Journal of Applied Biology & Biotechnology
Available online at http://www.jabonline.in

Diversity and potential function of bacteria from soil and cowpea root nodules across two agroecological zones in Tharaka Nithi, Kenya

Youstina Nasr Salib Habashy[1], Stephen Mwangi Githiri[2], Ezekiel Mugendi Njeru[3]

[1]Department of Molecular Biology and Biotechnology, Pan African University Institute for Basic Sciences, Technology and Innovation, Nairobi, Kenya.
[2]Department of Horticulture and Food Security, Jomo Kenyatta University of Agriculture and Technology, Juja, Kenya.
[3]Department of Biochemistry, Microbiology and Biotechnology, Kenyatta University, Nairobi, Kenya.

Doi: http://doi.org/10.7324/jabb.2026.323367

Supplementary Materials

Diversity and Potential Function of Bacteria from Soil and Cowpea Root Nodules across two Agroecological Zones in Tharaka Nithi, Kenya

Table S1. Sequencing depth statistics

Table S1a. Per-sample sequencing statistics (n = 20; see Methods 2.3).

Sample ID	Group	Total Reads	Denoised Reads	Merged Reads	Non-Chimeric Reads
N1	Non-arid soil	111361	91979	55999	50306
N2	Non-arid soil	106794	87585	49792	44355
N3	Non-arid soil	112695	95245	46935	40723
N4	Non-arid soil	106432	90739	55681	50216
N5	Non-arid soil	128250	109284	70221	60745
NN1	Non-arid nodule	72431	71018	70540	65039
NN2	Non-arid nodule	84920	83265	82651	75834
NN3	Non-arid nodule	92553	91270	90765	83377
NN4	Non-arid nodule	72025	70568	70185	63901
NN5	Non-arid nodule	78327	76864	76455	70224
NS1	Semi-arid nodule	81252	79311	78488	69668
NS2	Semi-arid nodule	97988	96137	95223	87729
NS3	Semi-arid nodule	91007	88846	87480	78750
NS4	Semi-arid nodule	78241	76856	76263	70397
NS5	Semi-arid nodule	57278	55403	54944	48962
S1	Semi-arid soil	97842	82099	43399	38209
S2	Semi-arid soil	105690	86085	44085	37756
S3	Semi-arid soil	104821	86963	48408	41254
S4	Semi-arid soil	84665	70307	38854	32635
S5	Semi-arid soil	122630	100440	53672	46452

Table S1b. Sequencing depth summary by sample group.

Group	n	Lowest	Highest	Mean
Non-arid soil	5	40,723	60,745	49,269
Non-arid nodule	5	63,901	83,377	71,675
Semi-arid soil	5	32,635	46,452	39,261
Semi-arid nodule	5	48,962	87,729	71,101
All samples	20	32,635	87,729	57,827

Figure S1. Rarefaction curves

Rarefaction curves for Chao1, Observed Features, and Shannon diversity across all 20 samples. Most samples approached a plateau across all three metrics, indicating that sequencing depth captured the major patterns of bacterial diversity. Several Chao1 and Observed Features curves remained slightly upward at the highest read depths, suggesting additional sequencing could recover a limited number of rare taxa. Shannon diversity remained consistently lower in nodules than in soil throughout, reflecting genuine biological specialization rather than incomplete sampling.

.

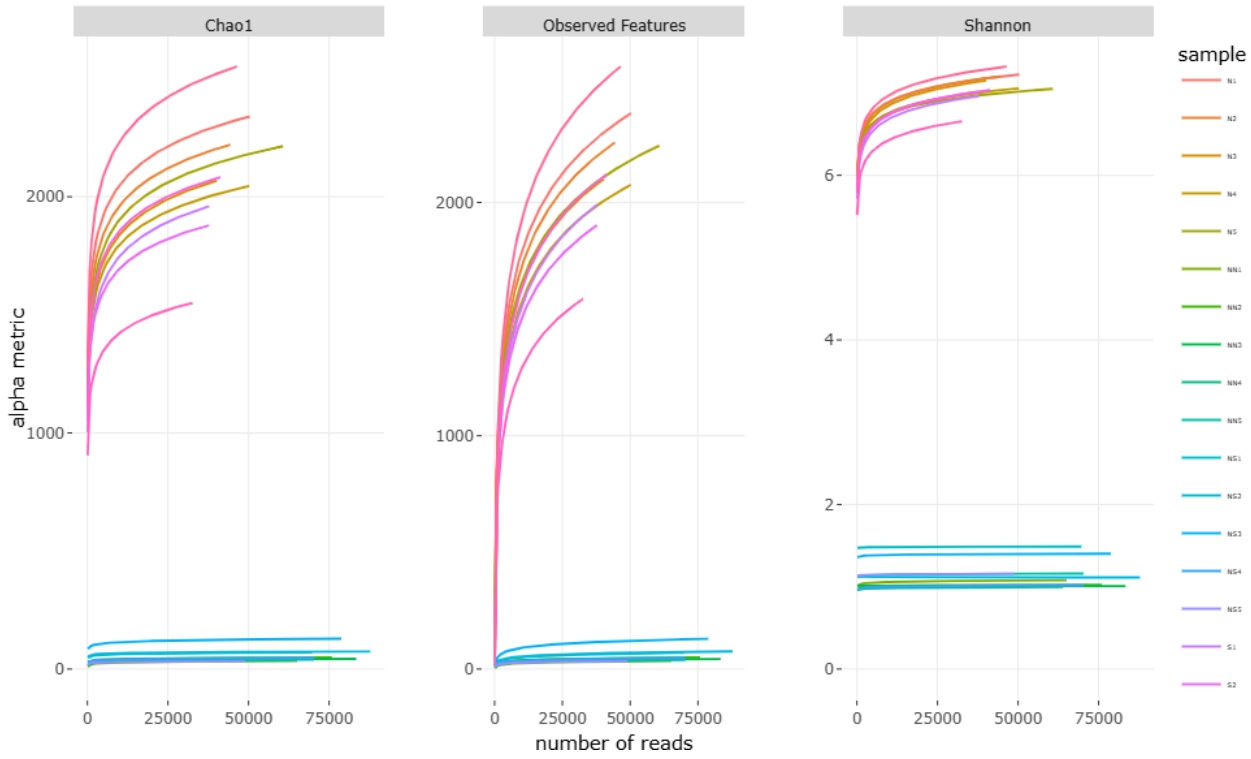


Table S2. Alpha diversity indices by sample group

Mean ± SD of Observed ASVs, Chao1, Shannon, and Gini–Simpson indices per group. Different compact letters (a, b) indicate statistically significant differences (Kruskal–Wallis followed by pairwise Wilcoxon rank-sum tests with Benjamini–Hochberg FDR correction, $P < 0.05$); groups sharing the same letter are not significantly different.

