Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Soil pH associations with rhizosphere microbial communities and functional potential across Allium victorialis habitats

Jingxin Wang
Chinese Academy of Agricultural Sciences
Hui Zhang
Heilongjiang Academy of Agricultural Sciences
Jiangping Song
Chinese Academy of Agricultural Sciences
Xingliang Wang
Heilongjiang Academy of Agricultural Sciences
Huixia Jia
Chinese Academy of Agricultural Sciences
Xiaohui Zhang
Chinese Academy of Agricultural Sciences
Ruifeng Wang
Chinese Academy of Agricultural Sciences
Yue Huang
Heilongjiang Academy of Agricultural Sciences
Wenlong Yang
Chinese Academy of Agricultural Sciences
Haiping Wang
wanghaiping@caas.cn

Chinese Academy of Agricultural Sciences

Research Article

Keywords: Soil pH, Rhizosphere microbiome, Microbial ecology, Metagenomics, Functional potential, Allium victorialis

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11593193/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Soil pH associations with rhizosphere microbial communities and functional potential across *Allium victorialis* habitats

Jingxin Wang[1#], Hui Zhang[2#], Jiangping Song[1#], Xingliang Wang[2], Huixia Jia[1], Xiaohui Zhang[1], Ruifeng Wang[1], Yue Huang[2], Wenlong Yang[1], Haiping Wang[1*]

1 State Key Laboratory of Vegetable Biobreeding, Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing 100081, China

2 Horticultural Branch, Heilongjiang Academy of Agricultural Sciences, Harbin 150040, China

# These authors contributed equally to this work.

*Correspondence: Haiping Wang (wanghaiping@caas.cn).

Abstract

Background: *Allium victorialis* is a wild vegetable and medicinal plant in northeastern China. The relationships between environmental variation, its rhizosphere microbiome and plant phenotypic traits remain poorly understood. We used shotgun metagenomic sequencing of 125 composite rhizosphere soil samples from 25 sites in Heilongjiang Province to characterize microbial communities and their associations with soil pH, geography and plant traits.

Results: Annotation identified 10,906 species-level taxa spanning bacteria, archaea, fungi and viruses. Community dissimilarities were associated with sampling site in all four groups, although dispersion also differed among sites for bacteria, archaea and fungi. Site-level Mantel tests supported an association between pH differences and archaeal community dissimilarity ($r = 0.547$, $q = 0.0160$ across 32 tests). Two co-abundance modules were negatively associated with site-mean pH, with support from random-intercept models separating between-site and within-site variation. Observed taxon richness decreased with site-mean pH in a random-intercept model (slope = −693.86 taxa per pH unit, $q = 0.00524$), although sequencing depth was not included in the model. Shannon and Simpson diversity did not show significant between-site associations after adjustment. Screening 2,127 module members against 89 nutrient-

transformation KOs identified 140 candidate taxa. Functional annotations indicated nutrient transformation potential, including several DMSP-related markers.

Conclusions: Rhizosphere community composition and co-abundance patterns varied across *A. victorialis* habitats and were associated with soil pH at the between-site level. The results support group-specific environmental associations and functional potential, while richness estimates remain sensitive to sampling depth. The observational design does not establish independent pH effects, pathway activity or microbial contributions to plant adaptation.

Keywords: Soil pH; Rhizosphere microbiome; Microbial ecology; Metagenomics; Functional potential; *Allium victorialis*

Background

Allium victorialis L. is a perennial wild vegetable and medicinal plant. Extracts and isolated compounds have shown antioxidant and anti-neuroinflammatory activities in experimental assays [1, 2]. Habitat factors have been examined in relation to the nutrient composition of wild *A. victorialis* in the Changbai Mountains [3]. More broadly, habitat variation is relevant to conservation planning for medicinal plants [4].

Plant reintroduction studies report variable establishment success [5], and aboveground–belowground linkages can inform ecological restoration [6]. The rhizosphere microbiome contributes to plant nutrition, stress responses and pathogen suppression [7]. Experiments with other rare or endangered plants have linked soil microbiomes to plant performance and population persistence [8]. In Wollemi pine (*Wollemia nobilis*), soil microbial community variation was associated with seedling growth independently of species-specific plant–soil feedback [9]. These findings motivate study of belowground communities without establishing the same mechanisms in *A. victorialis*.

Soil pH is an important correlate of rhizosphere bacterial community variation. A controlled study in peat-based growing substrates found that substrate pH and plant species influenced bacterial communities [10]; work in acidic crop soil also documented pH-associated changes in rhizosphere bacterial diversity and composition [11]. These studies provide context rather than evidence for a particular soil preference of *A. victorialis*. Cultivation studies in *Fritillaria cirrhosa* linked changes in rhizosphere communities with

changes in metabolites [12]. Research on diseased *Panax notoginseng* rhizospheres associated declining beneficial bacterial groups with reduced autotoxin-degradation functions [13]. Whether comparable relationships occur in wild *A. victorialis* requires direct evaluation.

A germination study reported failure of *A. victorialis* seeds to germinate under its tested conditions despite viable embryos [14]. *Herbaspirillum rhizosphaerae* was isolated from the rhizosphere of *A. victorialis* var. *platyphyllum* [15], demonstrating a cultivable rhizosphere association without establishing host specificity. Here, we used shotgun metagenomics across 25 sites in Heilongjiang Province to characterize bacterial, archaeal, fungal and viral annotations, examine associations with pH, geography and plant traits, and describe nutrient-transformation potential. We distinguished between-site from within-site associations where supported by the design and treated taxon–KO correlations as candidate relationships rather than evidence of gene carriage or activity.

Methods

Site selection, plant materials, and rhizosphere soil sampling

In June 2024, rhizosphere soil was collected from wild populations of *Allium victorialis* at 25 sites in Heilongjiang Province, China (Additional file 1). Xingliang Wang (Horticultural Branch, Heilongjiang Academy of Agricultural Sciences, Harbin, China) conducted the field sampling and identified the plants in the field. For metagenomic sequencing, rhizosphere soil from 25 plants at each site was pooled in groups of five, yielding five composite samples per site and 125 composite samples from 625 plants in total. Plant material was also collected during the same fieldwork for subsequent research.

All sampling sites were outside designated protected areas. Under the rules applicable to these sites in June 2024, no special collection permit was required to collect rhizosphere soil or plant material. No voucher specimen from the original sampling was deposited in a herbarium or other reference collection, and no voucher accession number is available.

Plant phenotypic traits and soil physicochemical properties

Leaf length (LL), leaf width (LW), petiole length (PL) and plant height (PH), measured in cm, were averaged across the five plants corresponding to each composite. Soil pH was measured in composite soil.

Leaf color (LC) was an overall plant-group greenness rating: 1, light green; 2, green; 3, dark green (Additional file 2).

DNA extraction, library preparation, and metagenomic sequencing

DNA was extracted using the VAMNE Stool/Soil DNA Extraction Kit-BOX 2 (DM401-C2; Vazyme). Libraries were prepared from 200 ng DNA using the Hieff NGS OnePot Pro DNA Library Prep Kit V2 (12195ES96; Yeasen Biotechnology), with enzymatic fragmentation to 250–300 bp, end repair, A-tailing, adapter ligation, purification, size selection and PCR amplification.

Library concentration and fragment size were assessed using Qubit 4.0 and Qsep400 instruments, respectively. Wuhan Metware Biotechnology Co., Ltd. performed 150-bp paired-end sequencing on an Illumina NovaSeq 6000 (Illumina, San Diego, CA, USA; Additional file 3).

Metagenomic assembly and gene catalog construction

Reads were quality-controlled using fastp [16] without host-reference filtering and assembled individually using MEGAHIT v1.2.9 [17]. Bowtie2 [18] mapped reads to each assembly; unmapped paired-end reads were pooled for additional assembly. QUAST [19] evaluated assemblies, and MetaGeneMark v3.38 [20] predicted coding sequences, excluding those shorter than 100 nt.

CD-HIT v4.8.1 [21] generated a non-redundant gene catalog. Bowtie2 mapped reads to the catalog, and genes with mapped-read counts ≤ 2 were excluded from abundance estimation. Gene counts were divided by gene length and normalized to their within-sample sum; abundances of genes assigned to each taxon were summed.

DIAMOND v2.1.9.163 [22] searched predicted proteins against microbial NR sequences (May 2022; blastp, E-value $\leq 10^{-5}$). Hits within ten times the minimum E-value were retained for lowest common ancestor assignment. Functional annotation used KEGG (May 2024), eggNOG v5.0, CAZy (October 2023) and GO (August 2022). DiTing v0.7 [23] used HMM-based annotation and gene TPM values: genes assigned to each KO were summed, and pathway scores followed pathway-specific formulas. Diagrams used the mean of five composites per site.

Bioinformatics and statistical analyses

Combined-community alpha diversity was calculated from species-level read counts after excluding Others, using R 4.2.0, phyloseq 1.40.0 [24] and vegan 2.6-2. Observed richness counted taxa with positive counts; Shannon used natural logarithms and Simpson was defined as $1 - \Sigma p_i^2$. No rarefaction or sequencing-depth adjustment was applied. Gaussian models fitted by restricted maximum likelihood included a random site intercept and separate fixed effects for site-mean and site-centered pH. Benjamini–Hochberg (BH) correction [25] covered three indices separately for each effect level. Sensitivity checks included site-mean correlations, residual diagnostics, leave-one-site-out analyses and whole-site bootstrapping. Detailed computations, diagnostics and the complete alpha-diversity effect plot (Fig. S1) are provided in Additional file 4.

Co-abundance modules were identified using R 4.2.2 and WGCNA 1.71 [26] after excluding Others and filtering features (power = 4; minimum module size = 30; merging cut height = 0.25; Fig. S2). Across 25 sites, eigengenes and continuous metadata were averaged, whereas LC used the median. Associations used Pearson correlation, except Spearman correlation for LC, with BH correction across 88 tests. Random-intercept models separated between-site and within-site module associations with pH, LL and PL (Additional file 5).

Species-level Bray–Curtis PCoA was performed separately for archaea, bacteria, fungi and viruses. PERMANOVA [27] tested the site factor using 9,999 unrestricted sample-label permutations; PERMDISP [28] assessed distances to site spatial medians. BH correction covered four groups separately for each test family (Additional file 6).

All turquoise and pink module members were screened against 89 nutrient-transformation KOs using Pearson correlations of $\log_{10}(x + 10^{-10})$-transformed relative abundances across 125 samples. Associations were retained at $|r| > 0.7$ and BH $q < 0.001$ across all comparisons within each cycle. Sensitivity analyses used site means of log-transformed abundances and site-demeaned values, with separate within-cycle corrections (Additional files 7 and 8).

