Journal of Microbiology (2022) Vol. 60, No. 10, pp. 986–997
DOI 10.1007/s12275-022-2363-x
eISSN 1976-3794
pISSN 1225-8873

Microbial co-occurrence network in the rhizosphere microbiome: its association with physicochemical properties and soybean yield at a regional scale[§]

Sarbjeet Niraula[#], Meaghan Rose, and Woo-Suk Chang[*]

Department of Biology, University of Texas at Arlington, Arlington, Texas 76019, USA
[#]Present address: Division of Vascular Surgery, Department of Surgery, Northwestern University Feinberg School of Medicine, Chicago, Illinois 60611, USA

(Received Aug 12, 2022 / Revised Aug 28, 2022 / Accepted Aug 29, 2022)

Microbial communities in the rhizosphere play a crucial role in determining plant growth and crop yield. A few studies have been performed to evaluate the diversity and co-occurrence patterns of rhizosphere microbiomes in soybean (*Glycine max*) at a regional scale. Here, we used a culture-independent method to compare the bacterial communities of the soybean rhizosphere between Nebraska (NE), a high-yield state, and Oklahoma (OK), a low-yield state. It is well known that the rhizosphere microbiome is a subset of microbes that ultimately get colonized by microbial communities from the surrounding bulk soil. Therefore, we hypothesized that differences in the soybean yield are attributed to the variations in the rhizosphere microbes at taxonomic, functional, and community levels. In addition, soil physicochemical properties were also evaluated from each sampling site for comparative study. Our result showed that distinct clusters were formed between NE and OK in terms of their soil physicochemical property. Among 3 primary nutrients (i.e., nitrogen, phosphorus, and potassium), potassium is more positively correlated with the high-yield state NE samples. We also attempted to identify keystone communities that significantly affected the soybean yield using co-occurrence network patterns. Network analysis revealed that communities formed distinct clusters in which members of modules having significantly positive correlations with the soybean yield were more abundant in NE than OK. In addition, we identified the most influential bacteria for the soybean yield in the identified modules. For instance, included are class *Anaerolineae*, family *Micromonosporaceae*, genus *Plantomyces*, and genus *Nitrospira* in the most complex module (ME9) and genus *Rhizobium* in ME23. This research would help to further identify a way to increase soybean yield in low-yield states in the U.S. as well as worldwide by reconstructing the microbial communities in the rhizosphere.

***Keywords*:** microbial co-occurrence network, soybean rhizosphere, microbiome, soybean yield, soil physicochemical properties, microbial communities

Introduction

The soybean (*Glycine max*) rhizosphere, the soil region found in the immediate vicinity of the plant root, harbors highly dynamic microbial communities that interact with the host (Hartmann *et al.*, 2008). As a soybean endosymbiont, not only does *Bradyrhizobium japonicum* fulfill more than half of the nitrogen requirement by the host, but it also provides other plant growth promoting traits in the rhizosphere that greatly influence growth, vigor, and eventually yield of soybeans (Salvagiotti *et al.*, 2008). The prominence of the symbiotic nitrogen fixation associated with soybean is so large that it accounts for 77% of the nitrogen fixed by leguminous crops globally (Herridge *et al.*, 2008). In addition to the nitrogen fixation, the host performance is also influenced by other interacting biotic and abiotic factors.

Rhizosphere microbes are selected from the microbial community in locally adapted near bulk soil in response to the plant root exudates (Sugiyama, 2019). Therefore, changes brought in the indigenous bulk soil community due to environmental factors influence species assembly and composition in the rhizosphere (Philippot *et al.*, 2013; Mendes *et al.*, 2014). Numerous studies have been performed to investigate the diversity and composition of rhizosphere microbiomes under different experimental conditions, including soybean genotype, growth stages, soil parameters, and geographical locations (Jin *et al.*, 2009; Sugiyama *et al.*, 2014; Rascovan *et al.*, 2016). Most of the studies focus on identifying taxa abundances, alpha and beta diversities, and correlations with environmental factors. However, microorganisms develop more complex interactions within their natural habitat, involving mutualism, competition, parasitism, and commensalism (Faust and Raes, 2012) with the host, which may not be possible to be understood by traditional methods.

