Spatial Niche Differentiation and Growth-Promoting Functions of Bacterial Microbiota in Smilax china L

Qian Wang
[illegible]

Wenjuan Wang
[illegible]

Ziyu Wei
[illegible]

Heming Wu
[illegible]

Xiaofang Song
[illegible]/[illegible]

Xuan Zhang
[illegible]

Gaoling Wu
[illegible]/[illegible]

Zhiyong Wang
wzhy660305@sina.com

[illegible]/[illegible]

Research Article

Keywords: Bacterial community, Smilax china L., Horizontal distribution

Posted Date: June 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9562315/v1

Additional Declarations: No competing interests reported.

Spatial Niche Differentiation and Growth-Promoting Functions of Bacterial Microbiota in *Smilax china* L.

Qian Wang[1,2], Wenjuan Wang[3], Ziyu Wei[5], Heming Wu[4], Xiaofang Song[2], Xuan Zhang[5], Gaoling Wu[2*], Zhiyong Wang[2*]

(1. Yangtze University, Jingzhou 434023, Hubei, China; 2. Xianning Polytechnic/ Hubei Province South Characteristic Agricultural Industry Technology Research Institute, Xianning 437100, Hubei, China; 3. Hubei University of Science and Technology, Xianning 437100, Hubei, China; 4. Hubei Furen Pharmaceutical Co., Ltd., Xianning 437100, Hubei, China; 5. Xianning Public Inspection Center, Xianning 437100, Hubei, China)

Abstract

An extremely crucial role is exerted by bacteria in plant growth and development. To investigate the specific distribution patterns of bacteria from control check soil (CK) to rhizosphere soil (RS), plant roots (R), stems (T) and leaves (L), as well as their effects on plant growth and development, *Smilax china* L. with typical rhizomes was selected as the research material in this study. Samples of CK, RS, R, T and L were collected from 10 sampling sites of plants that had been transplanted and established for 3 years. On the basis of determining the conventional physicochemical properties and culturable microbial contents of the soil, the V5-V7 regions of the bacterial 16S rRNA gene were amplified from the total DNA of each sample, followed by high-throughput sequencing. Subsequently, bioinformatics analysis was conducted using QIIME2. **Results**: Significant differences were observed in the changes of soil physicochemical properties and culturable microorganisms induced by the growth of *S. china* L. The highest species richness and α-diversity level were detected in RS, whereas more dominant taxa were identified in R. The community composition was stable in each spatial niche, and the taxa with high abundance in T and L were basically consistent with the dominant taxa. The top 10 genera mainly played the role of "connectors" in the entire network system, and no obvious network core taxa or peripheral taxa were

Received: 2026- -.

Funding: Central Government Guided Local Science and Technology Development Fund Project (No. Z135050009017-429); School-level Research Project of Xianning Vocational and Technical College (No. 2025005).

Author Information: Qian Wang (1994–), male, Bachelor of Science, Lecturer at Xianning Polytechnic Research direction: applied biotechnology. E-mail: 91970025@qq.com; Wenjuan Wang (1985–), female, Master of Medicine, Lecturer at Hubei University of Science and Technology. Research direction: integrated traditional Chinese and Western medicine. E-mail: wenjuanwang@hbust.edu.cn.

*These authors contributed equally to this work.

Corresponding author: Zhiyong Wang, male, Professor, Master degree. Research direction: applied biotechnology. E-mail: wzhy660305@sina.com; **Co-corresponding authors**: Gaoling Wu, male, Professor, Doctoral degree. Research direction: Chinese medicinal materials. E-mail: wugaoling888@163.com.

found in the system. The endophytes in the above-ground parts of *S. china* L. supplemented nutrients for the plant through metabolism. The endophytes in rhizomes degraded harmful substances via metabolic processes and produced certain plant growth regulators to promote cell division. **Conclusion**: The growth of *S. china* L. causes changes in soil physicochemical properties, which in turn induce alterations in the spatial distribution of bacterial communities. These changes enable endophytic bacteria in the plant to supplement nutrients for different plant parts and degrade harmful components through metabolism, thereby enhancing the plant's competitiveness in diverse environments. In addition, some endophytes produce growth regulators to promote cell division, which may be one of the reasons for rhizome enlargement and the improvement of its commercial value.

Key points

Bacterial α-diversity peaks in rhizosphere soil and decreases toward aboveground tissues; roots act as a transitional hub.

Dominant genera show strong tissue specificity: methylotrophs dominate shoots, while Burkholderia-complex and rhizobia enrich roots.

Bacteria exhibit niche-adapted functions; root isoprene and cell-division pathways may promote rhizome enlargement of Smilax china L.

Key words: Bacterial community; *Smilax china* L.; Horizontal distribution

Introduction

Plant-associated bacterial communities form a continuous ecological system across bulk soil, Plant-associated bacteria function as an integral “second genome” that governs plant nutrient acquisition, stress resilience, and developmental processes, and they exhibit strong niche differentiation across the soil–plant continuum (Berendsen et al., 2012; Müller et al., 2016). The rhizosphere typically supports the highest microbial diversity, whereas endophytic communities are tightly regulated by host filtering and display distinct tissue-specific functional zonation (Reinhold-Hurek et al., 2015; Xiong et al., 2021). In model and crop plants, bacterial assembly is modulated by root exudates, tissue microenvironments, and host traits, with aboveground bacteria favoring photosynthetic efficiency and root-associated taxa driving nutrient mineralization and pathogen suppression (Bulgarelli et al., 2013; Madhaiyan et al., 2015). However, such comprehensive analyses remain limited for perennial medicinal plants.

Smilax china L. is a medicinally important perennial liana widely distributed in southern China. Its enlarged rhizome is officially recognized in the Chinese Pharmacopoeia for anti-inflammatory and immunomodulatory activities (Chinese Pharmacopoeia Commission, 2020). As a flagship authentic medicinal crop in Hubei Province, large-scale cultivation is constrained by long growth cycles, inconsistent yields, and variable active component contents. Current research has focused on phytochemistry and pharmacology, with microbiological studies restricted to rhizome endophytes

(Li et al., 2019; Zhao et al., 2022). The spatial organization, assembly drivers, and functional roles of bacterial communities across bulk soil, rhizosphere, root, stem, and leaf remain largely uncharacterized.

To fill this knowledge gap, we examined bacterial communities across the soil–plant continuum of 3-year-old cultivated *S. china* using 16S rRNA high-throughput sequencing, soil physicochemical assays, and microbial quantification. This study aimed to define spatial niche partitioning, identify key functional taxa, and clarify microbial mechanisms linked to rhizome development. Our results establish a microecological framework for optimizing cultivation practices and improving yield and quality of this valuable medicinal plant through beneficial microbial interventions.

Materials and Methods

Materials

Sampling Time and Location

The sampling was performed on July 21, 2024, at the authentic medicinal plant cultivation base managed by Hubei Furen Pharmaceutical Co., Ltd., located at 29°16′50″ N, 113°54′25″ E in Xianning, Hubei Province, China.

Overview of the Sampling Site

The experimental site has a subtropical humid monsoon climate with four distinct seasons, large diurnal temperature variation, and strong summer winds. The annual average temperature is 11.7 °C, annual precipitation is approximately 1681 mm, and relative humidity is high. The plantation was established with 3-year-old artificially transplanted *S. china* L. and covers an area of approximately 13.34 ha. The soil type is typical acidic red soil locally used for medicinal plant cultivation.

Sample Collection and Processing

Ten sampling points were randomly selected along an "S"-shaped route in the north-south direction. Samples of leaves (L), stems (T), underground tubers (R), rhizosphere soil (RS), and bulk soil (CK, approximately 30 cm away from the plant roots) were collected separately at each point and numbered sequentially according to the sampling order (e.g., CK1–CK10). After soil sampling, the sampling pits were immediately backfilled with surrounding soil. The collected soil samples were thoroughly mixed, air-dried to a humidity of 20%–30% promptly, and sieved through a 2 mm sieve. Subsequent determination and analysis were carried out as soon as possible (Zhang, 2007).

Ten representative sampling points were selected along an S-shaped transect to ensure uniform spatial distribution. At each point, five compartment samples were collected: Bulk soil (CK): soil collected 30 cm away from plant roots at 0–20 cm depth; Rhizosphere soil (RS): soil tightly adhering to root surfaces after gentle shaking; Root (R): healthy fibrous and main roots; Stem (T): middle and upper healthy stems; Leaf (L): fully expanded mature leaves. Plant tissue samples were transported with dry ice to maintain sample integrity. Soil samples were homogenized, passed

through a 2 mm sieve, and partially air-dried to 20%–30% moisture content for physicochemical analysis. All samples for microbial analysis were stored at −80 °C until DNA extraction.

Reagents

MagBeads FastDNA Kit for Soil (116564384) (MP Biomedicals); Q5® High-Fidelity DNAPolymerase（M0491L）（NEB）; Quant-iT PicoGreen dsDNAAssay Kit（P7589）（Invitrogen）; other conventional reagents.

Instruments

Multi-sample Tissue Grinder (Model Tissuelyser-48, Jingxin Industrial Development Co., Ltd., Shanghai, China); Gel Electrophoresis Apparatus (Model DYY-6C, Liuyi Biotechnology Co., Ltd., Beijing, China); Nanodrop 2000 Ultra-micro Spectrophotometer; PCR Instrument (Model ANALYTIKJENA qTOWER 2.2, Analytik Jena AG, Germany).