Group	Observed	Chao1	Shannon	Simpson
Nodule_NonArid	38.20 ± 5.72 (b)	39.20 ± 5.72 (b)	1.52 ± 0.10 (b)	0.58 ± 0.02 (b)
Nodule_SemiArid	67.20 ± 34.52 (b)	68.36 ± 34.58 (b)	1.78 ± 0.29 (b)	0.61 ± 0.05 (b)
Soil_NonArid	2075.60 ± 114.22 (a)	2080.19 ± 113.02 (a)	10.15 ± 0.11 (a)	1.00 ± 0.00 (a)
Soil_SemiArid	1911.00 ± 342.06 (a)	1919.19 ± 347.06 (a)	9.94 ± 0.33 (a)	1.00 ± 0.00 (a)

Table S3. PICRUSt2 prediction reliability (weighted NSTI)

Mean weighted Nearest Sequenced Taxon Index (NSTI) per sample (Table S3a) and per group (Table S3b), generated using PICRUSt2 (v2.5.2) via the Galaxy platform. Lower NSTI values indicate closer relatedness to reference genomes and higher confidence in predicted functional profiles; NSTI values below ~0.03 are generally considered highly reliable, whereas values above ~0.15–0.2 warrant more cautious interpretation.

Sample ID	Weighted NSTI
N1	0.2558
N2	0.2349
N3	0.2148
N4	0.2690
N5	0.2909
NN1	0.0269
NN2	0.0322
NN3	0.0288
NN4	0.0322
NN5	0.0290
NS1	0.0348
NS2	0.0285
NS3	0.0312
NS4	0.0330
NS5	0.0309
S1	0.2144
S2	0.2110
S3	0.2158
S4	0.2211
S5	0.2420

Group	n	Mean ± SD weighted NSTI
Non-arid soil	5	0.254 ± 0.030
Non-arid nodule	5	0.030 ± 0.002
Semi-arid soil	5	0.221 ± 0.012
Semi-arid nodule	5	0.032 ± 0.002
All soil samples	10	0.237 ± 0.028
All nodule samples	10	0.031 ± 0.002

Table S4. Pairwise comparisons of alpha diversity indices among sample groups

Pairwise Wilcoxon rank-sum tests among the four sample groups (Soil_NonArid, Soil_SemiArid, Nodule_NonArid, Nodule_SemiArid) for each alpha diversity index (Observed ASVs, Chao1, Shannon, Simpson), with Benjamini–Hochberg false discovery rate (FDR) correction for multiple comparisons ($p < 0.05$; ns = not significant).*

Pairwise Comparison	Measure	P-value (FDR-adjusted)	Significance
Soil_NonArid – Nodule_NonArid	Observed	0.0119	*
Soil_NonArid – Nodule_SemiArid	Observed	0.0119	*
Soil_NonArid – Soil_SemiArid	Observed	0.3095	ns
Nodule_NonArid – Nodule_SemiArid	Observed	0.2424	ns
Nodule_NonArid – Soil_SemiArid	Observed	0.0119	*
Nodule_SemiArid – Soil_SemiArid	Observed	0.0119	*
Soil_NonArid – Nodule_NonArid	Chao1	0.0119	*
Soil_NonArid – Nodule_SemiArid	Chao1	0.0119	*
Soil_NonArid – Soil_SemiArid	Chao1	0.3095	ns
Nodule_NonArid – Nodule_SemiArid	Chao1	0.2424	ns
Nodule_NonArid – Soil_SemiArid	Chao1	0.0119	*
Nodule_SemiArid – Soil_SemiArid	Chao1	0.0119	*
Soil_NonArid – Nodule_NonArid	Shannon	0.0119	*
Soil_NonArid – Nodule_SemiArid	Shannon	0.0119	*
Soil_NonArid – Soil_SemiArid	Shannon	0.1905	ns
Nodule_NonArid – Nodule_SemiArid	Shannon	0.1905	ns
Nodule_NonArid – Soil_SemiArid	Shannon	0.0119	*
Nodule_SemiArid – Soil_SemiArid	Shannon	0.0119	*
Soil_NonArid – Nodule_NonArid	Simpson	0.0119	*
Soil_NonArid – Nodule_SemiArid	Simpson	0.0119	*
Soil_NonArid – Soil_SemiArid	Simpson	0.1345	ns
Nodule_NonArid – Nodule_SemiArid	Simpson	0.2424	ns
Nodule_NonArid – Soil_SemiArid	Simpson	0.0119	*
Nodule_SemiArid – Soil_SemiArid	Simpson	0.0119	*

Table S5. Raw per-farm soil physicochemical values

Individual, independently measured physicochemical values for all 10 farm-level soil samples (N1–N5, S1–S5). All physicochemical and microbiome-based analyses reported in the main text use the full 10-farm dataset (n = 5 per zone; see Methods 2.3). These raw per-farm values underlie the zone-level mean ± SD summary statistics reported in Table 1

Sample ID	Zone	pH	EC (mmhos/cm)	MC (%)	N (%)	C (%)	K (mg/Kg)	P (mg/kg)	Mg (mg/Kg)	Ca (mg/Kg)	Fe (mg/Kg)	Zn (mg/Kg)
N1	Non-arid	6.37	0.156	5.000	0.222	2.200	83.902	24.715	31.200	499.200	1.240	1.345
N2	Non-arid	5.77	0.138	4.400	0.198	2.000	58.616	22.523	31.600	368.300	1.101	1.054
N3	Non-arid	5.55	0.143	4.500	0.140	1.400	33.221	21.612	32.300	486.900	0.998	0.879
N4	Non-arid	5.40	0.144	5.800	0.187	1.900	31.888	18.004	26.300	448.000	1.110	0.912
N5	Non-arid	5.59	0.059	5.500	0.163	1.600	82.568	17.903	25.400	358.900	1.112	0.978
S1	Semi-arid	5.54	0.077	2.400	0.105	1.100	14.659	8.284	21.200	313.200	1.063	0.765
S2	Semi-arid	5.52	0.101	2.500	0.082	0.800	11.991	5.076	21.500	365.600	0.878	0.864
S3	Semi-arid	5.75	0.057	3.100	0.047	0.500	21.218	13.494	20.500	375.700	0.754	0.723
S4	Semi-arid	5.88	0.037	3.200	0.117	1.200	28.999	17.770	22.900	298.900	0.912	0.894
S5	Semi-arid	5.86	0.041	2.000	0.140	1.400	27.886	15.798	23.400	311.000	0.699	0.812

Table S6. LEfSe statistical results for the top 50 taxonomic biomarkers (LDA score > 2.0), reported at every taxonomic rank identified (phylum through species) for each significant lineage. Underlies Figure 7. Significance: * P.adj < 0.05; ** P.adj < 0.01 (Benjamini–Hochberg FDR correction).