Microbial network analyses, especially co-occurrence networks, have been increasingly used to investigate microbial community structures in terrestrial and aquatic habitats such as soil and the ocean (Barberán *et al.*, 2012; Mendes *et al.*, 2014; Sugiyama *et al.*, 2014; Cram *et al.*, 2015). One study has

*For correspondence. E-mail: wschang@uta.edu; Tel.: +1-817-272-3280; Fax: +1-817-272-2855
[§]Supplemental material for this article may be found at https://doi.org/10.1007/s12275-022-2363-x.

demonstrated that soil microbial communities show non-random co-occurrence patterns and can be modular in nature (Barberán *et al.*, 2012). The identified modules represent densely inter-connected nodes that can be interpreted as groups of taxa with overlapping niches. The co-occurring taxa are more closely related than other taxa that would be expected to occur by chance (Newman and Girvan, 2004; Faust and Raes, 2012). Although interpretations of these networks are not straightforward, they allow us to predict hub species and their interaction in a particular niche.

Networks have been more prevalent in determination of protein interactions and prediction of gene functions from RNA-seq experiments by identifying gene significance and hub genes. Similar approaches have been employed to identify taxa significance and keystone species using co-occurrence networks with in-silico sequence data (Berry and Widder, 2014; Tao *et al.*, 2018). High-throughput sequencing technologies have facilitated rapid and efficient microbial community analysis by direct sequencing of 16S ribosomal RNA (rRNA) genes, a small subunit rRNA gene in prokaryotes. Molecular features of the 16S rRNA gene including its ubiquitousness, high degree of functional constancy, larger size with many domains, and mutations at very different rates, make it the ultimate molecular chronometer (Woese, 1987; Mizrahi-Man *et al.*, 2013).

In this study, we performed a comprehensive evaluation of the microbial diversity and predicted functions of microbial communities in the soybean rhizosphere by combining microbiome data and soil physicochemical properties from the soybean farms of Nebraska (NE) and Oklahoma (OK) as high and low yielding states, respectively. In addition, we identified strongly co-occurring groups of taxa matching to the sample traits. By analyzing microbial communities at taxonomic and functional levels, we aimed to identify key drivers for the difference in the soybean yield. This approach will also help us decipher rhizosphere colonization and interaction patterns of microbial communities under different environmental conditions.

Materials and Methods

Rhizosphere soil collection and physicochemical property evaluation

Rhizosphere soil samples were collected from a total of 12 sampling sites of soybean fields in NE and OK (6 sites per each state) during growing season (e.g., R1-R2 stages) in 2010 (Fig. 1). It has been reported that the microbial populations reach the highest level at soybean R2 stage (Meng *et al.*, 2010). Rhizosphere soil sampling was performed as previously described (Naim, 1965; Luster *et al.*, 2009) with a slight modification. Briefly, after pulling out the soybean plants from the field, rhizosphere soil was obtained by shaking the root system for 5 min in the air. Any plant debris and residual roots in the soil were removed subsequently. Samples were collected in triplets, transported in ice and stored at -80°C until processed (Supplementary data Fig. S1). Rhizosphere soil samples from each triplet were mixed in equal proportion and composite soil samples were sent to Soil, Water and Forage Testing