Mantel tests used R 4.2.0 and ggcor 0.9.8.1 across 25 sites, with mean abundances and continuous metadata and median LC. Pearson Mantel correlations related phylum-level Bray–Curtis dissimilarities to

individual-variable Euclidean distances, with BH correction across 32 tests. The metadata heatmap shows descriptive Pearson correlations and unadjusted *P* values.

RDA used original-scale abundances of 151 phyla across 125 samples (R 3.5.1; vegan 2.5-6), retaining predictors with VIF < 10. Complementary RDA of 25 site-mean profiles excluded LL through iterative VIF filtering and retained seven predictors. Overall and marginal tests used 9,999 permutations; BH correction covered seven marginal tests. Envfit R^2 represented vector fits rather than independent explanatory fractions. The Python implementation is described in Additional file 4.

The 140 candidate taxa were analysed using Pearson correlation (platform threshold 0.6; nominal $P < 0.05$) and Louvain community detection for Zi–Pi analysis [29]. Independent BH filtering ($|r| > 0.6$; $q < 0.05$ across all pairs) reproduced the platform's nodes and edges. Platform community assignments and roles were used in Figure 6 (Additional file 8).

Results

Variation in plant phenotypic traits across sampling sites

Plant samples and corresponding rhizosphere soils were collected from 25 locations across Heilongjiang Province, China. Plant height (PH), leaf length (LL), leaf width (LW) and petiole length (PL) varied among sites (Fig. 1A). Each plotted composite-sample trait record represented the mean of the corresponding five plants.

Plant height varied substantially among sites, with the highest site mean observed at site UU (49.72 cm) and the lowest at site OO (24.86 cm). Leaf length showed similar geographic variation, with sites UU and HH exhibiting the highest site means (34.96 and 34.30 cm, respectively), whereas site OO showed the lowest value (18.37 cm). Leaf width also varied among locations, with site UU displaying the highest value (7.38 cm) and site OO the lowest (3.70 cm). Petiole length showed distinct spatial variation, reaching the maximum at site HH (14.98 cm) and the minimum at site BB (4.60 cm). Collectively, these summaries describe variation in plant morphology among the sampled sites.

Principal component analysis (PCA) summarized phenotypic variation across the 125 composite-sample trait records from the 25 sampling sites (Fig. 1B). The first two principal components explained 75.3% of the total variation, with PC1 and PC2 accounting for 56.5% and 18.8%, respectively. The ordination

showed differences and overlap among site profiles; it was interpreted as a descriptive summary rather than a test of independent geographic effects.

Sample-level Pearson correlations among the four morphological traits are summarized descriptively in Figure 1C. Plant height covaried positively with leaf length ($r = 0.81$), leaf width ($r = 0.64$) and petiole length ($r = 0.54$). Leaf length was positively correlated with petiole length ($r = 0.76$) and leaf width ($r = 0.57$), whereas the petiole length–leaf width correlation was weaker ($r = 0.17$). These pooled correlations combine between-site and within-site variation.

Spatial variation in rhizosphere microbial communities

Metagenomic annotation identified 10,906 species-level taxa: 10,444 bacteria, 106 archaea, 308 fungi and 48 viruses (Additional file 9). These are annotations detected by a total-DNA metagenomic workflow, not direct measurements of cell counts, viability or activity.

At the phylum level, microbial communities exhibited pronounced group-specific compositional patterns. Bacterial assemblages were dominated by Pseudomonadota, Actinomycetota, and Acidobacteriota, representing prominent soil-associated annotations. Fungal communities were dominated by Mucoromycota, Ascomycota, and Basidiomycota, whereas archaeal communities were characterized by Euryarchaeota, Nitrososphaerota, and Bathyarchaeota. Viral assemblages were mainly represented by Uroviricota and Pisuviricota (Fig. 2A).

At the genus level, bacterial communities exhibited relatively consistent taxonomic patterns across sampling sites, with *Bradyrhizobium*, *Rhodoplanes*, and *Mycobacterium* representing the dominant and widely distributed bacterial genera (Fig. 2B). Fungal communities were primarily dominated by *Rhizopus*, *Aspergillus*, and *Termitomyces*, whereas archaeal communities included *Haloferax*, *Nitrososphaera*, and *Halorubrum*, together with other archaeal annotations. In contrast, viral communities displayed greater variation in dominant genera among sampling sites, with *Iflavirus*, *Minunavirus*, and *Gammaretrovirus* representing major viral groups.

Species-level Bray–Curtis dissimilarities were associated with sampling site in bacteria (PERMANOVA $R^2 = 0.873$), archaea (0.890), fungi (0.838) and viruses (0.833; all $P = q = 0.0001$ under 9,999 permutations). Multivariate dispersion also differed among sites for bacteria, archaea and fungi ($q = 0.0012$, 0.0026 and 0.0043), but no significant dispersion difference was detected for viruses ($q = 0.598$). The first three

PERMANOVA results therefore cannot be attributed solely to shifts in community centroids. These tests describe differences among sites and do not test a pH effect (Fig. 3; Additional file 6).

Soil pH ranged from 4.65 to 6.91 (Additional file 2). Across 25 site means, observed richness was negatively correlated with pH ($r = -0.594$, $P = 0.00175$, $q = 0.00524$). Shannon showed weaker support ($r = -0.404$, $P = 0.0452$, $q = 0.0678$), and Simpson was not significantly associated with pH ($r = -0.217$, $P = 0.298$, $q = 0.298$). These q values cover the three diversity outcomes; the associations were not adjusted for sequencing depth.

The random-intercept model estimated a between-site richness slope of −693.86 taxa per pH unit (95% CI −1099.24 to −288.47; $P = 0.00175$, $q = 0.00524$), compared with a within-site slope of −33.94 (95% CI −195.56 to 127.67; $P = 0.678$). The whole-site bootstrap interval for the between-site slope was −1155.63 to −390.83. Residual diagnostics indicated departures from normality, so model-based intervals were interpreted alongside the bootstrap results. For Shannon diversity, the model-based between-site q value was 0.0678 although the unadjusted bootstrap interval excluded zero; this difference did not support an unqualified claim of a robust association. Observed-richness effects are shown in Figure 4A; the complete three-index comparison is provided in Figure S1 and Additional file 10.

Association analysis of the rhizosphere microbiota with environmental factors and plant phenotypes

Geographic and trait associations are summarized in Figures 4 and 5. Climatic and soil-resource variation can accompany geographic patterns in microbial diversity [30], so coordinate associations do not establish a direct geographic mechanism or separate unmeasured covariates.

Co-abundance analysis identified 11 assigned modules and a grey group (Fig. 4B). Across 25 sites, turquoise and pink eigengenes were negatively correlated with mean pH ($r = -0.761$ and −0.755; both $q = 0.000374$ across 88 module–metadata tests). Random-intercept models supported between-site pH slopes of −0.1452 (95% CI −0.1986 to −0.0918) and −0.1478 (−0.2031 to −0.0925) in exported eigengene units per pH unit (both $q = 0.000140$ across 33 between-site tests). Within-site pH associations were not significant ($P = 0.823$ and 0.177). Whole-site bootstrap intervals supported the between-site associations despite non-Gaussian residuals. Corresponding LL and PL associations were not supported by the site-structured tests. Figure 4C shows the pH-related correlations between |MM| and |GS| across taxa in the

turquoise and pink modules and does not provide independent evidence of sample-level trait effects. Complete module statistics are provided in Additional file 5.

The site-level Mantel results supported an association between archaeal community dissimilarity and pH differences across 25 sites ($r = 0.547$, $P = 0.001$, $q = 0.0160$; Fig. 5A). Longitude differences were associated with bacterial ($r = 0.343$, $P = 0.002$, $q = 0.0213$) and fungal community dissimilarities ($r = 0.359$, $P = 0.001$, $q = 0.0160$). These were the only three associations retained at $q < 0.05$ across the 32 comparisons (Additional file 6). Archaeal dissimilarity was associated with LC before correction ($P = 0.041$), but this association did not remain significant after BH adjustment ($q = 0.328$). Site aggregation addresses the sampling hierarchy but does not separate pH from spatial covariation or other environmental factors.

The original 125-sample RDA of 151 microbial phyla retained all eight variables at VIF < 10. RDA1 and RDA2 accounted for 35.8254% and 7.2288% of total variation (Fig. 5B). Source envfit values for longitude, pH and latitude were $R^2 = 0.498$, 0.465 and 0.396; these are vector fits rather than independent explanatory fractions. In the supplementary 25-site RDA, LL was removed by iterative VIF filtering and seven predictors were retained. The first two constrained axes accounted for 43.90% and 8.71% of total variation (Fig. S3). The overall model was significant ($F = 3.378$, $P = 0.0016$; adjusted $R^2 = 0.410$), but the marginal pH test conditional on the other retained variables was not significant ($P = 0.0678$, $q = 0.152$). No retained term had $q < 0.05$ across the seven marginal tests. Thus, this RDA did not establish an independent pH contribution.

Positions of phyla relative to vectors in the original ordination describe multivariate associations. They were not interpreted as evidence that an individual phylum directly affects plant growth.

Candidate taxon associations with nutrient transformation genes

Screening 2,127 module members against 89 KOs retained three nitrogen-associated taxa, 109 carbon-associated taxa and 43 sulfur-associated taxa at $|r| > 0.7$ and within-cycle $q < 0.001$ (Additional files 7 and 8). These comprised 3, 164 and 43 associations involving two nitrogen, 16 carbon and four sulfur KOs, respectively, and 140 distinct taxa in total. All 210 retained associations had the same direction and $q < 0.05$ in the between-site sensitivity analysis. Only 195 also met the original strict between-site $|r| > 0.7$ and $q < 0.001$ criteria (0 nitrogen, 155 carbon and 40 sulfur associations); 151 had same-sign within-site $q <$

0.05. These different thresholds are reported separately and do not establish genomic linkage between taxa and KOs.

The candidate set comprised 123 turquoise and 17 pink taxa (Additional file 8). Six candidates had no retained taxon–taxon edges, leaving 134 displayed nodes and 5,231 positive edges. The network included 109 Actinomycetota (81.3%), 13 Pseudomonadota (9.7%), seven Acidobacteriota (5.2%) and five other or unclassified taxa (3.7%; Fig. 6A–B). The maximum node degree was 110. The Louvain partition yielded communities of 17, 67 and 50 nodes; these communities are distinct from WGCNA modules.

The platform Zi and Pi output classified all 134 network nodes as broadly peripheral, including 117 peripheral and 17 ultra-peripheral nodes. No module hubs or connectors were identified under the specified criteria (Fig. 6C). These topological classifications do not validate ecological keystone functions or microbial effects on plant growth. The network remains an exploratory graph of sample-level correlations and does not constitute a test of direct microbial interactions.