Taxon	Group	LDA score	P (unadjusted)	P (FDR-adjusted)	Sig.
g__Bradyrhizobium	Nodule_NonArid	5.66	0.0016	0.0072	**
f__Nitrobacteraceae	Nodule_NonArid	5.65	0.0016	0.0072	**
o__Hyphomicrobiales	Nodule_NonArid	5.65	0.0022	0.0091	**
s__Bradyrhizobium viridifuturi	Nodule_NonArid	5.64	0.0037	0.0135	*
c__Alphaproteobacteria	Nodule_NonArid	5.62	0.0026	0.0101	*
p__Pseudomonadota	Nodule_SemiArid	5.59	0.0035	0.0128	*
s__Bradyrhizobium yuanmingense	Nodule_SemiArid	5.59	0.0015	0.0070	**
k__Pseudomonadati	Nodule_SemiArid	5.42	0.0024	0.0095	**
k__Bacillati	Soil_SemiArid	5.42	0.0024	0.0095	**
p__Actinomycetota	Soil_SemiArid	5.26	0.0012	0.0068	**
c__Actinomycetes	Soil_NonArid	4.96	0.0012	0.0068	**
p__Chloroflexota	Soil_NonArid	4.83	0.0012	0.0068	**
p__Acidobacteriota	Soil_NonArid	4.77	0.0016	0.0072	**
c__Rubrobacteria	Soil_SemiArid	4.72	0.0008	0.0068	**
o__Rubrobacterales	Soil_SemiArid	4.72	0.0008	0.0068	**
f__Rubrobacteraceae	Soil_SemiArid	4.72	0.0008	0.0068	**
g__Rubrobacter	Soil_SemiArid	4.72	0.0008	0.0068	**

s__Rubrobacter spartanus	Soil_SemiArid	4.71	0.0026	0.0101	*
p__Bacillota	Soil_SemiArid	4.67	0.0019	0.0081	**
c__Bacilli	Soil_SemiArid	4.63	0.0019	0.0081	**
f__Rhodoligotrophaceae	Nodule_SemiArid	4.63	0.0019	0.0081	**
g__Rhodoligotrophos	Nodule_SemiArid	4.63	0.0019	0.0081	**
s__Rhodoligotrophos appendicifer	Nodule_SemiArid	4.63	0.0019	0.0081	**
o__Bacillales	Soil_SemiArid	4.62	0.0012	0.0068	**
c__Ktedonobacteria	Soil_NonArid	4.61	0.0006	0.0068	**
c__Terriglobia	Soil_NonArid	4.60	0.0008	0.0068	**
o__Ktedonobacterales	Soil_NonArid	4.58	0.0006	0.0068	**
c__Thermoleophilia	Soil_NonArid	4.58	0.0016	0.0074	**
f__Bacillaceae	Soil_SemiArid	4.55	0.0011	0.0068	**
f__Methylobacteriaceae	Soil_SemiArid	4.50	0.0009	0.0068	**
g__Microvirga	Soil_SemiArid	4.47	0.0009	0.0068	**
o__Gaiellales	Soil_NonArid	4.37	0.0010	0.0068	**
f__Gaiellaceae	Soil_NonArid	4.37	0.0010	0.0068	**
g__Gaiella	Soil_NonArid	4.37	0.0010	0.0068	**
s__Gaiella occulta	Soil_NonArid	4.37	0.0010	0.0068	**
p__Gemmatimonadota	Soil_NonArid	4.36	0.0011	0.0068	**
p__Verrucomicrobiota	Soil_NonArid	4.36	0.0012	0.0068	**
c__Gemmatimonadia	Soil_NonArid	4.36	0.0011	0.0068	**
o__Gemmatimonadales	Soil_NonArid	4.36	0.0011	0.0068	**
f__Gemmatimonadaceae	Soil_NonArid	4.36	0.0011	0.0068	**
o__Terriglobales	Soil_NonArid	4.33	0.0008	0.0068	**
f__Acidobacteriaceae	Soil_NonArid	4.33	0.0008	0.0068	**
c__Betaproteobacteria	Soil_NonArid	4.32	0.0014	0.0070	**
o__Streptosporangiales	Soil_NonArid	4.31	0.0006	0.0068	**
c__Chloroflexia	Soil_SemiArid	4.29	0.0012	0.0068	**
f__Hyphomicrobiaceae	Soil_NonArid	4.28	0.0006	0.0068	**
o__Bryobacterales	Soil_NonArid	4.27	0.0007	0.0068	**
f__Bryobacteraceae	Soil_NonArid	4.27	0.0007	0.0068	**
c__Vicinamibacteria	Soil_SemiArid	4.26	0.0010	0.0068	**
o__Vicinamibacterales	Soil_SemiArid	4.26	0.0010	0.0068	**

Table S7. CCA forward selection (ordistep) results

Step-wise forward selection of soil physicochemical variables for constraining the CCA model (vegan::ordistep, 999 permutations), primary model (n = 10). Table S7a shows the initial step (null model), testing each of the 11 variables individually; organic carbon (C) was the strongest and only significant predictor (F = 8.9505, P = 0.005). Table S7b shows the second step, after C was entered, testing whether any remaining variable further improved the model. No variable met the P < 0.05 threshold in Step 2 (moisture content was closest, P = 0.121), so selection stopped after organic carbon alone.

Table S7a. Step 1 — testing each variable against the null model (otu ~ 1).

Variable	AIC	F	P	Sig.
C (%)	23.006	8.9505	0.005	**
Fe (mg/Kg)	24.018	7.3181	0.007	**
N (%)	22.842	9.2294	0.008	**
MC (%)	22.572	9.7023	0.009	**
K (mg/Kg)	22.761	9.3696	0.009	**
Zn (mg/Kg)	25.116	5.7251	0.011	*
EC (mmhos/cm)	26.672	3.7474	0.064	.
Mg (mg/Kg)	27.032	3.3319	0.064	.
P (mg/kg)	27.843	2.4489	0.103	
Ca (mg/Kg)	27.949	2.3397	0.129	
pH	29.987	0.4330	0.674	ns

Table S7b. Step 2 — testing remaining variables after organic carbon entered the model (otu ~ C). None significant; selection stopped.