Analytical Methods

Analysis of Soil Physicochemical Properties and Determination of Culturable Microbial Counts in Soil

The soil pH value is determined by the electrode method; soil organic matter (OM) is measured using the potassium dichromate external heating method; total nitrogen (TN) in soil is analyzed via the Kjeldahl method; nitrate nitrogen (NO_3-N) is quantified by ultraviolet spectrophotometry; available phosphorus (A-P) is determined through the ammonium fluoride-hydrochloric acid extraction-molybdenum antimony anti-colorimetric method; available potassium (A-K) is measured using the ammonium acetate extraction-atomic absorption spectrophotometry; soil water content is determined by conventional oven-drying weighing and air-drying weighing methods (Zhang, 2007).

For the mixed soil samples from each sampling point, referring to *Microbiology Experiments* (Zhao, 2007), the counts of culturable bacteria, fungi, and actinomycetes in soil are determined by the dilution plate counting method using 1/10 TSB medium, modified Martin medium, and Gao's No. 1 medium, respectively.

Taking the content of each index in the samples as raw data, GenesCloud (https://www.genescloud.cn) is used to analyze the significance of differences in the content of each component and the correlation of data. In the significance analysis of differences, the Kruskal-Wallis test is performed for multiple group comparisons, and the Spearman rank correlation coefficient, a non-parametric statistical method suitable for non-normally distributed data, is used for correlation analysis between groups. A P value < 0.05 was considered statistically significant.

16S rRNA Gene High-Throughput Sequencing

Total DNA Extraction

Extraction of Total Soil DNA

Samples are taken out of the refrigerator, and an appropriate amount (0.2-0.5 g) is quickly weighed and added to a centrifuge tube containing extraction lysis buffer, followed by grinding with a tissue grinder at a frequency of 60 Hz. For the preprocessed samples, nucleic acids are extracted using the OMEGA Soil DNA Kit.

Extraction of Total DNA from Endophytic Bacteria

Surface-sterilized samples were subjected to total DNA extraction using a modified CTAB method. Briefly, approximately 500 mg of sample is added to 500 μL of 2×CTAB; a small amount of glass beads is added, and the mixture is homogenized at 25 Hz for 7 min using a tissue disruptor; 50 μL of lysozyme and 5 μL of RNase are added, followed by incubation at 37℃ for 20 min; 16 μL of proteinase K is added, and the mixture is gently inverted to mix (do not vortex vigorously!) and incubated at 55℃ for 1 h; 25 μL of 20% SDS is added, and the mixture is gently inverted to mix (do not vortex vigorously!) and incubated at 55℃ for 1 h; an equal volume of phenol: chloroform: isoamyl alcohol (25:24:1) is added, and the mixture is inverted to mix thoroughly to denature and precipitate proteins as much as possible, followed by centrifugation at 12000 rpm for 10 min; the supernatant is carefully transferred to a new 2 mL centrifuge tube without disturbing the middle white precipitate, 2 μL of RNase is added, and the mixture is allowed to stand at room temperature for 5 min; an equal volume of chloroform: isoamyl alcohol (24:1) is added, the mixture is inverted to mix (do not vortex vigorously), centrifuged at 12000 rpm for 10 min, and the supernatant is transferred to a new 1.5 mL centrifuge tube; 2/3 volume of isopropanol, 3 μL of glycogen (20 mg/mL), and 1/10 volume of 3 M sodium acetate are added, mixed well (white flocculent precipitate should be observed), and incubated at -20℃ for more than 30 min, followed by centrifugation at 12000 rpm for 10 min; the supernatant is discarded, the pellet is washed with freshly prepared 75% ethanol, centrifuged at 12000 rpm for 5-10 min, and allowed to stand at room temperature for 5 min until ethanol is completely volatilized; 50-100 μL of DNase-free water is added to dissolve the DNA, gently pipetted to mix, and allowed to stand at room temperature for 2-5 min until the DNA is completely dissolved. The molecular size of the extracted DNA is determined by 0.8% agarose gel electrophoresis, and the DNA is quantified using a Nanodrop spectrophotometer.

PCR Amplification and Library Preparation

The V5-V7 hypervariable regions of the bacterial 16S rRNA gene are amplified using the forward primer 799F (5′-AACMGGATTAGATACCCKG-3′) and the reverse primer 1193R (5′-ACGTCATCCCCACCTTCC-3′). This primer pair can effectively reduce the co-amplification level of mitochondrial 16S rRNA genes (Wang et al., 2018; Horton et al., 2014). The barcode in the forward primer is a 7-10 base oligonucleotide sequence used to distinguish different samples in the same library.

The PCR reaction system consists of 5 μL of 5× Reaction Buffer, 5 μL of 5× High GC Buffer, 0.25

μL of Q5 high-fidelity DNA polymerase, 2 μL of 10 mM deoxyribonucleoside triphosphates (dNTPs), 1 μL of 10 μM forward primer, 1 μL of 10 μM reverse primer, 2 μL of DNA template, and 8.75 μL of double-distilled water. The amplification program is set as follows: pre-denaturation at 98℃ for 5 minutes; 25 cycles of denaturation at 98℃ for 30 seconds, annealing at 53℃ for 30 seconds, and extension at 72℃ for 45 seconds; final extension at 72℃ for 5 minutes; and preservation at 12℃. The amplification products are subjected to 2% agarose gel electrophoresis, and the target fragments are excised and recovered using the Axygen Gel Extraction Kit.

High-Throughput Sequencing

The PCR amplification products are purified using Vazyme VAHTSTM DNA Clean Beads (Vazyme Biotech Co., Ltd., Nanjing, China) and quantified for nucleic acid concentration using the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen, Carlsbad, CA, USA). After quantifying individual samples, the amplicons are mixed in equimolar amounts and entrusted to Personalbio Technology Co., Ltd. (Shanghai, China) for paired-end 250 bp sequencing (2×250 bp) on the Illumina NovaSeq platform using the NovaSeq 6000 SP Sequencing Kit (500 cycles).

Bioinformatics Analysis

Microbiome bioinformatics analysis is performed using QIIME2 version 2019.4 (Bolyen et al., 2018) with minor adjustments according to the official tutorials (https://docs.qiime2.org/2019.4/tutorials/).

Raw Data Processing

The raw sequence data are demultiplexed using the demux plugin, followed by primer trimming using the cutadapt plugin (Martin, M., 2011). Subsequently, the sequences are subjected to quality filtering, denoising, merging, and chimera removal using the DADA2 plugin (Callahan et al., 2016). Non-singleton amplicon sequence variants (ASVs) are aligned using mafft (Katoh et al., 2002), and a phylogenetic tree is constructed using fasttree2 (Price et al., 2010).

Analysis of Species Composition, α-Diversity, and β-Diversity

Taxonomic annotation of ASVs (Amplicon Sequence Variants) is performed using the classify-sklearn naive Bayes classifier in the feature-classifier plugin (Bokulich et al., 2018) based on the SILVA Release 132 bacterial database (Koljalg et al., 2013). The diversity plugin is used to estimate the number of observed species, α-diversity indices (including Chao1 (Chao, 1984), Shannon index (Shannon, 1948a, b), and Good's coverage (Good, 1953)), and β-diversity indices (including weighted UniFrac (Lozupone et al., 2007), unweighted UniFrac (Lozupone et al., 2005), Jaccard distance, and Bray-Curtis dissimilarity). Each sample is rarefied to 34,298 sequences. Using the rarefied ASV/OTU table, the "qiime diversity core-metrics-phylogenetic" or "qiime diversity core-metrics" command is called to calculate four distance matrices (Jaccard, Bray-Curtis, unweighted UniFrac, and weighted UniFrac) or two distance matrices (Jaccard and Bray-Curtis), and Principal Coordinates Analysis (PCoA) is performed on these distance matrices (Ramette, 2007). When

constructing the species composition heatmap, samples are clustered by UPGMA (default clustering algorithm) based on the Euclidean distance of species composition data and arranged according to the clustering results by default; otherwise, they are arranged in the default order of samples/groups. Species are clustered by default, i.e., UPGMA clustering (default clustering algorithm) based on the Pearson correlation coefficient matrix of their composition data, and arranged according to the clustering results; otherwise, they are sorted by the average abundance of species in samples/groups. The Bray-Curtis distance algorithm is used, with roots (R) as the control, and the adonis test is performed to analyze inter-group differences (Clarke, 1993; Warton et al., 2012).