Metagenomic functional potential for nutrient transformations

Functional annotation summarized genes assigned to metabolic categories, including carbohydrate and amino acid metabolism (Fig. S4). Carbon, nitrogen and sulfur transformation categories were summarized using DiTing v0.7. The pathway diagrams in Figure 7 provide a framework for displaying these annotations. The classification inventory is distinct from the selected set of 89 KOs used in the taxon–KO association analysis.

The functional abundance table contained 6,994 standard KO identifiers and 14 named markers across 125 samples. Positive annotations included DsyB_prok, DSYB_euka, DddP, DddQ, DmdB, DmdC and AcuH, indicating potential for parts of DMSP-related transformations. DiTing includes a curated DMSP gene component alongside KEGG-based annotations [23]; marker detection does not establish complete pathways or activity. Canonical *nifH*, *nifD* and *nifK* KOs (K02588, K02586 and K02591), and the DmdA KO K17486, were not present in the available table. Consequently, complete canonical nitrogen fixation and a complete DMSP demethylation pathway were not inferred. Absence from this annotation table is not evidence of biological absence.

Discussion

This survey describes habitat-associated variation in the *A. victorialis* rhizosphere across 25 sites. Environmental heterogeneity can accompany variation in understory vegetation [31], while plant phenotypic responses can involve plasticity and genetic variation [32, 33]. The present design cannot separate these contributions. More generally, trait–function relationships can be context dependent [34, 35, 36]. Correlations among leaf length, plant height and petiole length therefore describe coordinated morphology without establishing adaptation, biomass allocation or microbial promotion of growth.

Soil pH has been associated with microbial variation in other systems [10, 37, 38, 39]. Here, archaeal community dissimilarity and two co-abundance modules were associated with pH between sites; the corresponding within-site module slopes were not significant. The observed-richness association was also primarily between sites, but it was not adjusted for sequencing depth. Whole-site resampling addresses sampling variation across sites and does not remove depth-related detection bias. Accordingly, this result concerns detected taxa under the analysed sampling conditions and cannot establish a depth-independent loss of microbial richness. Shannon support depended on inferential choices and Simpson showed no significant between-site association. The supplementary multivariable RDA did not isolate a significant marginal pH contribution. Rhizosphere functions can vary in space and time [40], and cultivation studies in other medicinal plants provide context [41], but these studies do not establish the same processes or a dominant causal pH effect in *A. victorialis*.

Metagenomic annotation identified genes associated with carbon, nitrogen and sulfur transformations. Correlations between these genes and module members identify candidate taxa for further study, but do not demonstrate that a correlated taxon carries or expresses a particular gene. Genomic linkage, isolate characterization or targeted experimental assays would be needed to evaluate those relationships. Likewise, co-occurrence network connectivity does not by itself establish ecological interactions or keystone function. Actinobacteria include taxa with plant growth-promoting traits [42] and acid-associated isolates [43], but these properties cannot be assigned to all detected taxa from their taxonomy alone. Studies of nutrient transformation pathways in other environments illustrate possible functional connections [44, 45], while pathway activity in the present system remains unmeasured.

Several limitations constrain interpretation. Sampling was observational, restricted to one province and one period. Site-level models and whole-site resampling addressed particular aspects of nesting without removing spatial autocorrelation or unmeasured environmental and host variation. The richness models did not include sequencing depth, and the supplied alpha output did not document a common-depth standardization; observed richness may therefore reflect differences in detection effort. Identical Chao1, ACE and observed values, together with coverage fixed at 1, do not independently establish sampling completeness. Residuals in the Gaussian richness and module models departed from normality; bootstrap results were interpreted alongside model-based tests. Site dispersions differed for bacteria, archaea and fungi, complicating interpretation of PERMANOVA location effects. Relative-abundance correlations are vulnerable to compositional and other shared-covariate effects and do not establish taxon–gene linkage. Soil nutrients, moisture, temperature and host genetic variation were not comprehensively measured. Total-DNA metagenomes provide incomplete coverage of viral diversity compared with dedicated viromes [46] and do not establish active infection or the life-strategy changes considered in other work [47]. Host-reference filtering was not performed, so effects of residual plant sequences on annotation and abundance estimates could not be excluded. No inoculation or controlled pH experiment was included, and gene abundance does not measure expression or process rates.

Conclusions

Across 25 *A. victorialis* habitats, soil pH differences were associated with archaeal community dissimilarity and site-mean pH was associated with two microbial co-abundance modules. The observed-richness association was primarily between sites and was not adjusted for sequencing depth; it therefore cannot establish a depth-independent decline in ecological richness. Shannon and Simpson did not support a uniform decline after the specified adjustment. Functional annotations identified nutrient-transformation potential, and taxon–KO screening generated candidates for future validation. These observational results do not establish independent causal pH effects, complete pathway activity or microbial promotion of plant adaptation.

Abbreviations

CAZy, Carbohydrate-Active enZYmes; DMSP, dimethylsulfoniopropionate; FDR, false discovery rate; GS, gene significance; KEGG, Kyoto Encyclopedia of Genes and Genomes; KO, KEGG Orthology; LC, leaf color; LL, leaf length; LW, leaf width; MM, module membership; PCA, principal component analysis; PCoA, principal coordinates analysis; PERMANOVA, permutational multivariate analysis of variance; PH, plant height; PL, petiole length; RDA, redundancy analysis; SRA, Sequence Read Archive; WGCNA, weighted gene co-expression network analysis.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The raw metagenomic sequencing reads are publicly available in the NCBI Sequence Read Archive under BioProject accession PRJNA1482359 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1482359). Sampling information, metadata, processed data and statistical results are provided in Additional files 1–3 and 5–11. Supplementary methods and figures are provided in Additional file 4. The complete species-level relative-abundance matrix is provided as a CSV file within the ZIP archive in Additional file 12. The species-level read-count matrix used to calculate alpha diversity, excluding the aggregate “Others” category, is provided in Additional file 13. The complete TPM matrix for KOs and named functional markers is provided in Additional file 14.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Key Research and Development Program of China (2021YFD1200200), the National Natural Science Foundation of China (32573022, 31872946, 32172566 and 32272731), the Innovation Engineering Project of the Chinese Academy of Agricultural Sciences (CAAS-ASTIP-2021-IVF), the Safe Preservation Project of Crop Germplasm Resources of MOF (2024NWB037), the National Horticultural Germplasm Centre Project (NHGRC2024-NH01), and the earmarked fund for CARS (CARS-22-A-12).

Authors' contributions

JW: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Visualization, Writing – original draft, and Writing – review and editing.

HZ, XW, XZ, RW, YH and WY: Investigation, Resources, and Data curation.

JS: Methodology, Software, Formal analysis, and Investigation.

HJ: Conceptualization and Resources.

HW: Conceptualization, Supervision, Project administration, Funding acquisition, and Writing – review and editing.

Acknowledgements

We thank all those who assisted with sample collection, laboratory work and data analysis.

References

1. Shirataki Y, Motohashi N, Tani S, Sunaga K, Sakagami H, Satoh K, et al. Antioxidative activity of *Allium victorialis* L. extracts. Anticancer Res. 2001;21:3331–9.

2. Woo KW, Moon E, Park SY, Kim SY, Lee KR. Flavonoid glycosides from the leaves of *Allium victorialis* var. *platyphyllum* and their anti-neuroinflammatory effects. Bioorg Med Chem Lett. 2012;22:7465–70. https://doi.org/10.1016/j.bmcl.2012.10.043.

3. Feng B, Xu G, Wang X, Su X, Liu J, Li C, et al. The relationship between habitat factors and the nutrient contents of wild *Allium victorialis* L. in the Changbai Mountains. Not Bot Horti Agrobo. 2023;51:13194. https://doi.org/10.15835/nbha51313194.

4. Xia C, Huang Y, Qi Y, Yang X, Xue T, Hu R, et al. Developing long-term conservation priority planning for medicinal plants in China by combining conservation status with diversity hotspot analyses and climate change prediction. BMC Biol. 2022;20:89. https://doi.org/10.1186/s12915-022-01285-4.

5. Godefroid S, Piazza C, Rossi G, Buord S, Stevens A-D, Aguraiuja R, et al. How successful are plant species reintroductions? Biol Conserv. 2011;144:672–82. https://doi.org/10.1016/j.biocon.2010.10.003.

6. Kardol P, Wardle DA. How understanding aboveground–belowground linkages can assist restoration ecology. Trends Ecol Evol. 2010;25:670–9. https://doi.org/10.1016/j.tree.2010.09.001.

7. Berendsen RL, Pieterse CMJ, Bakker PAHM. The rhizosphere microbiome and plant health. Trends Plant Sci. 2012;17:478–86. https://doi.org/10.1016/j.tplants.2012.04.001.

8. David AS, Quintana-Ascencio PF, Menges ES, Thapa-Magar KB, Afkhami ME, Searcy CA. Soil Microbiomes Underlie Population Persistence of an Endangered Plant Species. Am Nat. 2019;194:488–94. https://doi.org/10.1086/704684.

9. Rigg JL, Offord CA, Singh BK, Anderson I, Clarke S, Powell JR. Soil microbial communities influence seedling growth of a rare conifer independent of plant–soil feedback. Ecology. 2016;97:3346–58. https://doi.org/10.1002/ecy.1594.

10. Naik S, Chapin LJ, Jones ML. Plant species and pH interactions shape rhizosphere bacterial diversity in peat-based systems. Appl Soil Ecol. 2026;217:106659. https://doi.org/10.1016/j.apsoil.2025.106659.

11. Wan W, Tan J, Wang Y, Qin Y, He H, Wu H, et al. Responses of the rhizosphere bacterial community in acidic crop soil to pH: Changes in diversity, composition, interaction, and function. Sci Total Environ. 2020;700:134418. https://doi.org/10.1016/j.scitotenv.2019.134418.

12. Liu Z, Ying J, Liu C. Changes in Rhizosphere Soil Microorganisms and Metabolites during the Cultivation of *Fritillaria cirrhosa*. Biology. 2024;13:334. https://doi.org/10.3390/biology13050334.

13. Wang B, Lin Y, Yu W, Xia Q, Ali A, Wei F, et al. The loss of microbial autotoxin degradation functions is associated with the decline of beneficial bacterial agents induced by phenolic acids. Microbiol Spectr. 2023;11:e03380-22. https://doi.org/10.1128/spectrum.03380-22.

14. Winiarczyk K, Skrzypczak K, Jaroszuk-Sciseł J, Bocianowski J. Investigations of the capacity and strength of seed germination in *Allium victorialis* L. Acta Soc Bot Pol. 2014;83:219–28. https://doi.org/10.5586/asbp.2014.021.