Variable	AIC	F	P	Sig.
MC (%)	21.964	2.4882	0.121	ns
K (mg/Kg)	22.308	2.1679	0.155	ns
EC (mmhos/cm)	23.356	1.2554	0.272	ns
Ca (mg/Kg)	23.626	1.0353	0.339	ns
pH	23.617	1.0427	0.341	ns
Fe (mg/Kg)	23.437	1.1884	0.361	ns
Mg (mg/Kg)	23.801	0.8958	0.436	ns
N (%)	24.079	0.6798	0.466	ns
Zn (mg/Kg)	24.497	0.3654	0.714	ns
P (mg/kg)	24.566	0.3146	0.744	ns

Final model significance test: cca(otu_final ~ C(%)), permutation test (999 permutations): F = 8.950, Df = 1, ChiSquare = 0.2197, P = 0.006, explaining 52.8% of total inertia. Variance inflation factor (VIF) for C = 1 (no collinearity, as expected with a single-variable model).

Table S8. Full-model Variance Inflation Factor (VIF) diagnostic

Variance Inflation Factor (VIF) diagnostic for the 11 soil physicochemical variables (n = 10, primary model), alongside each variable's role in the final forward-selected CCA model (Table S7). The VIF column assesses collinearity among predictor variables only and is independent of ecological importance; organic carbon (C) shows a low, unremarkable VIF (1.94) despite being the sole statistically significant driver of bacterial community composition, illustrating that VIF and explanatory importance are separate statistical properties. Electrical conductivity (VIF = 20.29), potassium (VIF = 11.45), and zinc (VIF = 18.35) exceeded the conventional collinearity threshold of 10; this diagnostic is reported for transparency only, as VIF was not used to pre-filter variables prior to forward selection (see Methods 2.7).

Variable	VIF	Collinearity diagnostic	Role in final CCA model
C (%)	1.94	Low — acceptable	SIGNIFICANT DRIVER (retained, P = 0.006)
N (%)	2.12	Low — acceptable	Not retained (redundant with C)
P (mg/kg)	2.93	Low — acceptable	Not retained (redundant with C)
EC (mmhos/cm)	20.29	High — exceeds threshold (>10)	Not retained (redundant with C)
MC (%)	4.91	Low — acceptable	Not retained (closest, P = 0.121)
K (mg/Kg)	11.45	High — exceeds threshold (>10)	Not retained (redundant with C)
Mg (mg/Kg)	6.65	Moderate — acceptable (<10)	Not retained (redundant with C)
Ca (mg/Kg)	9.06	Moderate — acceptable (<10)	Not retained (redundant with C)
Fe (mg/Kg)	5.61	Low — acceptable	Not retained (redundant with C)
Zn (mg/Kg)	18.35	High — exceeds threshold (>10)	Not retained (redundant with C)
pH	3.26	Low — acceptable	Not significant even alone (P = 0.674)

Table S9. Pairwise PERMANOVA comparisons of microbial community composition (Bray-Curtis dissimilarity, 999 permutations, FDR-adjusted P-values; n = 20).

Pairwise PERMANOVA comparisons of bacterial community composition among the four sample groups (Bray–Curtis dissimilarity, 999 permutations, Benjamini–Hochberg-adjusted p-values). Asterisk (*) denotes statistical significance at $p < 0.05$ after FDR correction. Summary statistics (all padj ≤ 0.026; F = 4.05–24.67; R^2 = 0.34–0.76) are reported in the main text.

Groups	F	R^2	P	P (FDR-adjusted)	Sig.
Soil_NonArid vs Nodule_NonArid	13.21	0.623	0.008	0.0132	*
Soil_NonArid vs Nodule_SemiArid	11.92	0.598	0.011	0.0132	*
Soil_NonArid vs Soil_SemiArid	4.05	0.336	0.026	0.026	*
Nodule_NonArid vs Nodule_SemiArid	24.67	0.755	0.009	0.0132	*
Nodule_NonArid vs Soil_SemiArid	20.81	0.722	0.008	0.0132	*
Nodule_SemiArid vs Soil_SemiArid	18.25	0.695	0.008	0.0132	*

Table S10. Spearman correlation coefficients between top-15 bacterial genera and soil physicochemical variables (primary model, n = 10)

Full pairwise Spearman correlation results underlying Figure 5 (165 genus–variable combinations), corrected for multiple testing using the Benjamini–Hochberg false discovery rate (FDR) procedure. "P (raw)" is the unadjusted two-sided P-value; "P (FDR)" is the Benjamini–Hochberg-adjusted P-value across all 165 tests. No correlation remained significant after FDR correction in this primary model (0/165 at P (FDR) ≤ 0.05, versus 56/165 at unadjusted P < 0.05).

Genus	Variable	rho	P (raw)	P (FDR)	Sig.
Rubrobacter	pH	0.273	0.4483	0.5209	ns
Gaiella	pH	0.152	0.6818	0.7401	ns
Microvirga	pH	-0.115	0.7588	0.7925	ns
Vicinamibacter	pH	0.6	0.0731	0.1609	ns
Paludibaculum	pH	-0.115	0.7588	0.7925	ns
Chthoniobacter	pH	0.333	0.3488	0.4295	ns
Solirubrobacter	pH	-0.406	0.2474	0.3373	ns
Rhodoplanes	pH	-0.006	1.0	1.0	ns
Roseiflexus	pH	0.309	0.3871	0.4595	ns
Dehalogenimonas	pH	-0.527	0.1228	0.2179	ns
Gemmatimonas	pH	0.067	0.8648	0.8862	ns
Capillimicrobium	pH	-0.709	0.0275	0.1261	ns
Niallia	pH	0.139	0.7072	0.7577	ns
Streptomyces	pH	0.03	0.9457	0.9515	ns
Bradyrhizobium	pH	0.224	0.5367	0.6065	ns
Rubrobacter	EC	-0.733	0.0212	0.1204	ns
Gaiella	EC	0.527	0.1228	0.2179	ns
Microvirga	EC	-0.345	0.3305	0.4163	ns
Vicinamibacter	EC	-0.636	0.0544	0.1439	ns
Paludibaculum	EC	0.406	0.2474	0.3373	ns
Chthoniobacter	EC	0.321	0.3677	0.4461	ns
Solirubrobacter	EC	-0.127	0.7329	0.7802	ns
Rhodoplanes	EC	0.612	0.0665	0.1545	ns
Roseiflexus	EC	-0.406	0.2474	0.3373	ns
Dehalogenimonas	EC	0.612	0.0665	0.1545	ns