Analysis of Species Differences, Biomarker Species, and Correlation Networks

QIIME2 is used to evaluate the significance of differences in microbial community structure among groups through Permutational Multivariate Analysis of Variance (PERMANOVA) (McArdle and Anderson, 2001), Analysis of Similarities (ANOSIM) (Clarke, 1993; Warton et al., 2012), and Permutation Dispersion Test (Permdisp) (Anderson et al., 2006). Based on the occurrence of ASVs among samples/groups (ignoring relative abundance), the R package "VennDiagram" is used to draw Venn diagrams to show the shared and unique ASVs among samples/groups (Zaura et al., 2009). Random forest analysis (Breiman, 2001; Liaw and Wiener, 2002) is applied to distinguish samples from different groups under the default settings of QIIME2, where nested stratified k-fold cross-validation is used for automatic hyperparameter optimization and sample prediction (k = 5). Co-occurrence network analysis is performed via SparCC analysis. The pseudocount value in SparCC is set to 10^{-6}. The threshold of the correlation coefficient is determined to be 0.7 using a method based on random matrix theory (implemented in the R package RMThreshold). A co-occurrence network is constructed based on the correlation coefficients, where nodes represent ASVs and edges represent the correlations between these ASVs. Using R, according to the modular partitioning results of the current co-occurrence network, the Zi (within-module connectivity) and Pi (among-module connectivity) scores of each node in the network are calculated. The role of each node in the correlation network is then determined based on the Zi and Pi scores (Deng et al., 2012).

Analysis of Metabolic Differences

With reference to known microbial genome data, the 16S rRNA gene sequences are predicted in multiple functional databases (such as MetaCyc, KEGG, COG, etc.). PICRUSt2 software is used to predict the abundances of KEGG orthologs, EC enzyme classification numbers, and other related indices based on the abundance of marker gene sequences in the samples (Gavin M. Douglas, et al., preprint). The abundance values of metabolic pathways are obtained based on metabolic pathway databases and certain calculation methods. After obtaining the abundance data of metabolic pathways, the fitFeatureModel function is called to fit the distribution of each pathway/group using a zero-inflated log-normal model, and the significance of differences is determined using the fitting results of this model.

Results and Analysis

Soil Physicochemical Properties and Culturable Microbial Data

The main soil physicochemical properties and culturable microbial contents are shown in Table 1.

Raw data indicated that after the growth of *S. china* L., the soil available phosphorus (A-P) content and culturable bacterial count increased significantly, and the pH value increased obviously. However, the contents of soil organic matter (OM), total nitrogen (TN), and nitrate nitrogen (NO_3-N), as well as the counts of culturable fungi and actinomycetes, showed no obvious regular changes.

The tissues of *S. china* L. mainly contain compounds such as steroidal saponins and flavonoids, among which steroidal saponins—accounting for a relatively large proportion—are neutral. Deng Yuxing et al. predicted a total of 26 flavonoid biosynthetic enzyme gene models from the current *Salvia miltiorrhiza* genome assembly using systematic computational methods (Deng et al., 2018); the isoelectric points (pI) of the expression products of these genes range from 4.9 to 8.94, and a similar pattern is observed in *S. china* L. These substances are also secreted into the soil, thereby increasing the pH value of the originally acidic soil to a certain extent. This increase is beneficial to the growth of most bacteria, and generally, within the neutral pH range, the content of soil available phosphorus increases with the rise of pH value.

Table 1 Main Physicochemical Properties of Soil and Quantity of Culturable Microorganisms

Sample	OM, mg/kg	TN, g/kg	NO^{3-}, mg/kg	pH	AP-K, mg/kg	AP-P, mg/kg	Bac, 10^5cfu/g	Fungi, 10^4cfu/g	Act, 10^4cfu/g
CK-1	11.23±0.15	0.81±0.01	1.91±0.03	5.40±0.01	125.05±1.17	107.73±2.45	42.6±1.15	12.33±2.08	11.33±0.58
RS-1	10.77±0.12	0.97±0.02	1.47±0.03	5.53±0.01	74.21±2.07	248.58±2.45	67.67±3.79	6.67±1.15	14.67±2.31
CK-2	12.47±0.21	0.83±0.01	6.82±0.23	4.97±0.01	86.18±0.43	198.92±3.04	41.33±1.15	11.00±1.00	17.00±2.65
RS-2	10.03±0.06	0.77±0.02	3.21±0.20	5.33±0.01	78.59±0.61	263.84±1.51	50.00±2.00	12.00±2.00	10.00±2.00
CK-3	11.17±0.15	0.75±0.02	2.33±0.10	5.13±0.01	73.24±1.38	103.57±0.70	48.00±5.29	14.33±4.04	15.00±1.00
RS-3	10.07±0.14	0.71±0.02	3.58±0.19	5.37±0.01	92.96±1.71	109.11±1.01	62.67±2.52	9.00±3.46	9.67±2.52
CK-4	9.18±0.14	0.68±0.02	2.66±0.16	4.93±0.01	29.12±0.21	55.14±0.94	48.00±3.00	8.00±2.00	5.00±1.00
RS-4	6.17±0.06	0.34±0.01	1.02±0.16	5.23±0.01	27.55±0.33	75.68±1.56	60.67±1.15	9.00±2.65	5.00±0.00
CK-5	13.37±0.25	0.80±0.03	3.70±0.14	5.07±0.01	110.83±0.91	116.89±1.21	43.33±2.89	13.00±1.00	14.00±2.65
RS-5	10.43±0.15	0.72±0.01	2.69±0.13	5.43±0.01	81.48±2.00	172.73±1.37	57.67±6.81	8.00±3.46	6.33±1.53
CK-6	10.27±0.15	0.74±0.01	2.66±0.32	5.13±0.01	61.17±0.30	164.70±2.16	26.67±5.77	17.00±2.00	11.00±1.00
RS-6	9.83±0.06	0.65±0.01	5.72±0.15	5.37±0.01	77.08±0.40	198.99±1.92	31.33±1.15	3.00±1.00	3.33±0.58

CK-7	9.97±0.04	0.76±0.01	3.80±0.04	4.80±0.01	71.49±1.57	215.85±1.44	63.00±7.81	18.33±1.53	5.67±0.58
RS-7	9.80±0.04	0.67±0.01	1.14±0.07	5.17±0.01	85.21±1.33	264.75±1.40	101.67±7.64	22.00±6.00	6.67±0.58
CK-8	13.50±0.17	0.76±0.02	1.63±0.07	4.97±0.01	119.12±0.49	135.26±2.55	60.33±0.58	11.33±1.53	13.67±1.53
RS-8	11.07±0.15	0.65±0.03	2.86±0.12	5.27±0.01	90.43±3.94	165.67±1.93	65.67±4.04	14.00±1.73	12.67±1.53
CK-9	9.30±0.06	0.67±0.03	1.26±0.05	5.03±0.01	25.03±0.19	65.12±2.61	61.67±1.53	8.33±0.58	18.33±1.53
RS-9	7.85±0.06	0.44±0.02	1.08±0.09	5.33±0.01	35.26±0.74	99.98±2.13	76.67±4.04	6.67±0.58	12.67±1.53
CK-10	10.47±0.15	0.68±0.01	0.85±0.07	5.27±0.01	38.34±0.55	128.71±3.02	67.00±1.73	17.00±3.00	11.67±0.58
RS-10	4.86±0.06	0.63±0.01	1.10±0.10	5.53±0.01	28.06±0.74	175.89±0.79	75.00±1.00	12.67±1.53	8.67±1.53

The significance analysis of differences in routine detection data (Fig. 1 A, B) revealed that the differences in all detected indicators of soil physicochemical properties and culturable microorganisms between the CK and RS groups were less than 0.01 ($p < 0.01$), indicating extremely significant differences in all indicators between the two groups overall. Therefore, *S. china* L. exerted significant effects on soil physicochemical properties and culturable microorganisms during its growth.

The results of correlation analysis of raw routine detection data (Fig. 1 C) showed that soil organic matter (OM) was positively correlated with total nitrogen (TN), nitrate nitrogen ($NO^{3-}N$), and actinomycetes, but negatively correlated with bacteria; available potassium (A-K) was positively correlated with OM, TN, $NO^{3-}N$, and available phosphorus (A-P); TN was also positively correlated with $NO^{3-}N$, A-P, and actinomycetes, but negatively correlated with bacteria; pH value was only negatively correlated with fungi; bacteria were negatively correlated with OM, TN, and $NO^{3-}N$, while fungi were only negatively correlated with pH.