15. Jung S-Y, Lee M-H, Oh T-K, Yoon J-H. *Herbaspirillum rhizosphaerae* sp. nov., isolated from rhizosphere soil of *Allium victorialis* var. *platyphyllum*. Int J Syst Evol Microbiol. 2007;57:2284–8. https://doi.org/10.1099/ijs.0.64666-0.

16. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34:i884–90. https://doi.org/10.1093/bioinformatics/bty560.

17. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31:1674–6. https://doi.org/10.1093/bioinformatics/btv033.

18. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9. https://doi.org/10.1038/nmeth.1923.

19. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: Quality assessment tool for genome assemblies. Bioinformatics. 2013;29:1072–5. https://doi.org/10.1093/bioinformatics/btt086.

20. Zhu W, Lomsadze A, Borodovsky M. Ab initio gene identification in metagenomic sequences. Nucleic Acids Res. 2010;38:e132. https://doi.org/10.1093/nar/gkq275.

21. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28:3150–2. https://doi.org/10.1093/bioinformatics/bts565.

22. Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods. 2021;18:366–8. https://doi.org/10.1038/s41592-021-01101-x.

23. Xue CX, Lin H, Zhu XY, Liu J, Zhang Y, Rowley G, et al. DiTing: A pipeline to infer and compare biogeochemical pathways from metagenomic and metatranscriptomic data. Front Microbiol. 2021;12:698286. https://doi.org/10.3389/fmicb.2021.698286.

24. McMurdie PJ, Holmes S. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One. 2013;8:e61217. https://doi.org/10.1371/journal.pone.0061217.

25. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc Series B. 1995;57:289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

26. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559. https://doi.org/10.1186/1471-2105-9-559.

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

28. Anderson MJ. Distance-based tests for homogeneity of multivariate dispersions. Biometrics. 2006;62:245–53. https://doi.org/10.1111/j.1541-0420.2005.00440.x.

29. Guimerà R, Amaral LAN. Functional cartography of complex metabolic networks. Nature. 2005;433:895–900. https://doi.org/10.1038/nature03288.

30. Tian J, He N, Hale L, Niu S, Yu G, Liu Y, et al. Soil organic matter availability and climate drive latitudinal patterns in bacterial diversity from tropical to cold temperate forests. Funct Ecol. 2018;32:61–70. https://doi.org/10.1111/1365-2435.12952.

31. Su X, Wang M, Huang Z, Fu S, Chen HYH. Forest Understorey Vegetation: Colonization and the Availability and Heterogeneity of Resources. Forests. 2019;10:944. https://doi.org/10.3390/f10110944.

32. Nicotra AB, Atkin OK, Bonser SP, Davidson AM, Finnegan EJ, Mathesius U, et al. Plant phenotypic plasticity in a changing climate. Trends Plant Sci. 2010;15:684–92. https://doi.org/10.1016/j.tplants.2010.09.008.

33. Savolainen O, Lascoux M, Merilä J. Ecological genomics of local adaptation. Nat Rev Genet. 2013;14:807–20. https://doi.org/10.1038/nrg3522.

441 34. Funk JL, Larson JE, Ames GM, Butterfield BJ, Cavender-Bares J, Firn J, et al. Revisiting the Holy
442 Grail: using plant functional traits to understand ecological processes. Biol Rev Camb Philos Soc.
443 2017;92:1156–73. https://doi.org/10.1111/brv.12275.

444 35. Kerr KL, Fickle JC, Anderegg WRL. Decoupling of functional traits from intraspecific patterns of
445 growth and drought stress resistance. New Phytol. 2023;239:174–88. https://doi.org/10.1111/nph.18937.

446 36. Anderegg LDL. Why can't we predict traits from the environment? New Phytol. 2023;237:1998–2004.
447 https://doi.org/10.1111/nph.18586.

448 37. Zhang N, Ma Z, Li D, Ni H, Sun B, Liang Y. Soil pH Filters the Association Patterns of Aluminum-
449 Tolerant Microorganisms in Rice Paddies. mSystems. 2022;7:e01022-21.
450 https://doi.org/10.1128/msystems.01022-21.

451 38. Tripathi BM, Stegen JC, Kim M, Dong K, Adams JM, Lee YK. Soil pH mediates the balance between
452 stochastic and deterministic assembly of bacteria. ISME J. 2018;12:1072–83.
453 https://doi.org/10.1038/s41396-018-0082-4.

454 39. Fierer N, Jackson RB. The diversity and biogeography of soil bacterial communities. Proc Natl Acad
455 Sci USA. 2006;103:626–31. https://doi.org/10.1073/pnas.0507535103.

456 40. Nuccio EE, Starr E, Karaoz U, Brodie EL, Zhou J, Tringe SG, et al. Niche differentiation is spatially
457 and temporally regulated in the rhizosphere. ISME J. 2020;14:999–1014. https://doi.org/10.1038/s41396-
458 019-0582-x.

459 41. Gao D, Gao X, Wang Y, Huo H, Wu Y, Yang Z, et al. Effects of long-term continuous cultivation on
460 the structure and function of soil bacterial and fungal communities of *Fritillaria Cirrhosa* on the Qinghai-
461 Tibetan Plateau. Sci Rep. 2024;14:21291. https://doi.org/10.1038/s41598-024-70625-x.

462 42. Boukhatem ZF, Merabet C, Tsaki H. Plant Growth Promoting Actinobacteria, the Most Promising
463 Candidates as Bioinoculants? Front Agron. 2022;4:849911. https://doi.org/10.3389/fagro.2022.849911.

464 43. Poomthongdee N, Duangmal K, Pathom-aree W. Acidophilic actinomycetes from rhizosphere soil:
465 diversity and properties beneficial to plants. J Antibiot. 2015;68:106–14.
466 https://doi.org/10.1038/ja.2014.117.

44. Sun X, Kong T, Huang D, Chen Z, Zhang Y, Häggblom MM, et al. Microbial Sulfur and Arsenic Oxidation Facilitate the Establishment of Biocrusts during Reclamation of Degraded Mine Tailings. Environ Sci Technol. 2024;58:12441–53. https://doi.org/10.1021/acs.est.3c10945.

45. Yang J, Li Q, Chen T, Tao H, Jiang Y. Microbial regulatory mechanisms involved in groundwater arsenic enrichment: Synergistic interactions between key species and genes in C-N-S metabolism. Geosci Front. 2025;16:102132. https://doi.org/10.1016/j.gsf.2025.102132.

46. Santos-Medellin C, Zinke LA, Ter Horst AM, Gelardi DL, Parikh SJ, Emerson JB. Viromes outperform total metagenomes in revealing the spatiotemporal patterns of agricultural soil viral communities. ISME J. 2021;15:1956–70. https://doi.org/10.1038/s41396-021-00897-y.

47. Tong D, Ma B, Hu L, Li Y, Dahlgren RA, Xu J. Soil Virus Life-Strategy Conversion and Implications for Ecosystem and Soil Functions. Glob Chang Biol. 2025;31:e70460. https://doi.org/10.1111/gcb.70460.

Figure legends

Fig. 1. Phenotypic variation across 25 *Allium victorialis* sampling sites. (A) Distributions of leaf length, leaf width, plant height and petiole length, all measured in cm. Each site has five composite-sample trait records, each representing the mean of the corresponding five plants. (B) Descriptive PCA of the 125 trait records; PC1 and PC2 explain 56.5% and 18.8% of variation. (C) Descriptive Pearson correlations among the four traits. Pooled correlations combine within-site and between-site variation. Site codes are defined in Additional file 1.

Fig. 2. Rhizosphere microbial community composition across sampling sites. (A) Relative-abundance profiles of up to 10 dominant phyla in bacteria, fungi, archaea and viruses across 125 composite samples. Other displayed annotations are grouped as Other. The accompanying dendrograms show UPGMA clustering based on Bray–Curtis dissimilarities. (B) Site-level profiles of the 10 displayed dominant genera in each microbial group across sites AA–YY. Phylum and genus profiles are descriptive summaries.

Fig. 3. Species-level community ordination. Panels A–D show archaeal, bacterial, fungal and viral Bray–Curtis PCoA, respectively, across 125 composite samples coloured by site (25 sites; five samples per site). PERMANOVA tested the site factor using 9,999 unrestricted sample-label permutations; all four groups had $q = 0.0001$. PERMDISP q values were 0.0026 for archaea, 0.0012 for bacteria, 0.0043 for fungi and 0.598 for viruses. Axis percentages use the sum of all signed eigenvalues as denominator, matching the source platform plots. Between-site patterns do not establish a pH effect.

Fig. 4. Associations of soil pH with observed richness and microbial co-abundance modules. (A) Between-site and within-site pH slopes for combined-community observed richness from a random-intercept model. Points show estimated slopes; bars show model-based 95% confidence intervals; the dashed vertical line indicates zero. BH *q* values were calculated across three diversity outcomes separately for each effect level. Richness was not adjusted for sequencing depth. The complete three-outcome comparison is shown in Figure S1. (B) Correlations between 11 assigned module eigengenes and eight metadata variables across 25 sites. Eigengenes and continuous traits were averaged within sites; LC was summarized by its median. Pearson correlation was used except for LC, which used Spearman correlation. Cells show *r* (BH-adjusted *q*); adjustment covers all 88 tests. Grey is excluded because it contains unassigned features. (C) Relationships between |MM| and |GS| for soil pH in the turquoise (1,975 taxa) and pink (152 taxa) modules. Each point represents a taxon. These across-taxon correlations are descriptive and do not constitute site-adjusted module–trait tests. LL, leaf length; LW, leaf width; PL, petiole length; PH, plant height; LC, leaf color (1, light green; 2, green; 3, dark green).

Fig. 5. Microbial community associations with plant traits and environmental variables. (A) Pearson Mantel correlations between group-specific phylum-level Bray–Curtis dissimilarities and individual-variable Euclidean distances across 25 sites. Abundance profiles and continuous metadata were averaged within sites; LC used the site median. Line width represents Mantel *r* and line color represents BH-adjusted *q* across 32 comparisons. The metadata heatmap shows descriptive 25-site Pearson correlations; its stars indicate unadjusted *P* values and are distinct from the Mantel *q* values. (B) Descriptive RDA of all 151 microbial phyla across 125 composite samples nested within 25 sites. RDA1 and RDA2 account for 35.83% and 7.23% of total variation. LON denotes longitude and LAT denotes latitude. Vector fits are not interpreted as independent explanatory fractions. Site-level RDA inference is provided in Figure S3 and Additional file 6.