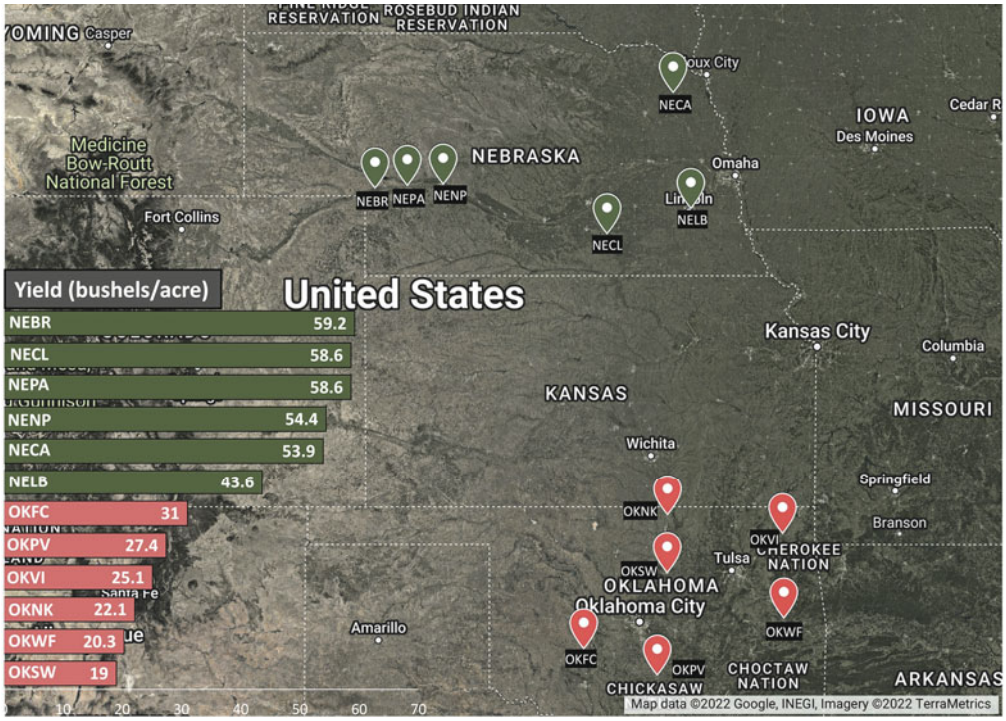

Fig. 1. Rhizosphere soil sampling sites in Nebraska (NE) and Oklahoma (OK) (Google Map, 2022). Horizontal bars represent soybean yield (bushels/acre). Location (e.g., county or city) abbreviation in NE: NEBR, Brule; NECL, Clay Center; NEPA, Paxton; NENP, North Platte; NECA, Concord; NELB, Lincoln. Location (e.g., county or city) abbreviation in OK: OKFC, Fort Cobb; OKPV, Pauls Valley; OKVI, Vinita; OKNK, Newkirk; OKWF, Webber Falls; OKSW, Stillwater.

Laboratory at Texas A&M AgriLife Extension Service (https://soiltesting.tamu.edu/) for their physicochemical properties. We also incorporated average precipitation and atmospheric temperature data during the cultivation period (e.g., May to September in 2010) for comparative analysis. Soybean yield was also monitored in each site (Fig. 1).

Library preparation and sequencing

DNA isolation was performed from each triplicate by using PowerSoil DNA isolation kit (Qiagen), following manufacturer's instruction. The concentration and purity of soil DNA was assessed using NanoDrop spectrophotometer and normalized DNA from each triplet was mixed in equal amount. Library preparation and sequencing was performed as previously described (Niraula *et al.*, 2021). Briefly, PCR amplification of the 16S rRNA V3-V4 hypervariable region was carried out in triplets of 25 ml reaction volume containing KAPA HiFi master mix and Illumina adaptor ligated universal primers 341F (5′-tcgtcggcagcgtcagatgtgtataagagacag CCTACGGGNGGCWGCAG-3′) and 806R (5′-gtctcgtgggc tcggagatgtgtataagagacagGGACTACNVGGGTWTCTAAT -3′). The thermocycler condition was maintained as follows; 94°C for 3 min, followed by 25 cycles of 94°C for 30 sec, 55°C for 30 sec, 72°C for 30 sec and a final extension of 72°C for 3 min. PCR products were extracted in 1% agarose gel and purified by using QIAquick Gel Extraction Kit (Qiagen). Purified PCR products were sent to the Genomic Sequencing and Analysis Facility (GSAF) at the University of Texas at Austin (https://research.utexas.edu/cbrs/cores/genomics/) for barcoding and sequencing using Illumina MiSeq sequencer to produce 300-bp paired end reads.

Sequence preprocessing and taxonomic assignment

Primer sequences were trimmed from raw sequences using cutadapt 2.4. QIIME2 (Bolyen *et al.*, 2018) followed by merging of paired end reads using vsearch. Sequences were further quality-filtered, dereplicated and clustered de-novo at 97% identity using vsearch plugin. Chimeric sequences were filtered using uchime-denovo in vsearch plugin and operational taxonomic unit (OTU) tables without borderline chimera were selected for further processing. OTUs with frequency 1 were filtered out to avoid possible sequencing artifacts and noise. Mitochondrial and chloroplast related sequences were filtered out and the OTUs with at least Phylum level annotation were retained. To assign the taxonomy, Greengenes (ver. 13.8) 97% OTUs database (Mcdonald *et al.*, 2012) was first trained with v3-v4 primer set and assigned using q2-feature-classifier (Bokulich *et al.*, 2018) against the trained database. To train the database, V3-V4 region of reference reads were extracted using our primer set followed by assigning taxonomy using naïve-bayes classifier in feature-classifier plugin. For comparative study, OTUs shared among at least 50% of the sampling sites from each state were considered for further analysis. Representative sequences were aligned using mafft (Katoh *et al.*, 2002) (via q2-alignment) to construct a phylogenetic tree with fasttree2 (Price *et al.*, 2010) (via q2-phylogeny). For alpha and beta diversity analysis, samples were rarified at the depth of 27,342 reads per sample.