Gemmatimonas	EC	0.491	0.1544	0.2523	ns
Capillimicrobium	EC	0.467	0.1782	0.2723	ns
Niallia	EC	-0.636	0.0544	0.1439	ns
Streptomyces	EC	-0.576	0.0878	0.1833	ns
Bradyrhizobium	EC	0.503	0.1434	0.2414	ns
Rubrobacter	MC	-0.782	0.0117	0.1056	ns
Gaiella	MC	0.382	0.279	0.3742	ns
Microvirga	MC	-0.6	0.0731	0.1609	ns
Vicinamibacter	MC	-0.709	0.0275	0.1261	ns
Paludibaculum	MC	0.455	0.1909	0.2838	ns
Chthoniobacter	MC	0.491	0.1544	0.2523	ns
Solirubrobacter	MC	-0.612	0.0665	0.1545	ns
Rhodoplanes	MC	0.467	0.1782	0.2723	ns
Roseiflexus	MC	-0.818	0.0068	0.1056	ns
Dehalogenimonas	MC	0.782	0.0117	0.1056	ns
Gemmatimonas	MC	0.709	0.0275	0.1261	ns
Capillimicrobium	MC	0.321	0.3677	0.4461	ns
Niallia	MC	-0.758	0.0159	0.1056	ns
Streptomyces	MC	-0.782	0.0117	0.1056	ns
Bradyrhizobium	MC	0.624	0.0602	0.1506	ns
Rubrobacter	N	-0.772	0.0089	0.1056	ns
Gaiella	N	0.657	0.0392	0.1439	ns
Microvirga	N	-0.76	0.0108	0.1056	ns
Vicinamibacter	N	-0.407	0.2427	0.3373	ns
Paludibaculum	N	0.547	0.1017	0.2025	ns
Chthoniobacter	N	0.809	0.0046	0.1056	ns
Solirubrobacter	N	-0.669	0.0345	0.1379	ns
Rhodoplanes	N	0.778	0.008	0.1056	ns
Roseiflexus	N	-0.316	0.3736	0.4466	ns
Dehalogenimonas	N	0.413	0.2351	0.3373	ns
Gemmatimonas	N	0.62	0.0558	0.1439	ns
Capillimicrobium	N	0.188	0.6021	0.6713	ns
Niallia	N	-0.815	0.0041	0.1056	ns
Streptomyces	N	-0.638	0.047	0.1439	ns
Bradyrhizobium	N	0.729	0.0166	0.1056	ns
Rubrobacter	C	-0.772	0.0089	0.1056	ns
Gaiella	C	0.657	0.0392	0.1439	ns
Microvirga	C	-0.76	0.0108	0.1056	ns
Vicinamibacter	C	-0.407	0.2427	0.3373	ns
Paludibaculum	C	0.547	0.1017	0.2025	ns
Chthoniobacter	C	0.809	0.0046	0.1056	ns
Solirubrobacter	C	-0.669	0.0345	0.1379	ns
Rhodoplanes	C	0.778	0.008	0.1056	ns
Roseiflexus	C	-0.316	0.3736	0.4466	ns
Dehalogenimonas	C	0.413	0.2351	0.3373	ns
Gemmatimonas	C	0.62	0.0558	0.1439	ns
Capillimicrobium	C	0.188	0.6021	0.6713	ns
Niallia	C	-0.815	0.0041	0.1056	ns
Streptomyces	C	-0.638	0.047	0.1439	ns
Bradyrhizobium	C	0.729	0.0166	0.1056	ns
Rubrobacter	K	-0.661	0.044	0.1439	ns
Gaiella	K	0.394	0.2629	0.3555	ns
Microvirga	K	-0.661	0.044	0.1439	ns
Vicinamibacter	K	-0.358	0.3128	0.4001	ns
Paludibaculum	K	0.418	0.2324	0.3373	ns
Chthoniobacter	K	0.624	0.0602	0.1506	ns
Solirubrobacter	K	-0.721	0.0242	0.1261	ns
Rhodoplanes	K	0.576	0.0878	0.1833	ns
Roseiflexus	K	-0.479	0.1661	0.2635	ns
Dehalogenimonas	K	0.455	0.1909	0.2838	ns
Gemmatimonas	K	0.552	0.1043	0.2025	ns
Capillimicrobium	K	0.042	0.9186	0.9299	ns
Niallia	K	-0.733	0.0212	0.1204	ns
Streptomyces	K	-0.721	0.0242	0.1261	ns
Bradyrhizobium	K	0.758	0.0159	0.1056	ns
Rubrobacter	P	-0.612	0.0665	0.1545	ns