Fig. 6. Taxonomic composition and topology of the candidate co-abundance network. (A) Phylum-level composition of the 134 retained nodes, calculated from node counts rather than abundance weights: 109 Actinomycetota, 13 Pseudomonadota, seven Acidobacteriota and five other or unclassified nodes. (B) The network contains 134 nodes and 5,231 positive associations among the 140 candidate taxa; six candidates without retained edges are omitted. Node size represents degree; node colors identify Louvain communities of 17, 67 and 50 nodes, consistent with panel C. (C) Within-module connectivity (Zi) and participation coefficients (Pi). All 134 nodes are peripheral under the four-role classification; the seven-role classification distinguishes 117 peripheral and 17 ultra-peripheral nodes. No module hubs, network hubs or connectors were identified. Network roles do not establish ecological functions or direct interactions.

Fig. 7. Schematic overview of annotation-based nutrient-transformation potential. Panels show (A) nitrogen, (B) sulfur, (C) carbon and (D) DMSP transformations. Arrows depict predefined metabolic relationships and do not necessarily indicate detection in this study. Grey dashed arrows distinguish the canonical nitrogen-fixation and DmdA-associated background steps; other arrow colors follow the company pathway schematic and do not encode detection status. Pie charts summarize annotation-based pathway abundance estimates. Site colors identify AA–YY, with each site represented by the mean of its five composite samples. Pie sectors represent the relative contributions of the displayed sites to a given pathway; circle sizes represent aggregated pathway abundance according to the panel-specific bins. Abundance estimates are derived from gene TPM values using pathway-specific DiTing v0.7 rules. Supporting marker annotations and abundance data are described in Additional file 11. The diagrams do not establish pathway completeness, microbial gene carriage or process activity.

Additional files

Additional file 1: Sampling sites (XLSX). Coordinates and sampling design for 25 sites.

Additional file 2: Plant traits and environmental metadata (XLSX). Metadata for 125 samples and 25-site summaries; LL, LW, PL and PH in cm, with LC coding.

Additional file 3: Sequencing quality (XLSX). Quality-control metrics for 125 samples and sequencing-yield summaries.

Additional file 4: Supplementary methods and figures (DOCX). Computational details and Figures S1–S4: alpha-diversity effects, WGCNA diagnostics, site-level RDA and functional annotation summaries.

Additional file 5: WGCNA modules and associations (XLSX). Module membership, eigengenes, site-level correlations and mixed-model results.

Additional file 6: Community statistical tests (XLSX). PERMANOVA, PERMDISP, Mantel and RDA results.

Additional file 7: Nutrient-transformation KOs (XLSX). The 89-KO input used for taxon–KO association screening.

Additional file 8: Candidate taxa and networks (XLSX). The 140 candidate taxa, taxon–KO associations, 134 network nodes, 5,231 edges and Zi–Pi roles.

Additional file 9: Taxonomic annotations (XLSX). Taxon identifiers, classifications and abundance-data descriptions.

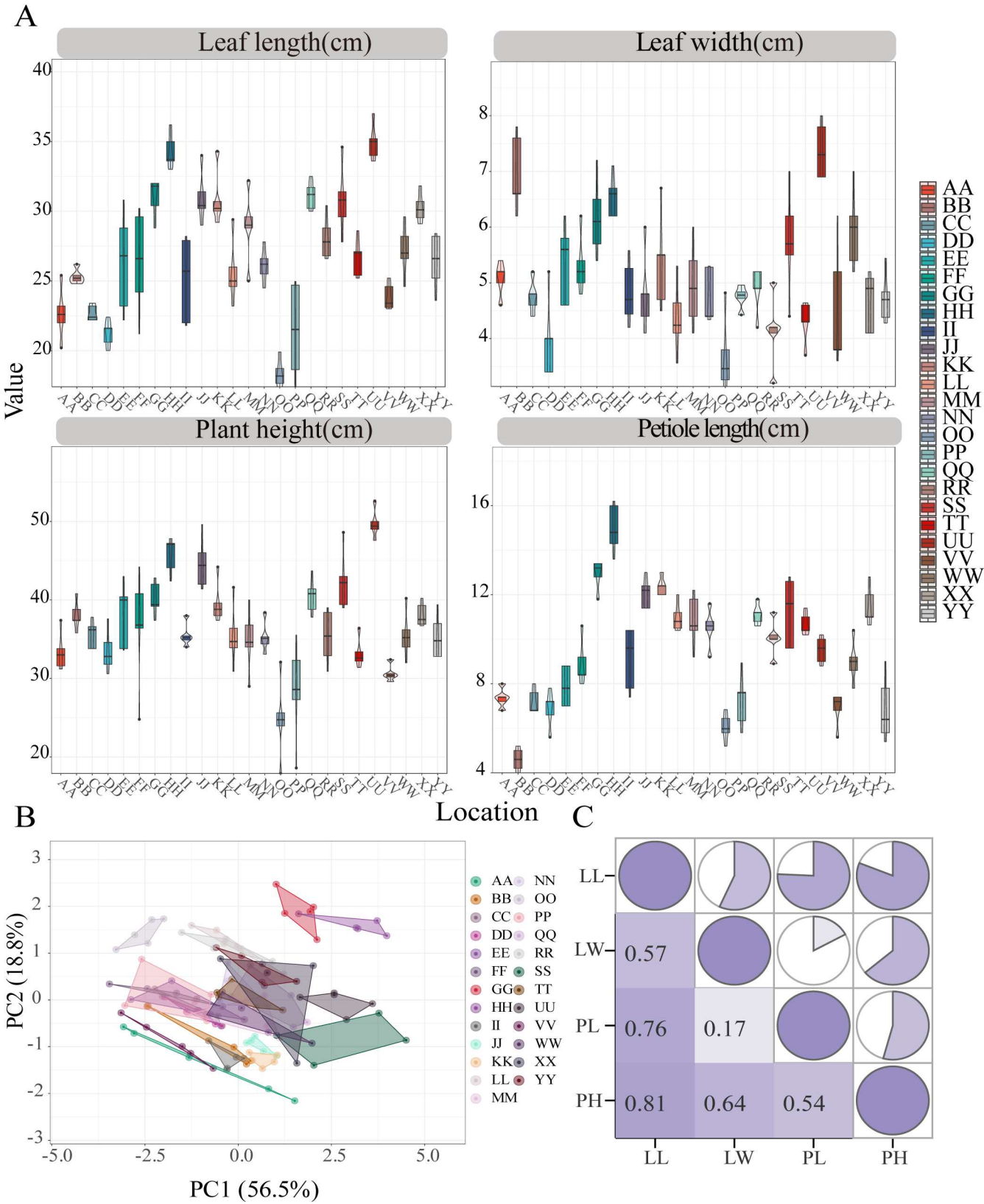
Additional file 10: Alpha diversity and model results (XLSX). Sample-level indices, input provenance, mixed-model results and sensitivity analyses.

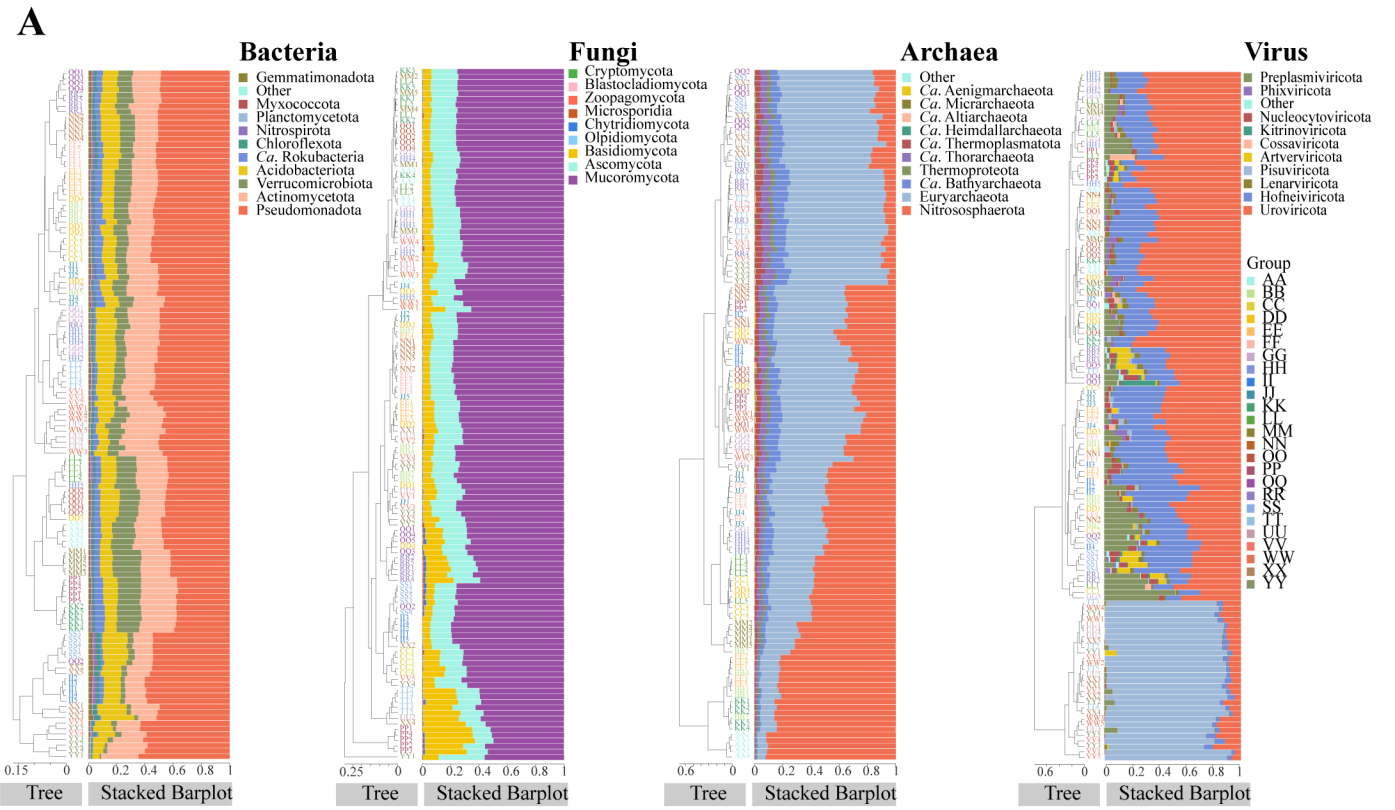
Additional file 11: Functional annotations and pathway abundance (XLSX). Marker evidence, pathway scores for 125 samples, 25-site means and site contributions.