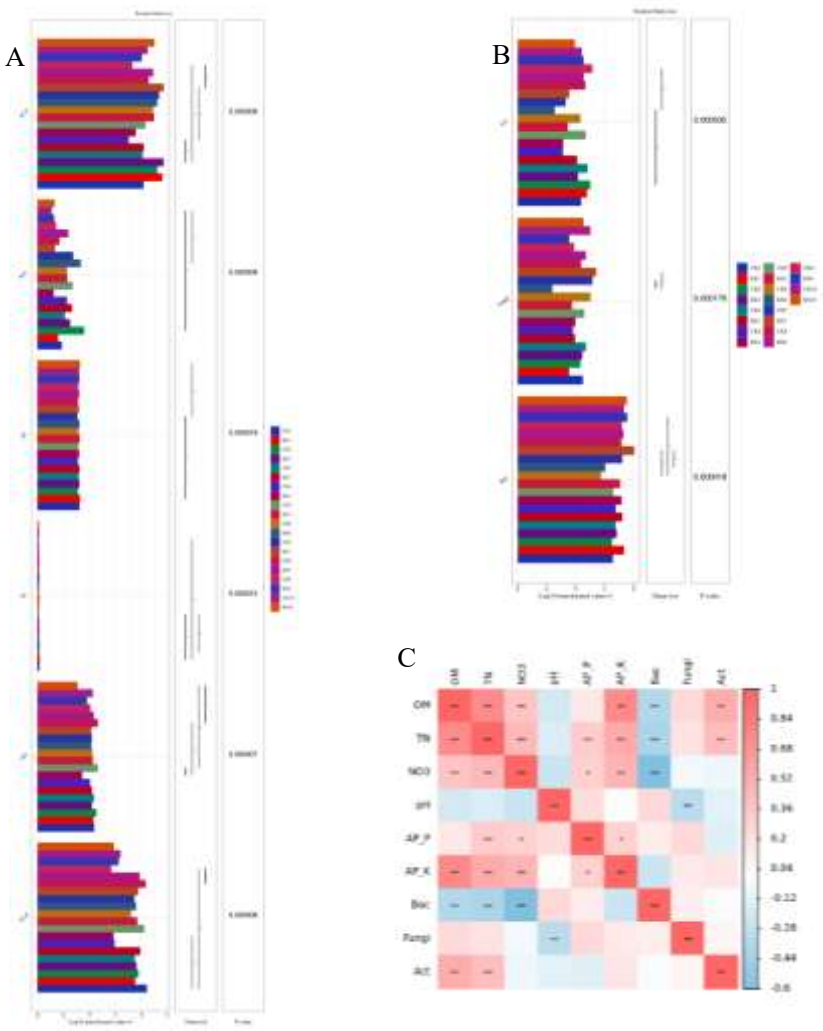

Fig. 1 Significance and Correlation Analysis of Soil Physicochemical Properties and Culturable Microorganisms

A. The significance analysis of differences in conventional physicochemical properties. Each column of data corresponds to a box plot subgraph, which displays the P-value of the overall inter-group difference obtained by the Kruskal-Wallis non-parametric test, as well as the markers of significance levels for pairwise post-hoc comparisons between groups derived from Dunn's test (*, P<0.05; **, P<0.01; ***, P<0.001).

B. The significance analysis of cultivable microbial differences (same as above).

C. The correlation analysis between physicochemical properties and cultivable microorganisms. The legend displays correlation coefficient values. Red indicates positive correlation, and blue indicates negative correlation. The shade of the color represents the strength of the correlation. Asterisks mark significant correlations (P < 0.05): * stands for p<0.05, ** for p<0.01, and *** for p<0.001.

Species Composition Analysis

Statistics of Taxonomic Units

After statistics of the feature table with singletons removed, the distribution of the number of

different bacterial taxonomic units in each sample-following averaging of samples within groups—is shown in Table 2. At the phylum, class, order, family, and genus levels, RS exhibited the richest taxonomic units, followed by CK, with R, T, and L decreasing sequentially. At the phylum level, the top 5 most abundant phyla included *Proteobacteria*, *Actinobacteriota*, *Acidobacteriota*, *Gemmatimonadota*, *Myxococcota*, and *Firmicutes*.

Table 2 Number Distribution of Bacterial Taxa at Different Taxonomic Levels in Each Sample

ID	Domain	phylum	class	order	Family	genus	Species
CK	1	17.9	34.6	78	106.2	142.4	44.9
RS	1	19.8	35.9	82.5	108.8	142.5	42.5
R	1	11.5	19.7	50.3	65.8	89	34.7
T	1	9.1	14.3	33.3	43	60.1	43.4
L	1	7.1	10.3	21.4	26.6	31.8	28.3

Species Composition Analysis

Excluding unidentified species, the top 20 most abundant species in the species composition heatmap included 14 distinct species (Fig. 2). Among these, 7 species belonged to the genus *Methylobacterium* (family *Methylobacteriaceae*), 5 species to the family *Rhizobiaceae* (2 species of *Aureimonas*, 1 species of *Rhizobium*, 1 species of *Mesorhizobium*, and 1 species of *Rhizobiales*), and 2 species of *Sphingomonas* (family *Sphingomonadaceae*).

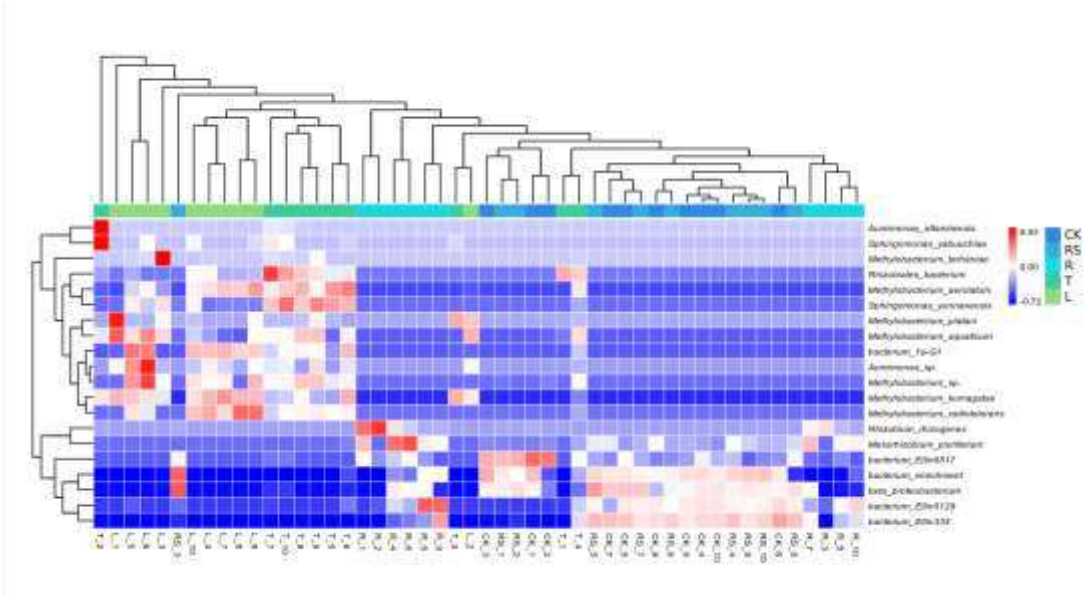

Fig. 2 Species Composition Heatmap

The horizontal axis of the graph represents each sample number, and the vertical axis indicates the relative abundance

of taxonomic units in each sample/group under the grouping scheme.

The seven species: *M. komagatae*, *M. radiotolerans*, *M. sp.*, *M. tarhaniae*, *M. platani*, *M. aerolatum*, and *M. aquaticum*—exhibited high abundances in both L (leaves) and T (stems). Studies have shown that phosphoribosyl pyrophosphate synthetase (PRS) from *Methylobacterium* regulates carbon flux redistribution mediated by the polyketide thiolase (PKT) pathway through its "metabolic valve" function, enabling efficient growth and phyllosphere colonization under low methanol conditions. These core phyllosphere microbial communities are characterized by methylotrophic metabolism, participate in biological nitrogen fixation, expand plant nutrient supply pathways, and enhance nutrient availability. They improve plant stress resistance by regulating host physiological processes (e.g., activating antioxidant systems) and can synergize with plants for pollution remediation, thereby promoting plant growth and yield improvement (Zhang et al., 2024).

The two species of the genus *Aureimonas* also showed high abundances in T and L; *Rhizobiales_bacterium* was highly abundant in T; *Rhizobium rhizogenes* was highly abundant in R (roots); and *Mesorhizobium plurifarium* was highly abundant in both R and RS (rhizosphere soil). As a *Liliales* plant, *S. china* L. lacks rhizobia-related receptors and downstream signaling molecules, and thus cannot respond to symbiotic signals from rhizobia to form nodule organs. However, rhizobia can survive independently of symbiosis, utilizing root exudates (e.g., carbohydrates, organic acids) as carbon sources for saprophytic growth on the root surface or in the intercellular spaces of root cortical cells. This relationship is beneficial to the growth of *S. china* L. The rhizobia accumulated in the roots of *S. china* L. may belong to such non-symbiotic strains. Some rhizobia can secrete effector proteins to transiently inhibit plant immune responses (e.g., suppressing reactive oxygen species (ROS) burst), thereby surviving locally in the roots without forming typical symbiotic structures (Wang et al., 2025). The endophytic microbial community in the roots of *S. china* L. may establish low-level colonization through this mechanism.

Sphingomonas yunnanensis and *Sphingomonas yabuuchiae* (family *Sphingomonadaceae*) exhibited high abundances in T and L. Bacteria of this genus exert positive effects on plants, including pollutant degradation, promotion of plant growth, inhibition of pathogenic bacteria, and biological control. These two major groups of bacteria synergistically form a stable microbial network, which can inhibit pathogen colonization, produce plant hormones (e.g., auxin) or phosphate-solubilizing substances, and directly stimulate plant growth and nutrient absorption (Cai et al., 2020).

Excluding unidentified species, the top 20 most abundant bacteria enhanced the competitive ability of *S. china* L. in different growth environments, promoted plant growth, and were conducive to improving its yield.

Alpha Diversity Analysis

Alpha diversity refers to indices reflecting species richness, diversity, and evenness in locally homogeneous habitats, and is also known as within-habitat diversity. Data from the three alpha

diversity indices showed no significant differences in bacterial community diversity between CK and RS, or among R, T, and L (Fig. 3).