Gaiella	P	0.661	0.044	0.1439	ns
Microvirga	P	-0.503	0.1434	0.2414	ns
Vicinamibacter	P	-0.297	0.407	0.4762	ns
Paludibaculum	P	0.236	0.5139	0.5888	ns
Chthoniobacter	P	0.6	0.0731	0.1609	ns
Solirubrobacter	P	-0.552	0.1043	0.2025	ns
Rhodoplanes	P	0.527	0.1228	0.2179	ns
Roseiflexus	P	-0.345	0.3305	0.4163	ns
Dehalogenimonas	P	0.358	0.3128	0.4001	ns
Gemmatimonas	P	0.539	0.1133	0.2149	ns
Capillimicrobium	P	0.139	0.7072	0.7577	ns
Niallia	P	-0.685	0.0351	0.1379	ns
Streptomyces	P	-0.6	0.0731	0.1609	ns
Bradyrhizobium	P	0.661	0.044	0.1439	ns
Rubrobacter	Mg	-0.673	0.0394	0.1439	ns
Gaiella	Mg	0.612	0.0665	0.1545	ns
Microvirga	Mg	-0.37	0.2956	0.3933	ns
Vicinamibacter	Mg	-0.479	0.1661	0.2635	ns
Paludibaculum	Mg	0.115	0.7588	0.7925	ns
Chthoniobacter	Mg	0.442	0.2042	0.3008	ns
Solirubrobacter	Mg	-0.261	0.4697	0.5419	ns
Rhodoplanes	Mg	0.527	0.1228	0.2179	ns
Roseiflexus	Mg	-0.358	0.3128	0.4001	ns
Dehalogenimonas	Mg	0.503	0.1434	0.2414	ns
Gemmatimonas	Mg	0.333	0.3488	0.4295	ns
Capillimicrobium	Mg	0.515	0.1328	0.2331	ns
Niallia	Mg	-0.661	0.044	0.1439	ns
Streptomyces	Mg	-0.479	0.1661	0.2635	ns
Bradyrhizobium	Mg	0.491	0.1544	0.2523	ns
Rubrobacter	Ca	-0.564	0.0958	0.1976	ns
Gaiella	Ca	0.333	0.3488	0.4295	ns
Microvirga	Ca	-0.055	0.8916	0.9082	ns
Vicinamibacter	Ca	-0.406	0.2474	0.3373	ns
Paludibaculum	Ca	0.164	0.6567	0.7176	ns
Chthoniobacter	Ca	0.091	0.8114	0.8368	ns
Solirubrobacter	Ca	-0.103	0.785	0.8146	ns
Rhodoplanes	Ca	0.297	0.407	0.4762	ns
Roseiflexus	Ca	-0.358	0.3128	0.4001	ns
Dehalogenimonas	Ca	0.455	0.1909	0.2838	ns
Gemmatimonas	Ca	0.503	0.1434	0.2414	ns
Capillimicrobium	Ca	0.224	0.5367	0.6065	ns
Niallia	Ca	-0.527	0.1228	0.2179	ns
Streptomyces	Ca	-0.648	0.049	0.1439	ns
Bradyrhizobium	Ca	0.576	0.0878	0.1833	ns
Rubrobacter	Fe	-0.758	0.0159	0.1056	ns
Gaiella	Fe	0.358	0.3128	0.4001	ns
Microvirga	Fe	-0.867	0.0027	0.1056	ns
Vicinamibacter	Fe	-0.648	0.049	0.1439	ns
Paludibaculum	Fe	0.77	0.0137	0.1056	ns
Chthoniobacter	Fe	0.636	0.0544	0.1439	ns
Solirubrobacter	Fe	-0.685	0.0351	0.1379	ns
Rhodoplanes	Fe	0.758	0.0159	0.1056	ns
Roseiflexus	Fe	-0.648	0.049	0.1439	ns
Dehalogenimonas	Fe	0.685	0.0351	0.1379	ns
Gemmatimonas	Fe	0.588	0.0802	0.1742	ns
Capillimicrobium	Fe	0.176	0.632	0.6952	ns
Niallia	Fe	-0.758	0.0159	0.1056	ns
Streptomyces	Fe	-0.709	0.0275	0.1261	ns
Bradyrhizobium	Fe	0.636	0.0544	0.1439	ns
Rubrobacter	Zn	-0.709	0.0275	0.1261	ns
Gaiella	Zn	0.552	0.1043	0.2025	ns
Microvirga	Zn	-0.77	0.0137	0.1056	ns
Vicinamibacter	Zn	-0.467	0.1782	0.2723	ns
Paludibaculum	Zn	0.527	0.1228	0.2179	ns
Chthoniobacter	Zn	0.842	0.0045	0.1056	ns
Solirubrobacter	Zn	-0.733	0.0212	0.1204	ns

Rhodoplanes	Zn	0.782	0.0117	0.1056	ns
Roseiflexus	Zn	-0.467	0.1782	0.2723	ns
Dehalogenimonas	Zn	0.539	0.1133	0.2149	ns
Gemmatimonas	Zn	0.685	0.0351	0.1379	ns
Capillimicrobium	Zn	0.176	0.632	0.6952	ns
Niallia	Zn	-0.636	0.0544	0.1439	ns
Streptomyces	Zn	-0.636	0.0544	0.1439	ns
Bradyrhizobium	Zn	0.648	0.049	0.1439	ns

Table S11. Statistical comparison of predicted PGP/stress-tolerance gene family abundances (Wilcoxon rank-sum tests, Benjamini–Hochberg FDR-corrected, n = 20)

Underlies Figure 8. "Group A" / "Group B" correspond to Non-Arid/Semi-Arid for the within-compartment comparisons and Soil/Nodule for the compartment comparison. FDR correction (Benjamini–Hochberg) applied jointly across all 24 tests. Trehalose (Drought) was excluded as no ASV carried a predicted copy of this gene family (K00710).

Comparison	PGP Function	P (raw)	P (FDR)	Sig.	Mean (Group A)	Mean (Group B)
Soil: Non-Arid vs Semi-Arid	ACC Deaminase (Stress)	0.0159	0.0293	*	4375.1	2496.4
Soil: Non-Arid vs Semi-Arid	Glycine Betaine (Salinity)	0.2222	0.2667	ns	88488.2	78990.2
Soil: Non-Arid vs Semi-Arid	HCN Production	0.0079	0.0173	*	15238.0	11112.5
Soil: Non-Arid vs Semi-Arid	IAA Production	0.0159	0.0293	*	70195.0	53063.0
Soil: Non-Arid vs Semi-Arid	N-Fixation	0.2222	0.2667	ns	2992.6	2554.4
Soil: Non-Arid vs Semi-Arid	P-Solubilization	0.3095	0.3537	ns	8686.2	7871.5
Soil: Non-Arid vs Semi-Arid	Proline (Drought/Salt)	0.0556	0.0889	ns	43555.2	35731.3
Soil: Non-Arid vs Semi-Arid	Siderophores	0.0317	0.0544	ns	166527.8	124576.3
Nodule: Non-Arid vs Semi-Arid	ACC Deaminase (Stress)	0.0593	0.0890	ns	2.3	15.4
Nodule: Non-Arid vs Semi-Arid	Glycine Betaine (Salinity)	0.2222	0.2667	ns	112833.3	142750.6
Nodule: Non-Arid vs Semi-Arid	HCN Production	1.0	1.0	ns	39302.7	38393.9
Nodule: Non-Arid vs Semi-Arid	IAA Production	0.0079	0.0173	*	105977.5	71356.1
Nodule: Non-Arid vs Semi-Arid	N-Fixation	0.0079	0.0173	*	5275.6	101051.1
Nodule: Non-Arid vs Semi-Arid	P-Solubilization	0.6905	0.7532	ns	37595.1	35197.1
Nodule: Non-Arid vs Semi-Arid	Proline (Drought/Salt)	0.8413	0.8778	ns	70073.9	69610.8
Nodule: Non-Arid vs Semi-Arid	Siderophores	0.0079	0.0173	*	225961.8	451254.5
Soil vs Nodule (all zones combined)	ACC Deaminase (Stress)	0.0002	0.0007	*	3435.8	8.8
Soil vs Nodule	Glycine Betaine	0.0000	0.0002	*	83739.2	127792.0

(all zones combined)	(Salinity)					
Soil vs Nodule (all zones combined)	HCN Production	0.0000	0.0001	*	13175.3	38848.3
Soil vs Nodule (all zones combined)	IAA Production	0.0039	0.0133	*	61629.0	88666.8
Soil vs Nodule (all zones combined)	N-Fixation	0.1230	0.1737	ns	2773.5	53163.4
Soil vs Nodule (all zones combined)	P-Solubilization	0.0000	0.0001	*	8278.8	36396.1
Soil vs Nodule (all zones combined)	Proline (Drought/Salt)	0.0000	0.0001	*	39643.2	69842.4
Soil vs Nodule (all zones combined)	Siderophores	0.0000	0.0001	*	145552.1	338608.2

Table S12. Putative species-level taxonomic assignments of ASVs classified within the genus Bradyrhizobium, with mean relative abundance by nodule zone

Note: values below are corrected mean relative abundances (%), calculated by averaging across n=5 biological replicates per nodule zone, verified directly against the raw ASV relative-abundance table. (Previous version of this table incorrectly reported the sum across replicates instead of the mean.) Species-level assignments are exploratory given the resolution limits of the V3–V4 region and should be interpreted as hypothesis-generating rather than confirmed identifications (see Discussion, Section 4).