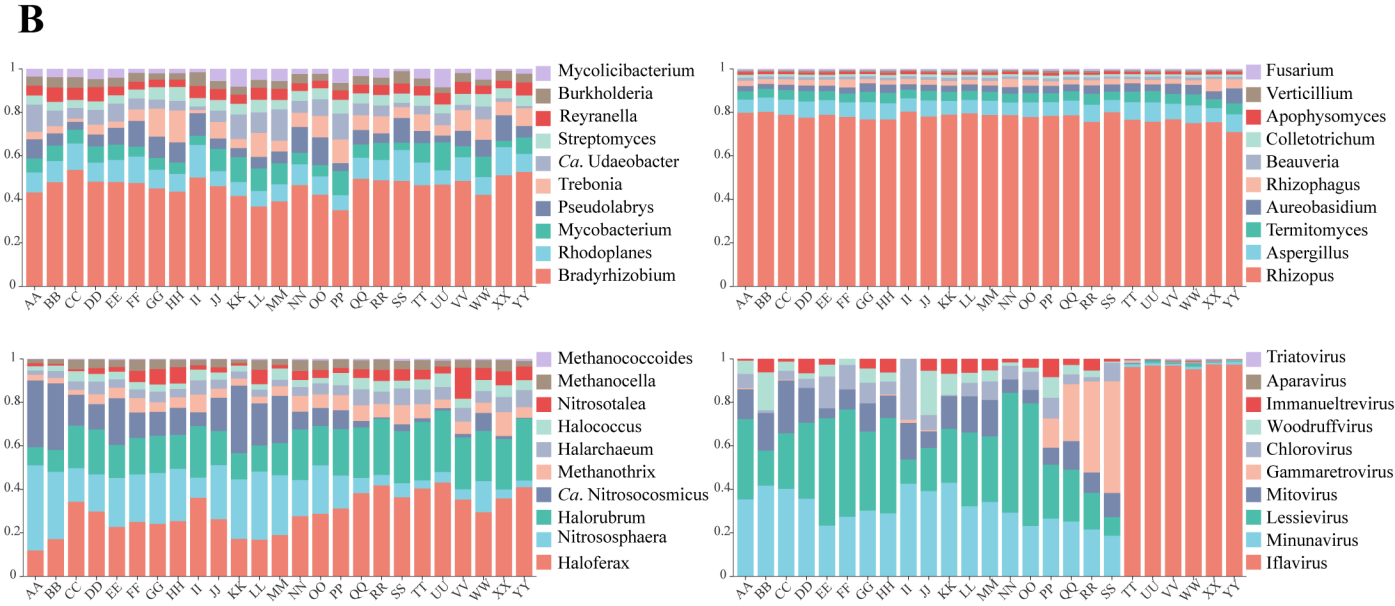
Additional file 12: Species relative abundance matrix (ZIP containing CSV). Relative abundances of 10,906 taxa across 125 samples, with taxonomic annotations.

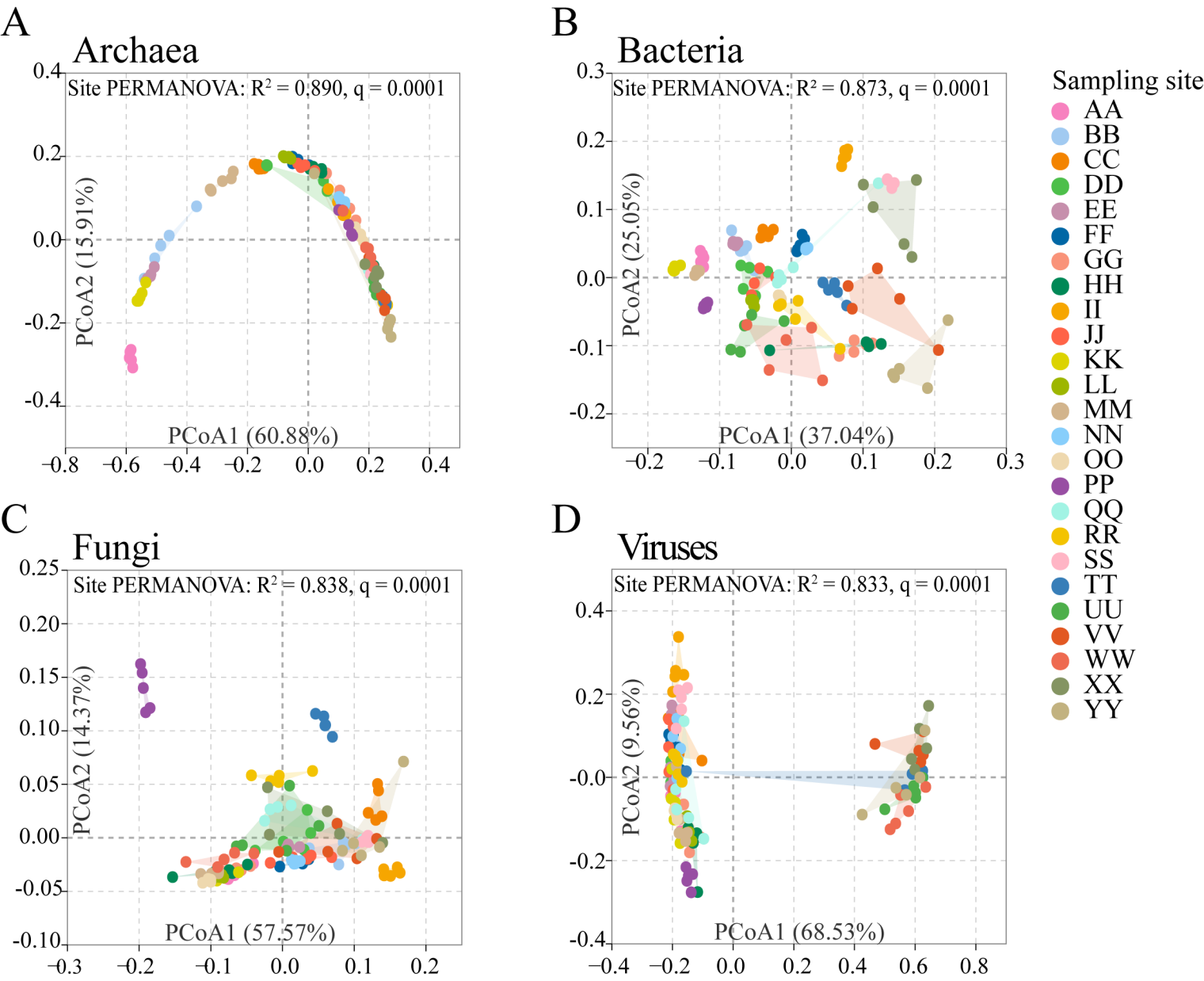
Additional file 13: Species-level read counts (CSV). Integer read counts for 10,906 taxa across 125 samples, excluding the aggregate Others category.

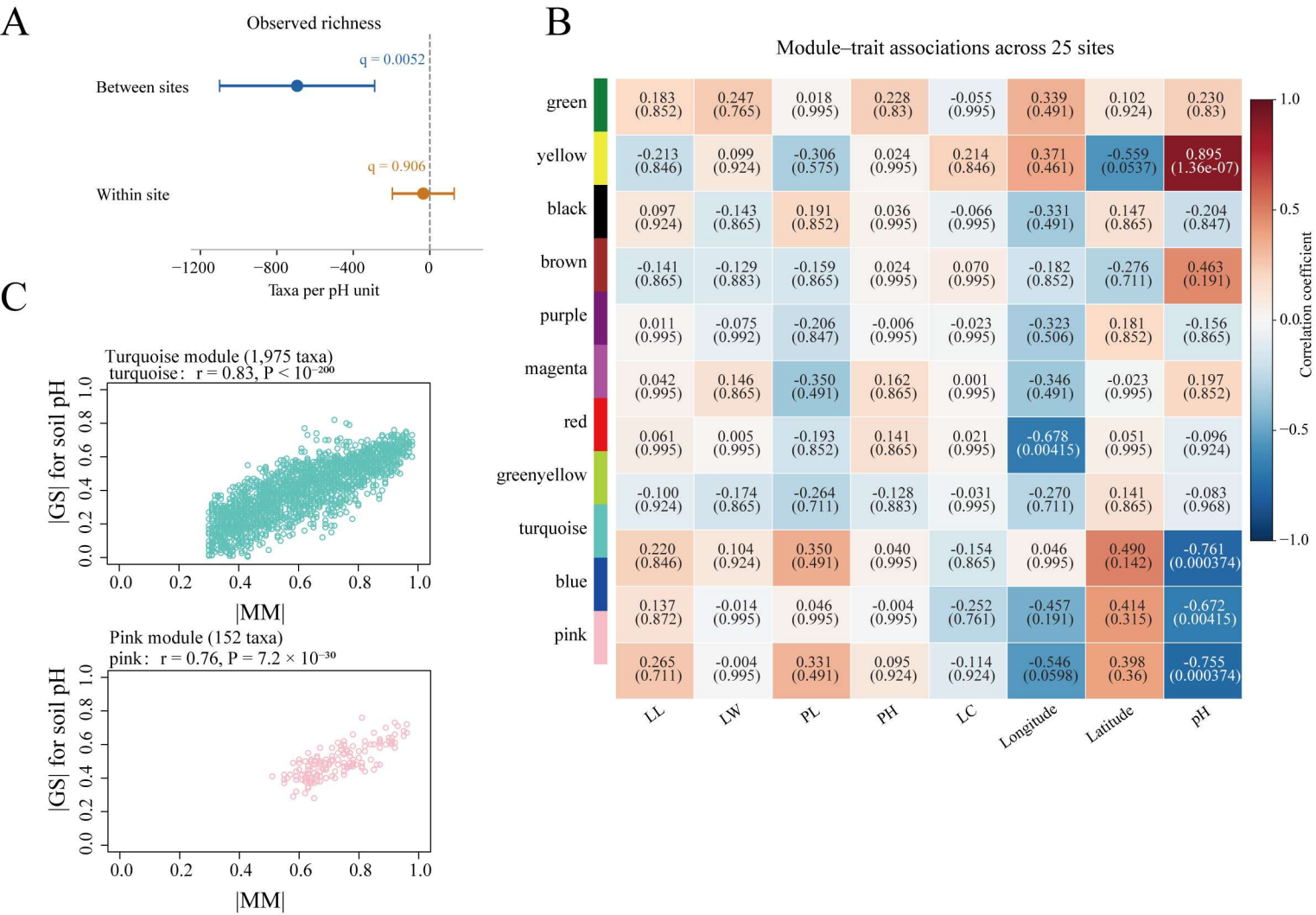
Additional file 14: Functional marker abundance matrix (CSV). TPM values for 6,994 KO identifiers and 14 named markers across 125 samples.