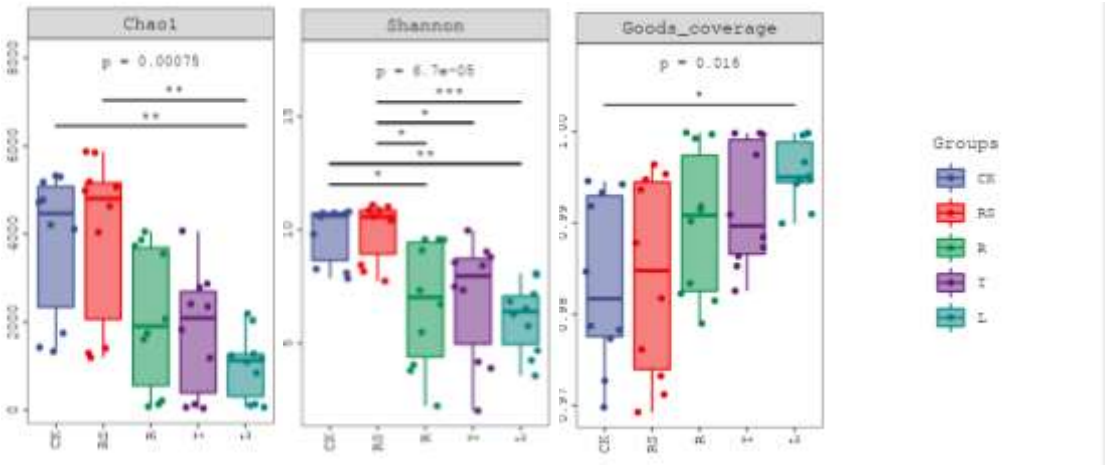

Figure 3 Alpha diversity of samples in each group

The abscissa represents group labels, and the ordinate represents the values of corresponding alpha diversity indices. The numbers under the diversity index labels are the p-values from the Kruskal-Wallis test.

The Chao1 index, representing species richness, was significantly higher in the RS group than in plant tissue groups ($p = 0.00075$). Pairwise comparisons further revealed that both CK and RS differed significantly from L ($p = 0.0044$ and $p = 0.005$, respectively), indicating remarkably higher bacterial richness in soil compartments than in aboveground tissues.

The Shannon index, which integrates species richness and abundance, also peaked in RS and showed an overall decreasing trend from soil to aboveground tissues ($p = 6.7 \times 10^{-5}$). CK and RS differed significantly from R, T, and L ($p < 0.05$), confirming strong host filtering effects that reduced bacterial diversity from the rhizosphere to internal plant tissues.

Good's coverage values exceeded 0.99 in all samples, indicating sufficient sequencing depth to capture the majority of bacterial taxa. Multiple-group comparison showed significant differences among compartments ($p = 0.016$), with CK and RS differing significantly from L ($p = 0.024$ and $p = 0.06$, respectively).

Overall, bacterial alpha diversity was highest in rhizosphere soil, followed by bulk soil, root, stem, and leaf in descending order. These results demonstrate a clear diversity gradient along the soil–plant continuum, with strong host selection gradually reducing bacterial diversity from the rhizosphere microenvironment to aboveground foliar tissues.

Beta Diversity Analysis

Beta diversity refers to the dissimilarity in species composition between different communities along environmental gradients or the rate of species turnover along such gradients, and is thus also

known as between-habitat diversity. The results of beta diversity analysis for each group are shown in Fig. 4.

PCoA Analysis

The results of Principal Coordinates Analysis (PCoA) are shown in Fig. 4-A. It can be observed that PCo1 was the core differential dimension, explaining 48.1% of the variation among samples and serving as a key driver of differences in community composition. Sample points in each group were relatively compact and clustered, indicating high stability of intra-group community composition.

There was a certain degree of overlap between CK and RS in both PCo1 and PCo2 dimensions, indicating no significant difference in community composition between these two groups, which formed a subgroup. The same was true for T and L; the two subgroups were widely separated along the PCo1 dimension, showing extremely significant differences.

Samples in the R group were positioned between the two subgroups along the PCo1 dimension and showed significant differences from both subgroups along the PCo2 dimension. Overall, these results indicated significant differences in community composition among the subgroups.

Hierarchical Clustering Analysis

The results of hierarchical clustering analysis are shown in Fig. 4-B. The left panel is a hierarchical clustering tree, where samples are clustered based on their similarity to each other; the shorter the branch length between two samples, the more similar they are. The right panel is a stacked bar chart of the top 20 genera in terms of abundance.

From the genus-level hierarchical clustering tree, T and L clustered into one major branch, RS and CK into another major branch, with relatively uniform community structure within each group. The R group clustered separately from the other two branches, indicating significant differences in microbial community structure among groups overall.

Furthermore, samples 1, 2, and 3 in each group showed significant differences in overall composition from other samples within the same group, each forming independent clustering branches. The distribution of the top 20 genera in the stacked bar chart across different samples also indicated significant differences among groups.

Inter-group Difference Analysis

The boxplot of inter-group difference analysis is shown in Fig. 4-C, with a p-value of 0.001 indicating statistically significant differences.

With R as the reference, the median, box range, and whisker extension range of the CK and RS groups were highly consistent, showing almost identical data distribution patterns and minimal intra-group variation.

The median values of the T and L groups were significantly higher than those of the CK and RS groups, with narrower box ranges and more concentrated data distribution; the box ranges of the T and L groups were highly close to their medians, indicating further reduced variation.

In summary, the bacterial community data showed the smallest intra-group variation in the CK and RS groups. The T and L groups also had minimal intra-group variation, with overall levels significantly higher than those of the previous two groups. Inter-group differences were mainly reflected in the stratification of "CK/RS vs R/T/L".

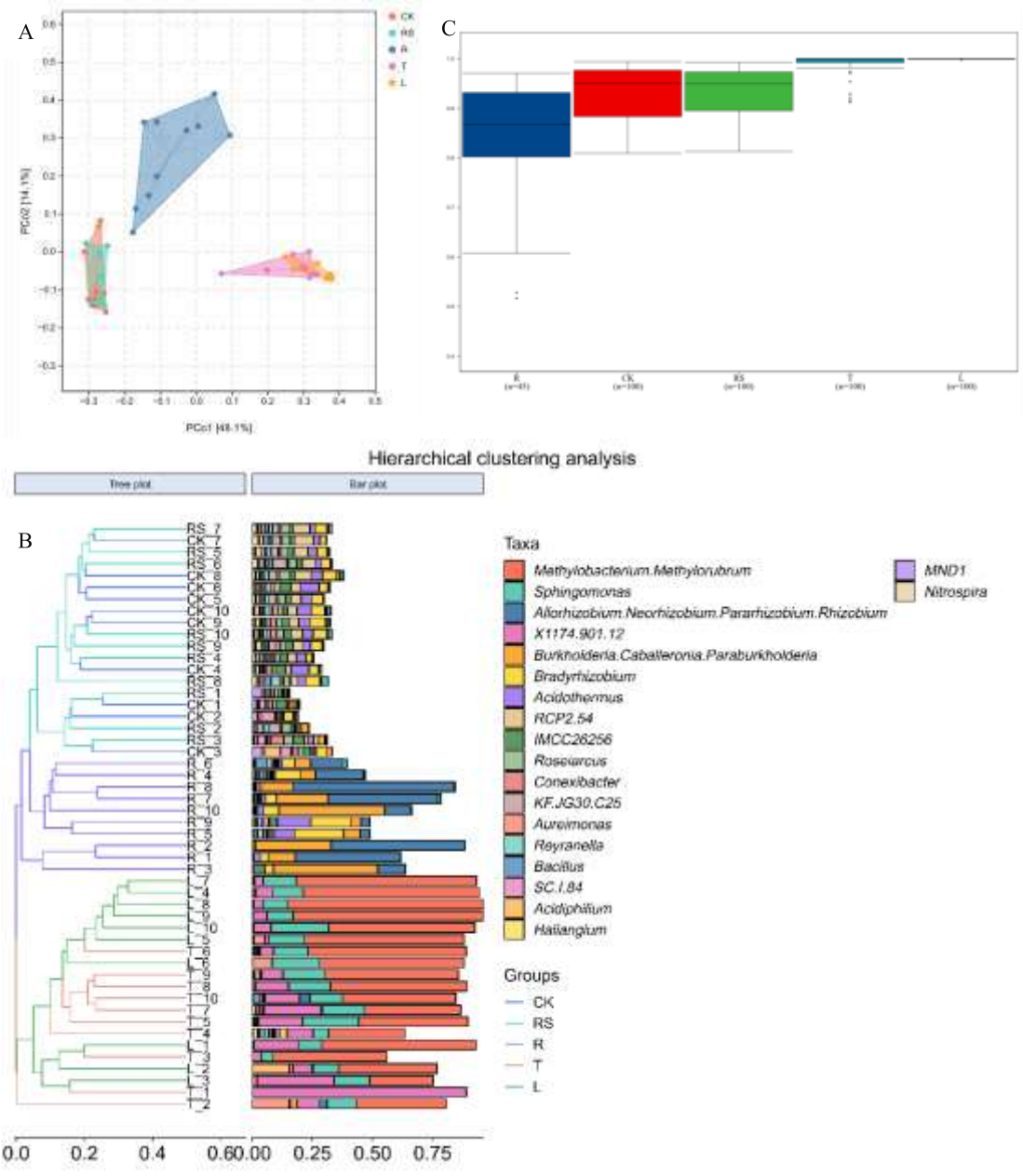

Fig. 4 Beta diversity of samples in each group

A. PCoA plot: Each point represents a sample, and points of different colors (ellipses with a confidence level of 0.95) indicate different groups. The percentages in parentheses on the axes represent the proportion of sample variation

(distance matrix) explained by the corresponding axis. B. Hierarchical clustering tree. The panel on the left is a hierarchical clustering tree diagram; the panel on the right is a stacked bar chart of the top 10 genera in abundance ranking. C. Inter-group difference analysis plot. X-axis: Represents different sample groups, with the number of samples in each group shown in parentheses. Y-axis: Represents the Good's coverage index.