ASV ID	Putative Species (BLAST)	Mean rel. abund., Non-Arid nodules (%)	Mean rel. abund., Semi-Arid nodules (%)
ASV_2	*Bradyrhizobium viridifuturi*	49.79	1.91
ASV_3	*Bradyrhizobium yuanmingense*	0.00	44.18
ASV_5	*Bradyrhizobium betae*	2.49	2.74
ASV_6	*Bradyrhizobium vignae*	0.00	1.63
ASV_147	*Bradyrhizobium viridifuturi*	0.11	0.00
ASV_145	*Bradyrhizobium yuanmingense*	0.00	0.11
ASV_678	*Bradyrhizobium vignae*	0.00	0.03
ASV_978	*Bradyrhizobium viridifuturi*	0.02	0.00
ASV_1144	*Bradyrhizobium betae*	0.00	0.02
ASV_2197	*Bradyrhizobium yuanmingense*	0.00	0.01
ASV_3018	*Bradyrhizobium yuanmingense*	0.00	0.01
ASV_3768	*Bradyrhizobium yuanmingense*	0.00	0.01

Table S13. ASV species-level taxa contributing to predicted plant growth-promoting (PGP) and stress-tolerance functions underlying Figure 8

Note: values below are corrected mean relative abundances (%): non-arid and semi-arid nodules averaged across n=5 replicates each; soil averaged across n=10 pooled samples. Verified against the raw ASV table. Species-level assignments are exploratory given the resolution limits of the V3–V4 region and should be interpreted as hypothesis-generating rather than confirmed strain-specific functional attributions (see Discussion, Section 4).

Species	Predicted PGP/Stress-Tolerance Function(s)	Mean rel. abund., Non-Arid nodules (%)	Mean rel. abund., Semi-Arid nodules (%)	Mean rel. abund., Soil (%)
Acidibacter ferrireducens	Siderophores	0.00	0.00	0.30
Actinoallomurus vinaceus	Siderophores	0.00	0.00	0.22
Actinomadura rifamycini	Siderophores, Glycine Betaine (Salinity), IAA Production	0.00	0.00	0.76

Arenimicrobium luteum	Siderophores	0.00	0.00	0.80
Arthrobacter pokkalii	IAA Production, Siderophores	0.00	0.00	0.41
Blastochloris viridis	Siderophores	0.00	0.00	0.57
Bradyrhizobium betae	Siderophores, P-Solubilization, Proline (Drought/Salt), HCN Production, IAA Production, Glycine Betaine (Salinity), N-Fixation	2.49	2.77	0.16
Bradyrhizobium vignae	HCN Production, IAA Production, Proline (Drought/Salt), Siderophores, Glycine Betaine (Salinity)	0.00	1.67	0.08
Bradyrhizobium viridifuturi	P-Solubilization, HCN Production, Proline (Drought/Salt), IAA Production, Siderophores, Glycine Betaine (Salinity)	49.92	1.91	0.71
Bradyrhizobium yuanmingense	IAA Production, P-Solubilization, HCN Production, Proline (Drought/Salt), Siderophores, Glycine Betaine (Salinity), N-Fixation	0.00	44.33	0.08
Brevitalea aridisoli	Siderophores	0.00	0.00	0.69
Capillimicrobium parvum	IAA Production, Glycine Betaine (Salinity), Siderophores	0.00	0.00	1.46
Chthoniobacter flavus	Siderophores, Proline (Drought/Salt), IAA Production, P-Solubilization	0.00	0.00	1.96
Gaiella occulta	IAA Production, Glycine Betaine (Salinity), Siderophores, Proline (Drought/Salt)	0.00	0.00	4.09
Gemmatimonas aurantiaca	Siderophores	0.00	0.00	0.74
Gemmatimonas phototrophica	Siderophores	0.00	0.00	0.81
Gemmatirosa kalamazoonensis	Siderophores	0.00	0.00	0.45
Luteitalea pratensis	Siderophores, IAA Production	0.00	0.00	0.58
Microvirga arsenatis	Siderophores, Glycine Betaine (Salinity)	0.00	0.00	0.40
Microvirga lenta	Siderophores	0.00	0.00	0.23
Microvirga massiliensis	Glycine Betaine (Salinity), Siderophores	0.00	0.00	0.46
Microvirga tunisiensis	Siderophores, Glycine Betaine (Salinity)	0.00	0.00	0.71
Mycobacterium aquaticum	Glycine Betaine (Salinity)	0.00	0.00	0.29
Mycolicibacterium pallens	Glycine Betaine (Salinity)	0.00	0.00	0.28
Neobacillus vireti	Siderophores	0.00	0.00	0.52
Niallia endozanthoxylica	Siderophores	0.00	0.00	0.53
Niallia nealsonii	Siderophores	0.00	0.00	0.71
Nitrosospira multiformis	Siderophores	0.00	0.00	0.38
Paeniglutamicibacter cryotolerans	IAA Production	0.00	0.00	0.29
Paludibaculum fermentans	Siderophores, Glycine Betaine (Salinity), IAA Production, Proline (Drought/Salt)	0.00	0.00	2.18
Petropleomorpha daqingensis	Siderophores	0.00	0.00	0.34
Pseudacidobacterium ailaaui	Siderophores, IAA Production	0.00	0.00	1.03
Rhodoplanes tepidicaeni	Siderophores, Glycine Betaine (Salinity), Proline (Drought/Salt), HCN Production, IAA Production	0.00	0.00	1.62
Roseicella frigidaeris	Siderophores	0.00	0.00	0.29
Roseiflexus castenholzii	Glycine Betaine (Salinity), Siderophores	0.00	0.00	1.68
Roseisolibacter agri	Siderophores	0.00	0.00	0.66
Rubrobacter spartanus	Glycine Betaine (Salinity),	0.00	0.00	5.51