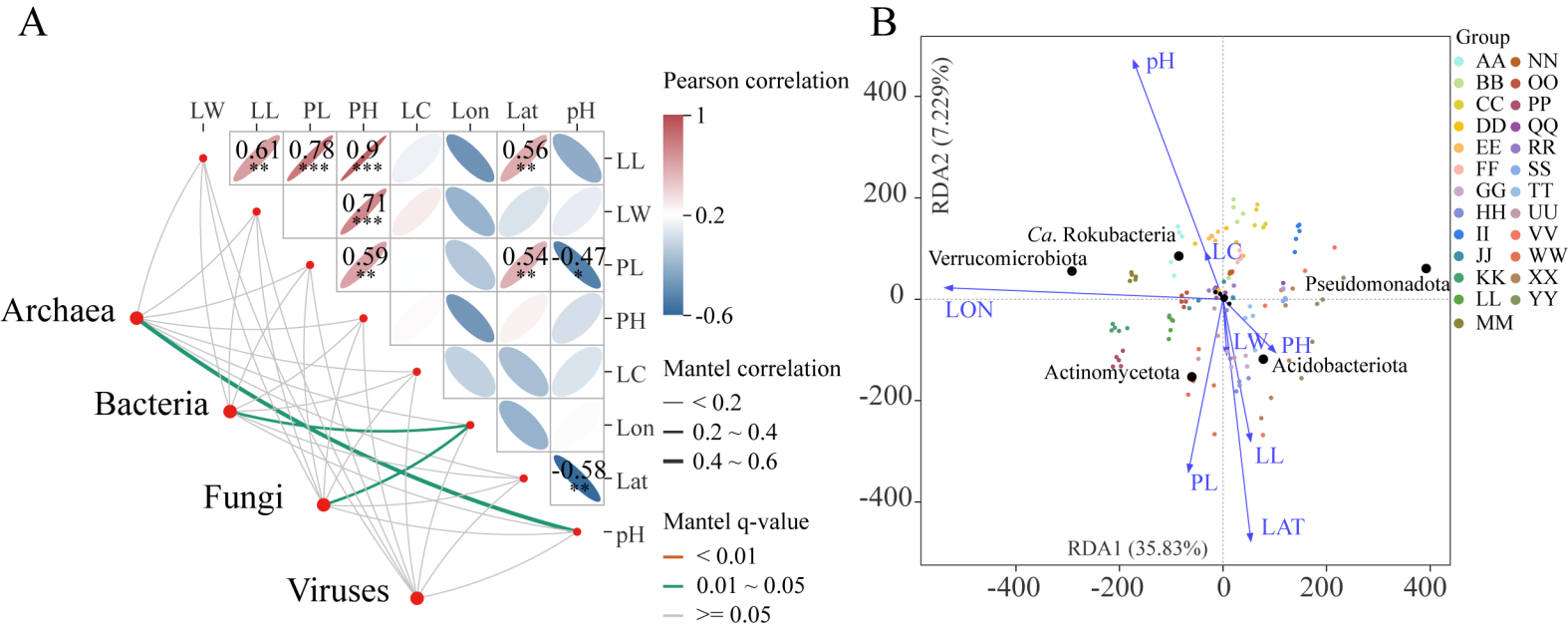

A
Leaf length(cm)
Leaf width(cm)
Plant height(cm)
Petiole length(cm)
Value
Location
AA
BB
CC
DD
EE
FF
GG
HH
II
JJ
KK
LL
MM
NN
OO
PP
QQ
RR
SS
TT
UU
VV
WW
XX
YY
B
PC2 (18.8%)
PC1 (56.5%)
C
LL
LW
PL
PH
0.57
0.76
0.17
0.81
0.64
0.54

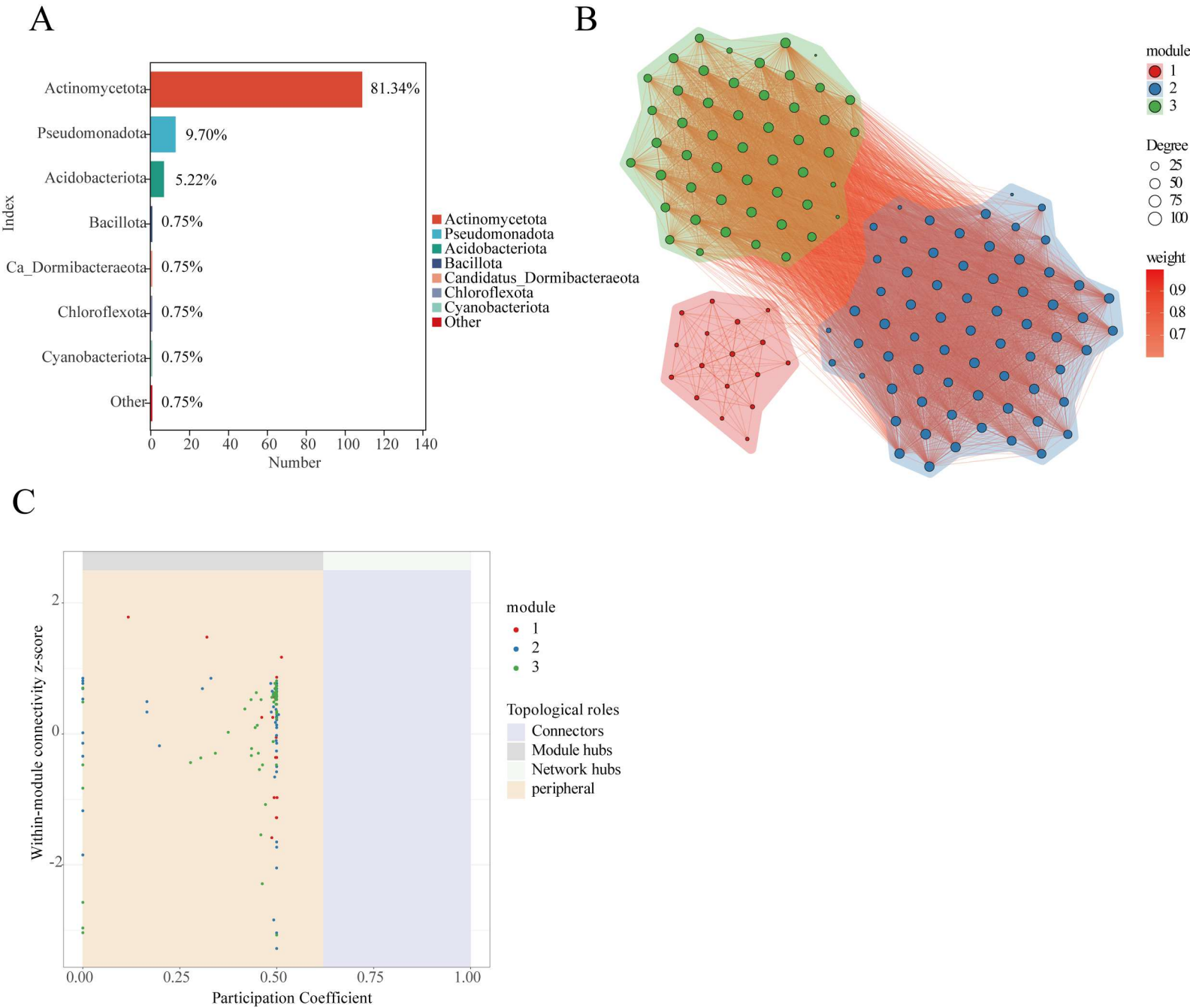

A
Bacteria
Gemmatimonadota
Other
Myxococcota
Planctomycetota
Nitrospirota
Chloroflexota
Ca. Rokubacteria
Acidobacteriota
Verrucomicrobiota
Actinomycetota
Pseudomonadota
Fungi
Cryptomycota
Blastocladiomycota
Zoopagomycota
Microsporidia
Chytridiomycota
Olpidiomycota
Basidiomycota
Ascomycota
Mucoromycota
Archaea
Other
Ca. Aenigmarchaeota
Ca. Micrarchaeota
Ca. Altiarchaeota
Ca. Heimdallarchaeota
Ca. Thermoplasmatota
Ca. Thorarchaeota
Thermoproteota
Ca. Bathyarchaeota
Euryarchaeota
Nitrososphaerota
Virus
Preplasmiviricota
Phixviricota
Other
Nucleocytoviricota
Kitrinoviricota
Cossaviricota
Artverviricota
Pisuviricota
Lenarviricota
Hofneiviricota
Uroviricota
Group
AA
BB
CC
DD
EE
FF
GG
HH
II
JJ
KK
LL
MM
NN
OO
PP
QQ
RR
SS
TT
UU
VV
WW
XX
YY
Tree
Stacked Barplot
Tree
Stacked Barplot
Tree
Stacked Barplot
Tree
Stacked Barplot