Species Difference Analysis and Biomarker Species

ASV Venn Diagram

The Venn diagram obtained when the occurrence probability of samples was > 0 is shown in Fig. 5A. As illustrated, there were 28,160, 29,895, 16,474, 12,991, and 5,728 ASVs in CK, RS, R, T, and L, respectively. This result is consistent with the species composition analysis, indicating that RS had the highest species richness, followed by a gradual decrease in the order of CK, R, T, and L.

A certain degree of overlap in ASVs was observed among all groups, with 17 shared ASVs across the five groups. Except for *Weissella cibaria* (a lactic acid bacterium with multiple probiotic properties), all other shared ASVs belonged to unculturable bacteria.

The Venn diagram also revealed that 1,283 ASVs were shared among CK, RS, and R, while 4,557 ASVs were shared between CK and RS, indicating smaller differences between these two groups. Among R, T, and L, 37 ASVs were shared, with 49 ASVs shared between R and L, and 1,654 ASVs shared between T and L, thus indicating smaller differences between T and L.

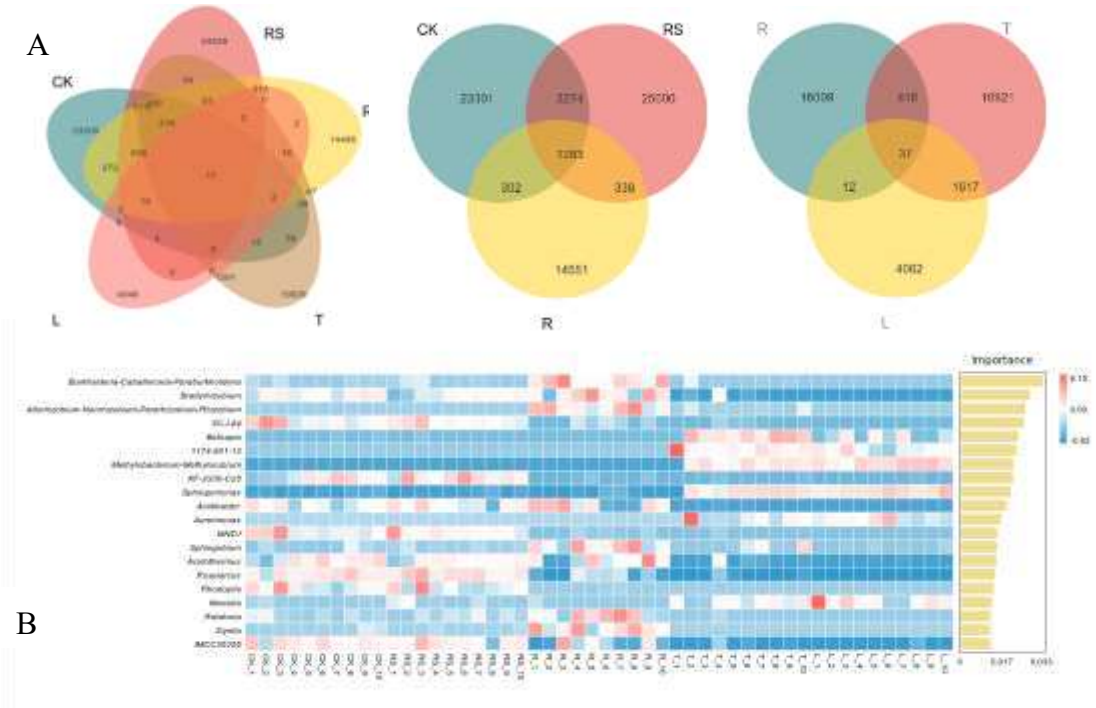

Fig. 5 Analysis Diagram of Species Differences and Biomarker Species

A. Venn diagram. Each color block represents a group, the overlapping area between color blocks indicates the ASVs/OTUs shared by the corresponding groups, and the number in each block represents the number of ASVs/OTUs contained in that block.

B. Random forest analysis identifying key biomarker genera that distinguish different sample groups. The horizontal axis represents each sample number, and the vertical axis represents the taxonomic names at the genus level of ASV/OTU. It indicates the normalized relative abundance, with red representing high abundance (maximum value of 6.15) and blue representing low abundance (minimum value of -0.82).

Random Forest Analysis

The random forest analysis diagram of samples is shown in Fig. 5B. The top 20 genera in abundance were unevenly distributed across different samples. *Methylobacterium*, *Methylorubrum*, and *Sphingomonas* were dominant genera in T and L; the former two had higher abundances in L and were almost absent in roots.

Bacteria of the genus *Sphingomonas* (family *Sphingomonadaceae*) had high abundances in stems and leaves. *Aureimonas* (family *Rhizobiaceae*) was also a dominant genus in T and L.

Belnapia (family *Acetobacteraceae*) was a dominant genus in T, while *Massilia* (family *Oxalobacteraceae*) was dominant in L. There was almost no overlap in dominant genera between T/L and R (except for *Sphingobium*).

The root (R) harbored numerous dominant genera. *Burkholderia*, *Caballeronia*, and *Paraburkholderia* (family *Burkholderiaceae*) were the most important dominant genera in R. *Allorhizobium*, *Neorhizobium*, *Pararhizobium*, and *Rhizobium* (family *Rhizobiaceae*), as well as *Bradyrhizobium* (family *Bradyrhizobiaceae*), were also key dominant genera.

Several genera of the family *Burkholderiaceae* (order *Burkholderiales*, class *Betaproteobacteria*)—*Burkholderia*, *Caballeronia*, and *Paraburkholderia*—were centered in R and their abundances decreased gradually toward RS, CK, T, and L. These genera are widely distributed in environments such as soil and plant rhizospheres, and some species can be used as biopesticides or organic compound degrading agents.

Sphingobium (order *Sphingomonadales*) is associated with "environmental pollutant degradation" and "ecological material cycling". *Dyella* (family *Hydrogenophilaceae*), centered on "degrading complex organic matter" and "adapting to extreme environments", shows clear application potential in environmental remediation, soil improvement, and plant growth promotion.

Acidibacter and *Acidothermus* also played certain important roles. Some species of *Ralstonia* (family *Burkholderiaceae*) are important pathogens that can cause plant root rot.

There were relatively more shared dominant genera between R and the two groups (RS, CK), such as *Acidibacter*, *Acidothermus*, *Bradyrhizobium*, and *Ralstonia*.

In addition to the above genera, *Roseiarcus* (family *Roseiarcaceae*, order *Rhodospirillales*) and *Rhodopila* (family *Rhodospirillaceae*) were unique dominant genera in RS and CK. Their physiological functions are closely related to the metabolic characteristics of "non-sulfur

photosynthetic bacteria", mainly reflected in energy utilization, material transformation, and ecological adaptation.

Among the unidentified genera, 1174-901-12 was more abundant in above-ground plant tissues, while IMCC26256, KF-JG30-C25, SC-I-84, and MND1 had the highest abundances in soil.

Overall results indicated that although RS had the highest species diversity among the five groups based on species composition analysis, R—acting as a key site connecting soil to plant tissues—harbored more dominant genera.

Correlation Network Analysis

Co-occurrence network analysis was performed to reveal topological roles and interaction patterns of the dominant bacterial genera across the soil–plant continuum of *S. china* L. Nodes were classified into four ecological roles (peripherals, connectors, module hubs, and network hubs) based on *Zi* (within-module connectivity) and *Pi* (among-module connectivity) scores.

As shown in Fig. 6, most of the top 10 abundant genera were distributed in the connector region (high *Pi*, low *Zi*), indicating that these dominant taxa primarily served as inter-module bridges to maintain connectivity across the entire network. These genera did not act as local keystone taxa (module hubs) or core global regulators (network hubs), but instead strengthened cross-module interactions to support community stability.

Notably, no obvious network hubs were detected in the bacterial community, suggesting the absence of dominant keystone taxa that control the whole network structure. This modular organization relying on connectors rather than central hubs enhances functional redundancy and environmental robustness, allowing the community to resist disturbance without catastrophic collapse.

Taken together, the bacterial network of *S. china* is maintained by connector-dominated modularity rather than centralized keystone taxa. This structural feature supports high stability and functional resilience, which is critical for persistent microbial colonization and synergistic plant–microbe interactions under field conditions.