Functional prediction

Functional potential of the microbial community was determined by using the Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) 2.0 (Douglas *et al.*, 2020). PICRUSt has been used to predict functional abundances based on marker gene sequences (Langille *et al.*, 2013). All default parameters were used to predict MetaCyc pathways from OTU representative sequences. The pathway abundance table was converted to phyloseq object and analyzed in R (4.2.0). We also ensured the nearest-sequenced taxon index (NSTI) value for each of the OTUs to be less than 2 using hsp.py function in PICRUSt 2.0.

Network construction

Co-occurrence networks were constructed from the OTUs in both NE and OK samples. For network construction, low-abundance OTUs with less than 0.1% relative abundance across all samples were excluded. Spearman correlation of relative abundance values were performed. Strongly correlated OTUs (> 0.85 and < -0.85) with Benjamini Hochberg (BH) adjusted P-value < 0.001 were selected. In the constructed networks, the nodes represent OTUs and edges represent significant pairwise association between nodes.

Networks were analyzed using igraph 1.3.4 package (Csárdi and Nepusz, 2006) in R. Correlation networks were clustered into modules based on the betweenness of each edge in the network using Girvan Newman algorithm (Newman and Girvan, 2004). Network topological properties such as modularity, clustering coefficient, edge density, diameter, average path length, betweenness centrality, and hub scores were also calculated. Modules containing at least 3 elements were selected from network for further analysis. WGCNA 1.71 package in R was used to calculate the module eigengenes, which represents the first principal component of the module as a representative of the taxonomic profiles of the respective module (Langfelder and Horvath, 2008). Module eigengenes were spearman-correlated with sample traits to identify the relationship of each module to the sample traits. Then, we selected modules that were significantly correlated with soybean yield to compare abundances of module members in NE and OK samples. Network visualization and graphics modifications were performed using Cytoscape 3.8.0 (Shannon *et al.*, 2003) and Gephi 0.9.6 (Bastian *et al.*, 2009).

Statistical analysis

Student's t-test was performed on log2 transformed environmental variables including soybean yield to calculate the difference between mean values. Principal component analysis (PCA) was performed based on variance-covariance matrix. Differences in microbial community (beta diversity) were evaluated by principal coordinate analysis (PCoA) based on Unweighted UniFrac distance matrix. Bioenv test was performed using q2-diversity to find the best subset of environmental variables that showed maximum rank-correlation with the distance matrices (Clarke and Ainsworth, 1993). In addition, two-sided Mantel test (Mantel, 1967) was conducted to test the significance of best spearman rank-correlated variables with the distance matrices using q2-diversity beta- correlation (Spearman, 1904). Permutational multivariate ANOVA

(PERMANOVA) was performed to assess the dissimilarity of microbial communities between two sampling groups. Likewise, module eigengenes were Spearman-correlated with sample traits (i.e., soil physicochemical properties, environmental variables, and soybean yield). Linear discriminant analysis effect size (LEfSe) was performed to identify biomarkers (discriminating features) at different taxonomic levels between two states using Galaxy Hutlab at http://huttenhower.sph.harvard.edu/galaxy/ (Segata *et al.*, 2012). Predicted functions were analyzed using run_deseq2 function in microbiomeMarker package (Cao, 2022) in R. Resulting plots were rendered using ggplot2 in R.

