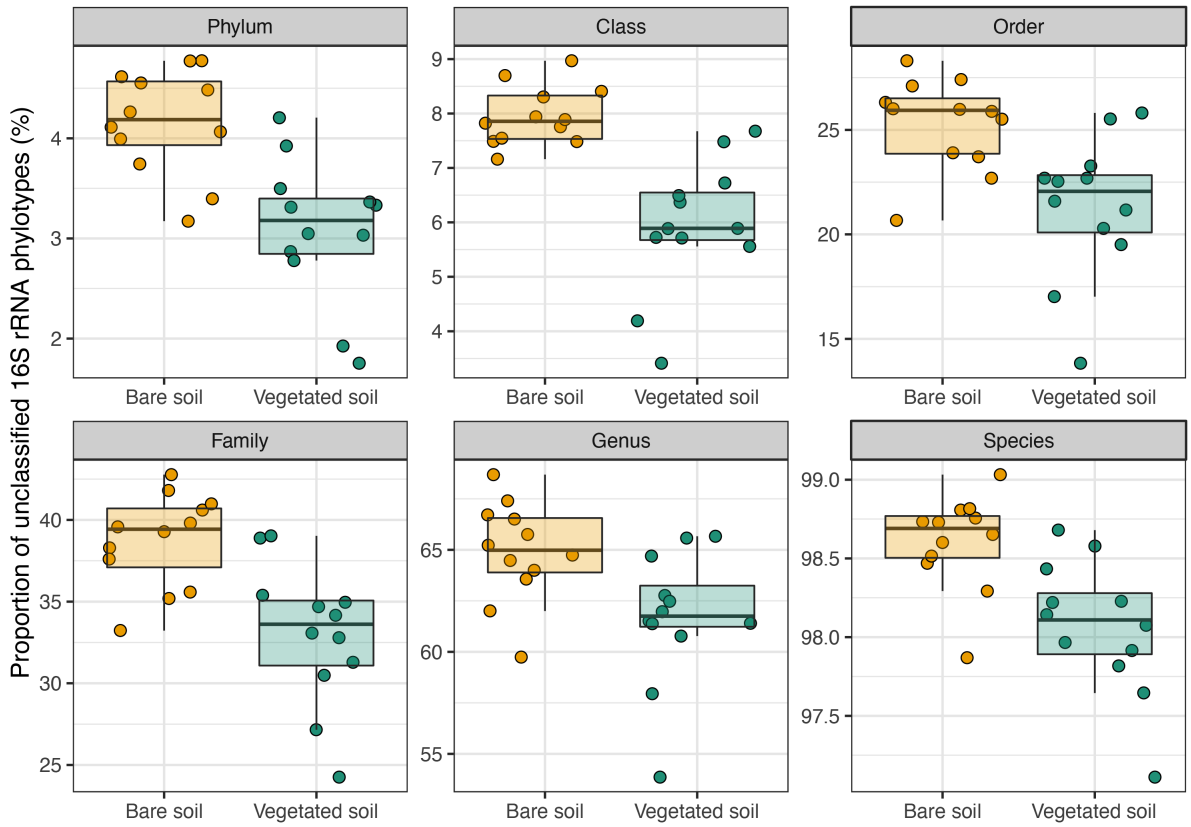

Figure S7. Comparison of the proportion of unclassified 16S rRNA phylotypes in bare and vegetated soils. Results of linear mixed-effects models: phylum ($F_{1,19}$ = 24.65, P = 8.6 × 10^{-5}, R^2 = 0.42), class ($F_{1,19}$ = 39.72, P = 4.8 × 10^{-6}, R^2 = 0.53), order ($F_{1,19}$ = 12.63, P = 0.002, R^2 = 0.33), family ($F_{1,19}$ = 17.43, P = 0.0005, R^2 = 0.38), genus ($F_{1,19}$ = 7.52, P = 0.01, R^2 = 0.25), species ($F_{1,19}$ = 12.74, P = 0.002, R^2 = 0.36). The medians in these box plots are as follows (bare soil vs vegetated soil): phylum (4.19% vs 3.18%), class (7.86% vs 5.89%), order (25.93% vs 22.06%), family (39.43% vs 33.61%), genus (64.98% vs 61.75%), species (98.69% vs 98.11%).

Table S1 Summary of metagenomic assembly

Site	Soil environment	Number of quality-filtered reads	Number of contigs	Longest contig (bp)	N50 (bp)
Site 1	Vegetated soil	162,554,158	1,031,022	47,112	761
	Vegetated soil	170,247,892	1,163,724	62,186	830
	Vegetated soil	199,325,530	1,412,983	69,088	834
	Bare soil	160,697,416	1,175,364	66,451	798
	Bare soil	69,356,892	384,400	9,540	718
	Bare soil	98,708,062	606,838	31,509	774
Site 2	Vegetated soil	167,463,710	1,182,611	71,651	817
	Vegetated soil	143,686,142	1,004,404	46,116	793
	Vegetated soil	69,333,926	408,079	49,440	818
	Bare soil	198,054,004	1,500,745	32,780	788
	Bare soil	356,676,854	3,075,022	64,148	964
	Bare soil	131,089,414	839,181	36,358	753
Site 3	Vegetated soil	149,046,426	963,633	63,852	787
	Vegetated soil	167,950,226	1,060,053	51,086	758
	Vegetated soil	99,200,172	548,597	60,823	809
	Bare soil	216,670,636	1,708,025	40,511	801
	Bare soil	106,564,126	599,647	27,538	715
	Bare soil	146,842,500	1,246,163	37,999	854

Table S1 (Continued)

	Soil environment	Number of quality-filtered reads	Number of contigs	Longest contig (bp)	N50 (bp)
	Vegetated soil	244,959,784	2,038,850	111,147	963
	Vegetated soil	167,979,738	1,214,599	107,649	892
	Vegetated soil	93,788,368	591,906	51,428	805
Site 4	Bare soil	349,946,810	3,012,666	144,877	960
	Bare soil	94,267,886	643,621	77,706	840
	Bare soil	141,476,438	1,093,625	67,066	870

Reference

1. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581–3.
2. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol. 2007;73:5261–7.
3. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res. 2013;41:D590–6.
4. Raes J, Korbel JO, Lercher MJ, von Mering C, Bork P. Prediction of effective genome size in metagenomic samples. Genome Biol. 2007;8:R10.
5. Pereira-Flores E, Glöckner FO, Fernandez-Guerra A. Fast and accurate average genome size and 16S rRNA gene average copy number computation in metagenomic data. BMC Bioinformatics. 2019;20:453.