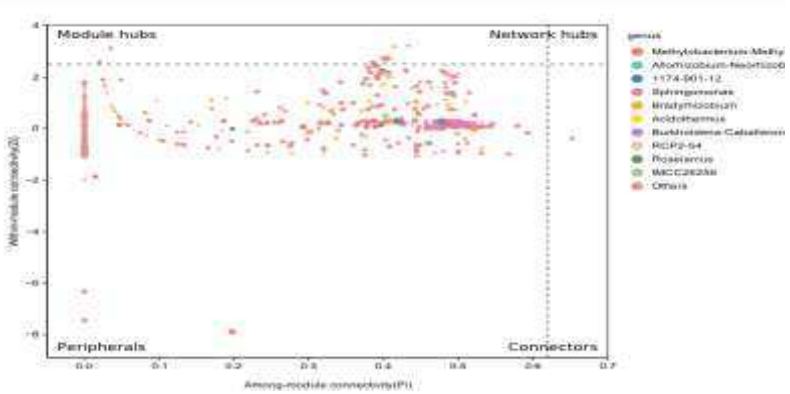

Fig. 6 Correlation Network Analysis Diagram

In the figure, the X and Y axes represent the Pi (among-module connectivity) and Zi (within-module connectivity) values of nodes in the network, respectively. Node size is proportional to its abundance (expressed as log2(CPM/n)), and nodes belonging to the top 10 most abundant genera are marked with different colors.

Metabolic Difference Analysis

Metabolic functions of bacterial communities across different compartments were predicted using PICRUSt2, with root (R) samples as the reference control (Fig. 7).

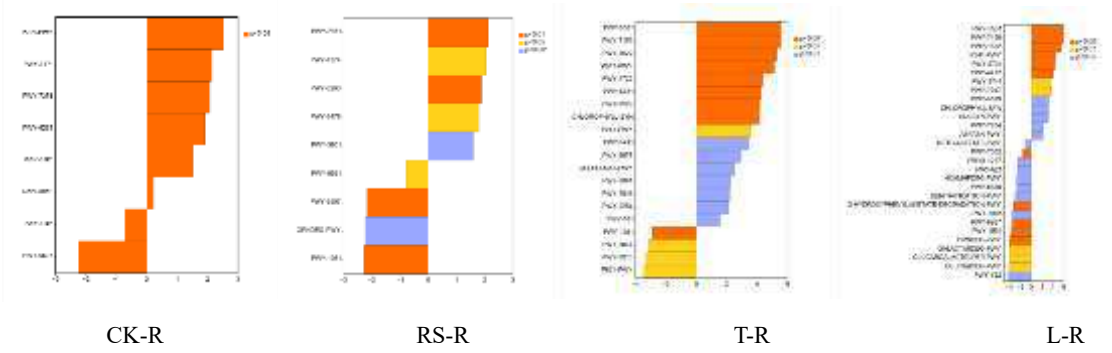

Fig. 7 Differential analysis of dominant bacterial genera among different sample compartments

The positive values of logFC (log2(fold change)) on the horizontal axis in the figure represent upregulation of the experimental group relative to the control group, while negative values indicate downregulation; the vertical axis represents different pathway/group labels, and the significance levels are displayed in different colors.

In bulk soil (CK) and rhizosphere soil (RS), pathways including PWY-7371, PWY-7374, and PWY-6263 were significantly up-regulated ($p < 0.05$). These pathways are mainly involved in vitamin K_2 biosynthesis, which promotes phosphorus and iron uptake by root cells and enhances plant stress resistance. Additionally, PWY-3801, PWY-6478 (RS), and P261-PWY (CK) were significantly enriched, contributing to enhanced carbon metabolism.

In stems (T) and leaves (L), pathways associated with chlorophyll synthesis (e.g., PWY-5531, PWY-7159, PWY-5529, CHLOROPHYLL-SYN) and sugar metabolism (e.g., P341-PWY, PWY-4722, PWY-6731, SUCSYN-PWY) were significantly or extremely significantly activated, indicating highly active photosynthesis and carbon fixation in aboveground tissues. Several supplementary pathways related to nitrogen metabolism, methyl transfer, and coenzyme synthesis were also enhanced, providing essential nutrients for plant growth.

In roots (R), multiple degradation pathways were highly expressed, including PWY-6957, GALACTARDEG-PWY, GLUCARDEG-PWY, and PWY-722, which degrade mandelic acid, D-glucaric acid, D-galactaric acid, and other potentially inhibitory compounds. Pathways such as PWY-1361 and 3-HYDROXYPHENYLACETATE-DEGRADATION-PWY were responsible for breaking down benzene-ring derivatives and harmful substances, reducing phytotoxicity (Wang et

al., 2022).

Notably, PWY-7391 (isoprene biosynthesis) and related cell-division pathways were specifically enriched in roots. These pathways promote cell proliferation and may directly contribute to rhizome enlargement—a key trait determining the medicinal and economic value of *S. china* L.

Overall, bacterial communities exhibited clear niche-adapted metabolic division of labor: soil and root bacteria supported nutrient acquisition and detoxification, while aboveground endophytes promoted photosynthesis and carbon accumulation. This functional coordination effectively boosts plant growth, stress tolerance, and rhizome development.

4 Discussion

Plant-associated bacteria act as a critical "second genome" that mediates nutrient acquisition, stress tolerance, and development (Berendsen et al., 2012; Müller et al., 2016). In the present study, we demonstrated that *S. china* L. significantly modifies rhizosphere properties and drives strong niche differentiation of bacterial communities across the soil–rhizosphere–root–stem–leaf continuum. These patterns align with the widely observed host-filtering effect along the soil–plant axis (Reinhold-Hurek et al., 2015; Xiong et al., 2021; Wang et al., 2019), in which microbial diversity gradually decreases from soil to aboveground tissues due to physical and biochemical filtering by the host.

The highest bacterial α-diversity in rhizosphere soil is consistent with numerous reports in medicinal and crop plants (Bulgarelli et al., 2013; Li et al., 2022). Root exudates and rhizodeposits provide abundant carbon sources and stimulate microbial proliferation, while elevated pH and available phosphorus further improve microhabitat quality. Clear separation between soil and endosphere communities confirms that roots serve as a key transitional hub, which has been recognized as a universal feature of plant microbiome assembly (Xiong et al., 2021).

Tissue-specific dominant genera further support functional partitioning. *Methylobacterium*, *Aureimonas*, and *Sphingomonas* in stems and leaves are typical methylotrophic and plant growth–promoting taxa that enhance photosynthesis and nitrogen supply (Yousaf et al., 2024). Roots were enriched in *Burkholderia–Caballeronia–Paraburkholderia* and rhizobia. Although *S. china* does not form nodules, these non-symbiotic diazotrophs can colonize intercellular spaces, mineralize nutrients, and promote growth via nitrogen fixation and phytohormone secretion (Li et al., 2022), which expands our understanding of growth-promotion mechanisms in non-leguminous medicinal plants.

Co-occurrence network analysis revealed that abundant genera mainly function as connectors rather than module or network hubs. This connector-dominated structure enhances modularity and functional redundancy, thereby improving community stability and disturbance resistance (Deng et al., 2012). The absence of dominant hub nodes suggests the community relies on collaborative interactions rather than keystone species, a trait beneficial for stable colonization under field conditions.

Functional prediction revealed clear division of labor: soil and root bacteria contribute to nutrient uptake and xenobiotic degradation, while aboveground taxa enhance carbon metabolism. Notably,

isoprene biosynthesis and cell-division pathways were significantly enriched in roots, which may accelerate cell proliferation and directly promote rhizome enlargement. This mechanism provides a novel microecological explanation for the formation of medicinal organs in *S. china*.

This study has several limitations. We only used 16S rRNA sequencing, which limits taxonomic and functional resolution; samples were collected from a single site and growth stage. Future studies should combine metagenomics, metabolomics, and isolate validation to explore causal relationships and develop beneficial microbial agents for cultivation optimization.

Conclusion

This study reveals clear spatial niche differentiation and functional division of bacterial communities across bulk soil, rhizosphere, root, stem, and leaf of *S. china* L. Bacterial α-diversity peaks in the rhizosphere and declines toward aboveground tissues, with roots acting as a transitional hub. Dominant genera show strong tissue specificity, and abundant taxa function as connectors to maintain network stability. Functional prediction demonstrates niche-adapted metabolic roles: soil/root bacteria enhance nutrient uptake and harmful compound degradation, while aboveground endophytes promote photosynthesis. Root-associated pathways related to isoprene biosynthesis and cell division may directly facilitate rhizome enlargement. These findings illustrate how *S. china* shapes a structurally stable and functionally complementary microbiome to improve nutrient utilization, stress tolerance, and rhizome development. This work provides a microecological foundation for optimizing cultivation and improving the yield and quality of this important medicinal plant via beneficial microbial regulation.

Declarations

Ethical approval

This study did not involve human participants or experimental animals, and ethical approval is not applicable.

Funding

This study was supported by the Central Government Guided Local Science and Technology Development Fund Project (No. Z135050009017-429) and the School-level Research Project of Xianning Vocational and Technical College (No. 2025005).

Data availability

All original experimental data, analyzed data, figures and relevant materials of this study are preserved in the laboratory of the corresponding author. The datasets are available from the corresponding author on reasonable academic request. No public datasets were used in this study.