	Siderophores, IAA Production, Proline (Drought/Salt)			
Skermanella mucosa	Glycine Betaine (Salinity), Siderophores	0.00	0.00	0.82
Solirubrobacter ginsenosidimutans	IAA Production, Siderophores, Glycine Betaine (Salinity)	0.00	0.00	1.31
Sphaerobacter thermophilus	Glycine Betaine (Salinity)	0.00	0.00	1.18
Sphingomonas limnosediminicola	Siderophores	0.00	0.01	0.63
Stella humosa	Siderophores, Glycine Betaine (Salinity)	0.00	0.00	0.64
Undibacter mobilis	Siderophores	0.00	0.00	0.24
Usitatibacter rugosus	Siderophores	0.00	0.00	0.29
Vicinamibacter silvestris	Proline (Drought/Salt), Siderophores, IAA Production, Glycine Betaine (Salinity)	0.00	0.00	2.40

Table S14. Putative species-level taxonomic assignments of ASVs classified within the stress-adapted soil genera Rubrobacter, Actinomadura, and Rhodoplanes, with mean relative abundance by soil aridity zone

Note: values below are corrected mean relative abundances (%), averaged across n=5 biological replicates per soil zone, verified against the raw ASV table. Species-level assignments are exploratory given the resolution limits of the V3–V4 region and should be interpreted as hypothesis-generating rather than confirmed identifications (see Discussion, Section 4).

ASV ID	**Putative Species (BLAST)**	**Genus**	**Mean rel. abund., Non-Arid soil (%)**	**Mean rel. abund., Semi-Arid soil (%)**
ASV_7	*Rubrobacter spartanus*	Rubrobacter	0.204	2.206
ASV_10	*Rubrobacter spartanus*	Rubrobacter	0.069	1.185
ASV_14	*Rubrobacter spartanus*	Rubrobacter	0.102	0.848
ASV_16	*Rubrobacter spartanus*	Rubrobacter	0.055	0.807
ASV_15	*Rhodoplanes tepidicaeni*	Rhodoplanes	0.664	0.030
ASV_24	*Rubrobacter spartanus*	Rubrobacter	0.079	0.573
ASV_30	*Rubrobacter spartanus*	Rubrobacter	0.092	0.492
ASV_33	*Rubrobacter spartanus*	Rubrobacter	0.046	0.501
ASV_37	*Rubrobacter spartanus*	Rubrobacter	0.051	0.462
ASV_25	*Rhodoplanes tepidicaeni*	Rhodoplanes	0.463	0.046
ASV_28	*Actinomadura rifamycini DSM 43936*	Actinomadura	0.445	0.000
ASV_43	*Rhodoplanes tepidicaeni*	Rhodoplanes	0.239	0.132
ASV_48	*Rhodoplanes tepidicaeni*	Rhodoplanes	0.315	0.000
ASV_72	*Rubrobacter spartanus*	Rubrobacter	0.029	0.278
ASV_84	*Rubrobacter spartanus*	Rubrobacter	0.014	0.283
ASV_70	*Rhodoplanes tepidicaeni*	Rhodoplanes	0.139	0.147
ASV_97	*Rubrobacter spartanus*	Rubrobacter	0.000	0.280
ASV_68	*Actinomadura rifamycini DSM 43936*	Actinomadura	0.252	0.000
ASV_119	*Rubrobacter spartanus*	Rubrobacter	0.024	0.211
ASV_144	*Rubrobacter tropicus*	Rubrobacter	0.062	0.157

Table S15. Complete PERMDISP (betadisper) results for multivariate dispersion homogeneity

Homogeneity of multivariate dispersion (Bray–Curtis distances) among the four sample groups, tested using PERMDISP (betadisper function, vegan package), with pairwise comparisons based on 999 permutations. These results underlie the PERMDISP statistics cited in the Discussion and confirm that the aridity-associated

compositional shift detected within the soil compartment (Soil_NonArid vs Soil_SemiArid) reflects genuine compositional differences rather than a dispersion artifact, since dispersion did not differ significantly for this comparison (P = 0.379).

Table S15a. Overall PERMDISP test

Test	Df	F	P (999 permutations)	Sig.
Overall (4 groups: Soil_NonArid, Soil_SemiArid, Nodule_NonArid, Nodule_SemiArid)	3, 16	23.63	0.001	**

Table S15b. Pairwise PERMDISP comparisons

Pairwise Comparison	P (permuted)	Sig.
Nodule_NonArid – Nodule_SemiArid	0.009	*
Nodule_NonArid – Soil_NonArid	0.002	*
Nodule_NonArid – Soil_SemiArid	0.001	*
Nodule_SemiArid – Soil_NonArid	0.004	*
Nodule_SemiArid – Soil_SemiArid	0.001	*
Soil_NonArid – Soil_SemiArid	0.379	ns

Table S16. Genus-level relative abundance (%) of the top 15 bacterial genera by sample group, with one-way ANOVA and Tukey HSD post hoc comparisons

Underlies Figure 6. Values are mean relative abundance (%) per group (n = 5 per group). Different lowercase letters within the same row indicate a statistically significant difference among groups (one-way ANOVA followed by Tukey's HSD post hoc test, P < 0.05); values sharing a letter are not significantly different. Genera aggregated at the genus level from ASVs after exclusion of Marileptolyngbya (chloroplast contamination).

Genus	Nodule_NonArid (%)	Nodule_SemiArid (%)	Soil_NonArid (%)	Soil_SemiArid (%)
Bradyrhizobium	90.91 a	89.67 a	2.02 b	0.54 b
Rhodoligotrophos	7.99 a	8.5 a	0.07 b	0.18 b
Rubrobacter	0 b	0 b	1.05 b	10.29 a
Gaiella	0 b	0 b	4.76 a	3.42 a
Microvirga	0 b	0.01 b	1.5 b	5.87 a
Vicinamibacter	0 c	0 c	1.93 b	2.88 a
Paludibaculum	0 b	0 b	2.91 a	1.45 ab
Chthoniobacter	0 b	0 b	3.05 a	0.88 b
Solirubrobacter	0 b	0 b	1.53 ab	2.17 a
Rhodoplanes	0 b	0 b	2.93 a	0.75 b
Roseiflexus	0 b	0 b	1.24 a	2.13 a
Dehalogenimonas	0 c	0 c	1.98 a	1.33 b
Gemmatimonas	0 b	0 b	2.26 a	0.85 b
Capillimicrobium	0 b	0 b	1.66 a	1.25 a
Streptomyces	0 c	0.08 c	0.84 b	1.9 a