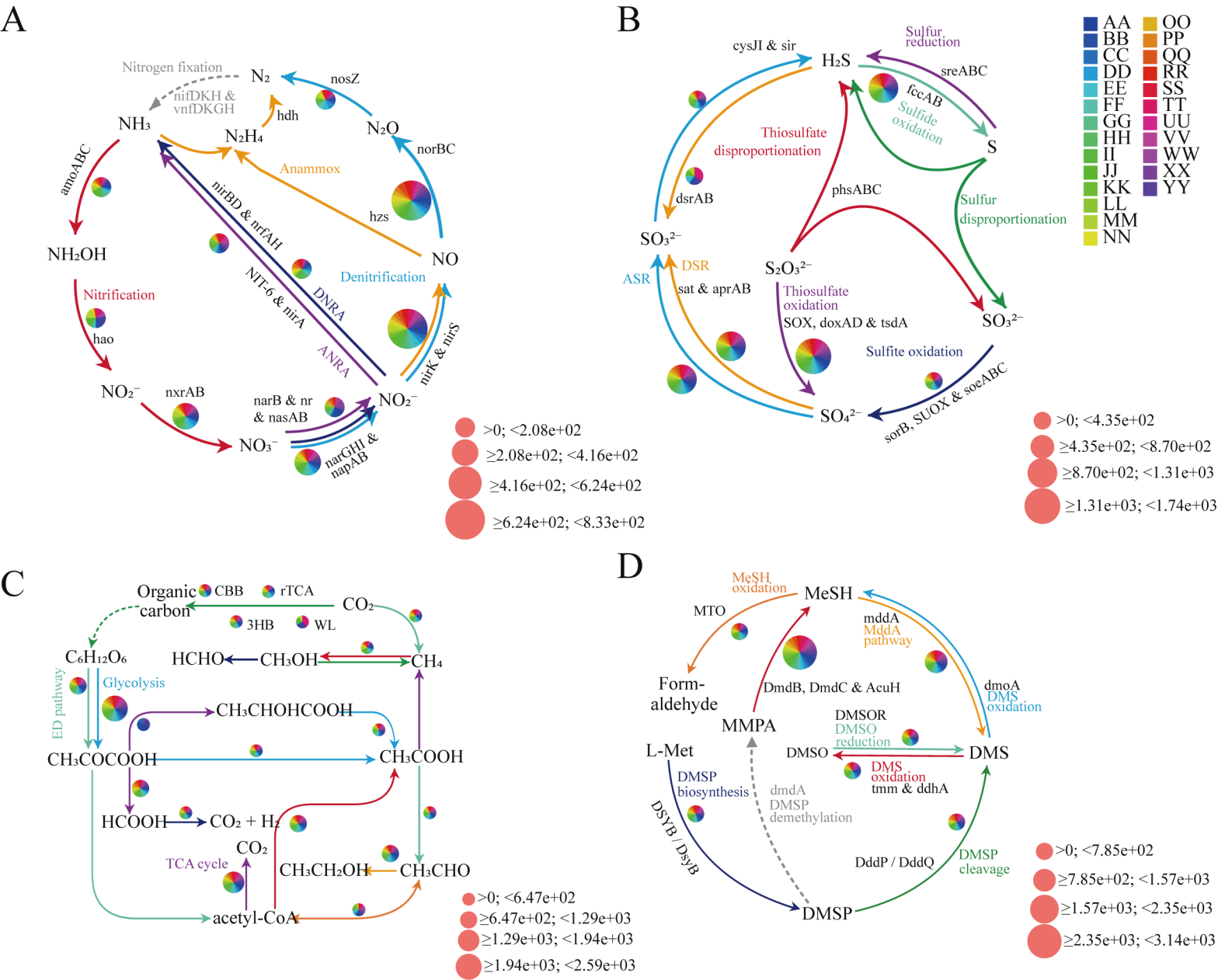

B
Mycolicibacterium
Burkholderia
Reyranella
Streptomyces
Ca. Udaeobacter
Trebonia
Pseudolabrys
Mycobacterium
Rhodoplanes
Bradyrhizobium
Fusarium
Verticillium
Apophysomyces
Colletotrichum
Beauveria
Rhizophagus
Aureobasidium
Termitomyces
Aspergillus
Rhizopus
Methanococcoides
Methanocella
Nitrosotalea
Halococcus
Halarchaeum
Methanothrix
Ca. Nitrosocosmicus
Halorubrum
Nitrososphaera
Haloferax
Triatovirus
Aparavirus
Immaneltrevirus
Woodruffvirus
Chlorovirus
Gammaretrovirus
Mitovirus
Lessievirus
Minunavirus
Iflavirus


A
Archaea
Site PERMANOVA: R² = 0.890, q = 0.0001
PCoA2 (15.91%)
PCoA1 (60.88%)
B
Bacteria
Site PERMANOVA: R² = 0.873, q = 0.0001
PCoA2 (25.05%)
PCoA1 (37.04%)
C
Fungi
Site PERMANOVA: R² = 0.838, q = 0.0001
PCoA2 (14.37%)
PCoA1 (57.57%)
D
Viruses
Site PERMANOVA: R² = 0.833, q = 0.0001
PCoA2 (9.56%)
PCoA1 (68.53%)
Sampling site
AA
BB
CC
DD
EE
FF
GG
HH
II
JJ
KK
LL
MM
NN
OO
PP
QQ
RR
SS
TT
UU
VV
WW
XX
YY


A
Observed richness
q = 0.0052
Between sites
q = 0.906
Within site
−1200
−800
−400
0
Taxa per pH unit
B
Module−trait associations across 25 sites
green
yellow
black
brown
purple
magenta
red
greenyellow
turquoise
blue
pink
0.183 (0.852)
0.247 (0.765)
0.018 (0.995)
0.228 (0.83)
-0.055 (0.995)
0.339 (0.491)
0.102 (0.924)
0.230 (0.83)
-0.213 (0.846)
0.099 (0.924)
-0.306 (0.575)
0.024 (0.995)
0.214 (0.846)
0.371 (0.461)
-0.559 (0.0537)
0.895 (1.36e-07)
0.097 (0.924)
-0.143 (0.865)
0.191 (0.852)
0.036 (0.995)
-0.066 (0.995)
-0.331 (0.491)
0.147 (0.865)
-0.204 (0.847)
-0.141 (0.865)
-0.129 (0.883)
-0.159 (0.865)
0.024 (0.995)
0.070 (0.995)
-0.182 (0.852)
-0.276 (0.711)
0.463 (0.191)
0.011 (0.995)
-0.075 (0.992)
-0.206 (0.847)
-0.006 (0.995)
-0.023 (0.995)
-0.323 (0.506)
0.181 (0.852)
-0.156 (0.865)
0.042 (0.995)
0.146 (0.865)
-0.350 (0.491)
0.162 (0.865)
0.001 (0.995)
-0.346 (0.491)
-0.023 (0.995)
0.197 (0.852)
0.061 (0.995)
0.005 (0.995)
-0.193 (0.852)
0.141 (0.865)
0.021 (0.995)
-0.678 (0.00415)
0.051 (0.995)
-0.096 (0.924)
-0.100 (0.924)
-0.174 (0.865)
-0.264 (0.711)
-0.128 (0.883)
-0.031 (0.995)
-0.270 (0.711)
0.141 (0.865)
-0.083 (0.968)
0.220 (0.846)
0.104 (0.924)
0.350 (0.491)
0.040 (0.995)
-0.154 (0.865)
0.046 (0.995)
0.490 (0.142)
-0.761 (0.000374)
0.137 (0.872)
-0.014 (0.995)
0.046 (0.995)
-0.004 (0.995)
-0.252 (0.761)
-0.457 (0.191)
0.414 (0.315)
-0.672 (0.00415)
0.265 (0.711)
-0.004 (0.995)
0.331 (0.491)
0.095 (0.924)
-0.114 (0.924)
-0.546 (0.0598)
0.398 (0.36)
-0.755 (0.000374)
LL
LW
PL
PH
LC
Longitude
Latitude
pH
1.0
0.5
0.0
−0.5
−1.0
Correlation coefficient
C
Turquoise module (1,975 taxa)
turquoise: r = 0.83, P < 10−200
|GS| for soil pH
|MM|
0.0
0.2
0.4
0.6
0.8
1.0
Pink module (152 taxa)
pink: r = 0.76, P = 7.2 × 10−30
|GS| for soil pH
|MM|


A
LW
LL
PL
PH
LC
Lon
Lat
pH
0.61
**
0.78
***
0.9
***
0.56
**
0.71
***
0.59
**
0.54
**
-0.47
*
-0.58
**
Archaea
Bacteria
Fungi
Viruses
Pearson correlation
1
0.2
-0.6
Mantel correlation
< 0.2
0.2 ~ 0.4
0.4 ~ 0.6
Mantel q-value
< 0.01
0.01 ~ 0.05
>= 0.05
B
pH
LC
LON
LW
PH
LL
PL
LAT
Ca. Rokubacteria
Verrucomicrobiota
Pseudomonadota
Actinomycetota
Acidobacteriota
RDA2 (7.229%)
RDA1 (35.83%)
400
200
0
-200
-400
Group
AA
BB
CC
DD
EE
FF
GG
HH
II
JJ
KK
LL
MM
NN
OO
PP
QQ
RR
SS
TT
UU
VV
WW
XX
YY


A
Actinomycetota
Pseudomonadota
Acidobacteriota
Bacillota
Ca_Dormibacteraeota
Chloroflexota
Cyanobacteriota
Other
81.34%
9.70%
5.22%
0.75%
0.75%
0.75%
0.75%
0.75%
Index
0
20
40
60
80
100
120
140
Number
Actinomycetota
Pseudomonadota
Acidobacteriota
Bacillota
Candidatus_Dormibacteraeota
Chloroflexota
Cyanobacteriota
Other
B
module
1
2
3
Degree
25
50
75
100
weight
0.9
0.8
0.7
C
Within-module connectivity z-score
2
0
-2
0.00
0.25
0.50
0.75
1.00
Participation Coefficient
module
1
2
3
Topological roles
Connectors
Module hubs
Network hubs
peripheral


A
Nitrogen fixation
N2
nifDKH & vnfDKGH
nosZ
NH3
hdh
N2O
N2H4
norBC
amoABC
Anammox
hzs
nirBD & nrfAH
NH2OH
NO
NIT-6 & nirA
Denitrification
DNRA
Nitrification
ANRA
hao
nirK & nirS
NO2−
nxrAB
narB & nr & nasAB
NO2−
NO3−
narGHJ & napAB
>0; <2.08e+02
≥2.08e+02; <4.16e+02
≥4.16e+02; <6.24e+02
≥6.24e+02; <8.33e+02
B
cysJI & sir
H2S
Sulfur reduction
sreABC
fccAB
Sulfide oxidation
S
Thiosulfate disproportionation
dsrAB
phsABC
Sulfur disproportionation
SO3 2−
DSR
ASR
sat & aprAB
S2O3 2−
Thiosulfate oxidation
SOX, doxAD & tsdA
SO3 2−
Sulfite oxidation
SO4 2−
sorB, SUOX & soeABC
AA BB CC DD EE FF GG HH II JJ KK LL MM NN
OO PP QQ RR SS TT UU VV WW XX YY
>0; <4.35e+02
≥4.35e+02; <8.70e+02
≥8.70e+02; <1.31e+03
≥1.31e+03; <1.74e+03
C
Organic carbon
CBB
rTCA
3HB
WL
CO2
C6H12O6
HCHO
CH3OH
CH4
ED pathway
Glycolysis
CH3CHOHCOOH
CH3COCOOH
CH3COOH
HCOOH
CO2 + H2
CO2
TCA cycle
CH3CH2OH
CH3CHO
acetyl-CoA
>0; <6.47e+02
≥6.47e+02; <1.29e+03
≥1.29e+03; <1.94e+03
≥1.94e+03; <2.59e+03
D
MeSH oxidation
MeSH
MTO
mddA MddA pathway
Form-aldehyde
DmdB, DmdC & AcuH
dmoA DMS oxidation
MMPA
DMSOR DMSO reduction
L-Met
DMSO
DMS
DMS oxidation
tmm & ddhA
DMSP biosynthesis
dmdA DMSP demethylation
DSYB / DsyB
DddP / DddQ
DMSP cleavage
DMSP
>0; <7.85e+02
≥7.85e+02; <1.57e+03
≥1.57e+03; <2.35e+03
≥2.35e+03; <3.14e+03

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.xlsx
- Additionalfile2.xlsx
- Additionalfile3.xlsx
- Additionalfile4.docx
- Additionalfile5.xlsx
- Additionalfile6.xlsx
- Additionalfile10.xlsx
- Additionalfile7.xlsx
- Additionalfile9.xlsx
- Additionalfile8.xlsx
- Additionalfile11.xlsx
- Additionalfile13.csv
- Additionalfile12.zip
- Additionalfile14.csv