References

Anderson MJ, Ellingsen KE, McArdle BH (2006) Multivariate dispersion as a measure of beta

585 diversity. Ecology Letters, 9(6): 683–693. doi:10.1111/j.1461-0248.2006.00926.x

586 Beckers B, Op De Beeck M, Weyens N, Boerjan W Vangronsveld J (2017) Structural variability and
587 niche differentiation in the rhizosphere and endosphere bacterial microbiome of field-grown
588 poplar trees. Microbiome, 5(1):25. https://doi.org/10.1186/s40168-014-0025-3

589 Bokulich N A, Kaehler BD, Ram RJ, Matthew D, Evan B, Rob K, Huttley GA, Gregory CJ (2018)
590 Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-
591 feature-classifier plugin. Microbiome, 6(1): 90. doi:10.1186/s40168-018-0470-z

592 Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Caporaso JG (2019) QIIME 2: Reproducible,
593 interactive, scalable, and extensible microbiome data science. Nature Biotechnology, 37(8): 852–
594 857. oi:10.1038/s41587-019-0209-9

595 Breiman L (2001) Random forests. Machine Learning, 45(1):5–32. doi:10.1023/A:1010933404324

596 Bulgarelli D, Schlaeppi K, Spaepen S, Ver Loren van Themaat E, Schulze-Lefert P. (2013). Structure
597 and functions of the bacterial microbiota of plants. Annual Review of Plant Biology, 64:807–838.
598 doi:10.1146/annurev-arplant-050312-120106

599 Cai XH, Wang RY, Hu SN, Li Y, Zhou W (2020) Research progress on *Sphingobacterium*. Genomics
600 and Applied Biology, 39(5):2096–2102. doi:10.13417/j.gab.039.002096

601 Callahan BJ, McMurdie PJ, Rosen M J, Han AW, Johnson AJA, Holmes SP (2016) DADA2: High-
602 resolution sample inference from Illumina amplicon data. Nature Methods, 13(7):581–583.
603 doi:10.1038/nmeth.3869

604 Chao A (1984) Nonparametric estimation of the number of classes in a population. Scandinavian
605 Journal of Statistics, 11(4):265–270. doi:10.2307/4615594

606 Chinese Pharmacopoeia Commission (2020) Pharmacopoeia of the People's Republic of China (Vol.
607 I). China Medical Science Press.

608 Clarke KR (1993) Non-parametric multivariate analyses of changes in community structure.
609 Australian Journal of Ecology, 18(1):117–143. doi:10.1046/j.1442-9993.2001.01070.x

610 Deng Y, Jiang YH, Yang Y, He ZL, Luo F, Zhou JZ (2012) Molecular ecological network analyses.
611 *BMC Bioinformatics*, 13(1):113. doi:10.1186/1471-2105-13-113

612 Deng YX, Li CL, Li HQ, Lu SF (2018) Identification and characterization of flavonoid biosynthetic
613 enzyme genes in *Salvia miltiorrhiza* (Lamiaceae). Molecules, 23(6):1467.

614 Douglas GM, Maffei VJ, Zaneveld J, Yurge SN, Brown JR, Taylor CM, Huttenhower C, Langille MGI
615 (2020) PICRUSt2: An improved and customizable approach for metagenome inference. Nature
616 Biotechnology, 38(5):586–591. doi:10.3390/molecules23061467

617 Good IJ (1953) The population frequency of species and the estimation of population parameters.
618 Biometrics, 40(2):237–246. doi:10.2307/2333250

619 Horton MW, Bodenhausen N, Beilsmith K, Meng D, Muegge BD, Subramanian S, Vetter MM, Vilhj
620 álmsson BJ, Nordborg M, Gordon JI (2014) Genome-wide association study of *Arabidopsis*
621 *thaliana* leaf microbial community. Nature Communications, 5:4688. doi:10.1038/ncomms6320

622 Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: A novel method for rapid multiple sequence
623 alignment based on fast Fourier transform. Nucleic Acids Research, 30(14):3059–3066.
624 doi:10.1093/nar/gkf436

625 Li CC (2020) Diversity, community structure of endophytic and soil microorganisms in dominant
626 grassland plants and their responses to environmental factors [D]. Beijing: University of
627 Chinese Academy of Sciences. doi:10.3389/fmicb.2021.736219

628 Liaw A & Wiener M (2002) Classification and regression by random Forest. *R News*, 2(3): 18–22.
629 doi:10.1017/CBO9780511815870.012

630 Lozupone C & Knight R (2005) UniFrac: A new phylogenetic method for comparing microbial
631 communities. Applied and Environmental Microbiology, 71(12):8228–8235.
632 doi:10.1128/AEM.00628-05

633 Lozupone CA, Hamady M, Kelley ST, Knight R (2007) Quantitative and qualitative β diversity
634 measures lead to different insights into factors that structure microbial communities. Applied and
635 Environmental Microbiology, 73(5):1576–1585. doi:10.1038/ismej.2016.148

636 Yousaf T, Saleem F, Andleeb S, Ali M, Haque MFU (2024) Methylotrophic bacteria from rice paddy
637 soils: mineral-nitrogen-utilizing isolates richness in bulk soil and rhizosphere. World J Microbiol
638 Biotechnol 40:188. https://doi.org/10.1007/s11274-024-04000-3

639 Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads.
640 EMBnet.journal, 17(1):10–12. https://doi.org/10.14806/ej.17.1.200

641 McArdle BH & Anderson MJ (2001) Fitting multivariate models to community data: A comment on
642 distance-based redundancy analysis. Ecology, 82(1):290–297. https://doi.org/10.1890/0012-
643 9658(2001)082[0290:FMMTCD]2.0.CO;2

644 Müller DB, Vogel C, Bai Y, Vorholt JA (2016) The plant microbiota: composition, functions, and
645 engineering. Nature Reviews Microbiology, 14(10): 614-626.
646 ttps://doi.org/10.1016/j.copbio.2021.07.003

647 Price MN, Dehal PS & Arkin AP (2010) FastTree 2 – Approximately maximum-likelihood trees for
648 large alignments. PLOS ONE, 5(3):e9490. https://doi.org/10.1371/journal.pone.0009490

649 Richardson AE, Barea JM, McNeill AM, Prigent-Combaret C (2009) Acquisition of phosphorus and

nitrogen in the rhizosphere and plant growth promotion by microorganisms. Plant and Soil, 321(1–2):305–339. https://doi.org/10.1007/s11104-016-2926-5

Shannon CE (1948). A mathematical theory of communication. Bell System Technical Journal, 27, 379–423, 623–656. https://doi.org/10.1002/j.1538-7305.1948.tb00917.x

Valentinuzzi F, Pii Y, Vigani G, Mimmo T, Terzano R, Crecchio C, Cesco S (2015) Phosphorus and iron deficiencies induce a metabolic reprogramming and affect the exudation traits of the woody plant *Fragaria* × *ananassa*. Journal of Experimental Botany, 66(20):6483–6494. https://doi.org/10.1093/jxb/erab386

Wang F, Men X, Zhang G, Liang K, Xin Y, Wang J, Li A, Zhang H, Liu H, Wu L (2018) Assessment of 16S rRNA gene primers for studying bacterial community structure and function of aging flue-cured tobaccos. AMB Express, 8:182. https://doi.org/10.1186/s13568-018-0713-1

Wang QZ, Geng SS, Wang LR, Wen ZQ, Sun XM, Huang H (2022) Bacterial mandelic acid degradation pathway and its application in biotechnology. Journal of Applied Microbiology, 133(2):273–286. https://doi.org/10.1007/s00253-021-11289-3

Wang GZ, van der Putten WH, Zhang JL (2025) Steering plant–soil feedback for sustainable agriculture. *Science*, 389(6758). https://doi.org/10.1126/science.ads2506

Wang J, Wang DM, Ren Y, Xu L, Wang Y (2019) Coupling relationships between soil microbes and soil nutrients under different hydrologic conditions in the riparian zone of the Lijiang River. Acta Ecologica Sinica, 39(8):2687–2695. https://doi.org/10.1016/j.geoderma.2023.114628

Warton DI, Wright ST & Wang Y (2012) Distance-based multivariate analyses confound location and dispersion effects. Methods in Ecology and Evolution, 3(1):89–101. https://doi.org/10.1111/j.2041-210X.2005.00012.x

Zaura E, Keijser BJF, Huse SM, Crielaard W (2009) Defining the healthy “core microbiome” of oral microbial communities. BMC Microbiology, 9:259. https://doi.org/10.1186/s12866-016-0868-3

Zhang C, Zhou DF, Wang MY, Song YZ, Zhang C, Zhang MM, Sun J, Yao L, Mo XH, Ma ZX, Yuan XJ, Shao Y, Wang HR, Dong SH, Bao K, Lu SH, Sadilek M, Kalyuzhnaya MG, Xing XH, Yang S (2024) Phosphoribosylpyrophosphate synthetase as a metabolic valve advances *Methylobacterium*/*Methylorubrum* phyllosphere colonization and plant growth. Nature Communications, 15(1):5969. https://doi.org/10.1038/s41467-024-50342-9

Zhang D (2007) Common Experimental Methods and Techniques in Ecology. Chemical Industry Press, Beijing.

Zhao B & He SJ (2007) Microbiology Experiment. Science Press, Beijing.

Zhao L, Hu X & Peng W (2022) Endophytic microbes and their roles in medicinal plant quality.

683 *Industrial Crops and Products*, 187:115389. https://doi.org/10.1007/978-981-99-9223-5_5

Supplementary Files

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

- Supplementaryfile1.docx