Results

Environmental variables

We incorporated a total of 17 traits including 14 soil physicochemical properties (i.e., organic matter, pH, nitrogen, phosphorus, potassium, calcium, magnesium, sulfur, sodium, iron, zinc, manganese, copper, and electrical conductivity), atmospheric temperature, precipitation, and crop yield. Among all sample traits, potassium ($P < 0.001$), calcium ($P < 0.01$), sodium ($P < 0.05$), electrical conductivity ($P < 0.05$), and soybean yield ($P < 0.001$) were significantly higher in NE soils, whereas the atmospheric temperature ($P < 0.001$) was significantly lower in NE compared to that in OK (Supplementary data Fig. S2). PCA analysis of environmental variables showed a distinct clustering of NE and OK samples (Fig. 2). Most of the variables, except for nitrogen, iron, manganese, and copper, are positively correlated with NE soils. Interestingly, among 3 primary nutrients (i.e., nitrogen, phosphorus, and potassium), potassium is more positively correlated with the NE cluster (Fig. 2), indicating that potassium seems to contribute more than the others for the high soybean yield. As expected, the higher temperature is correlated with the OK cluster (Fig. 2).

Sequencing summary

A total of 338,896 high quality reads with sizes ranging from 440 to 480 bp were obtained after primer removal, merging, quality filtration, denoising, and chimera filtration. Additionally, mitochondrial and chloroplast related OTUs were filtered out. For the comparative study, OTUs that were not common in at least 50% of sampling sites (i.e., 3 out of 6 sampling sties) in each state were excluded, resulting in a total of 8,784 OTUs in both NE and OK soil samples. This exclusion was necessary in order to capture representative OTUs from each state. Among them, 3,872 OTUs (i.e., 44.1% of total OTUs) were shared in both states (Fig. 3B) at 97% nucleotide sequence similarity.

Taxonomic composition and classification of bacteria

Sequences were classified into relative abundance of taxa up to the species level. At the phylum level, phyla with higher than 1% relative abundance were constructed, resulting in

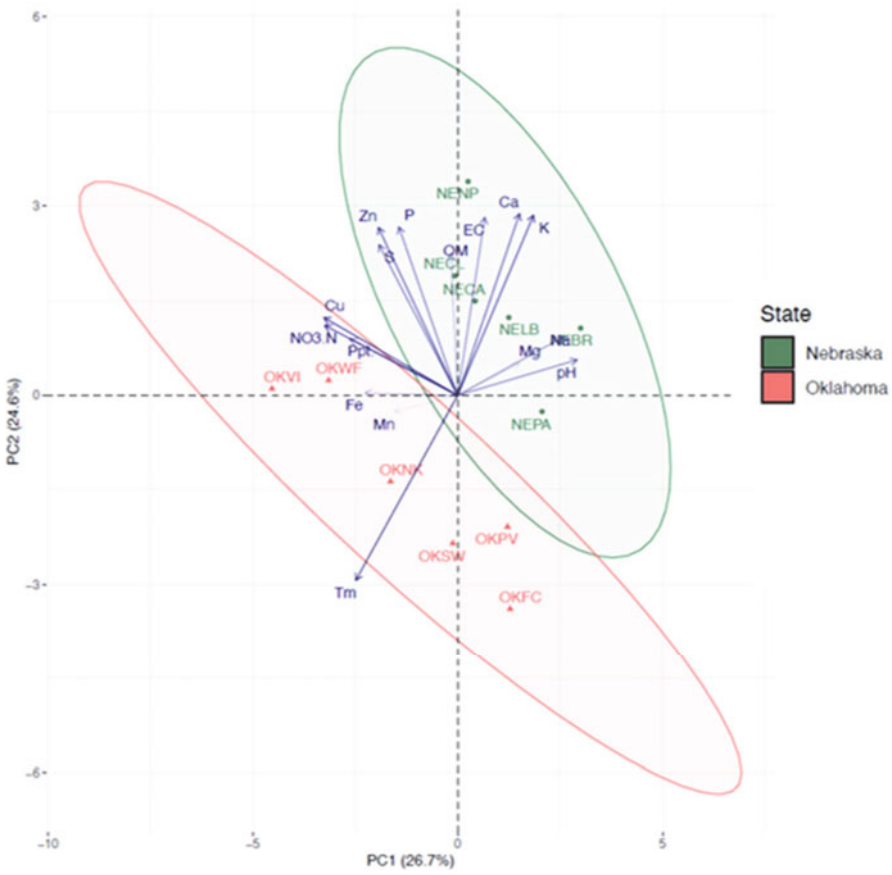

Fig. 2. Principal component analysis of the environmental variables in NE and OK sampling sites. Length of arrows indicates extent of variation, and darkness of arrows indicates strength of correlation.

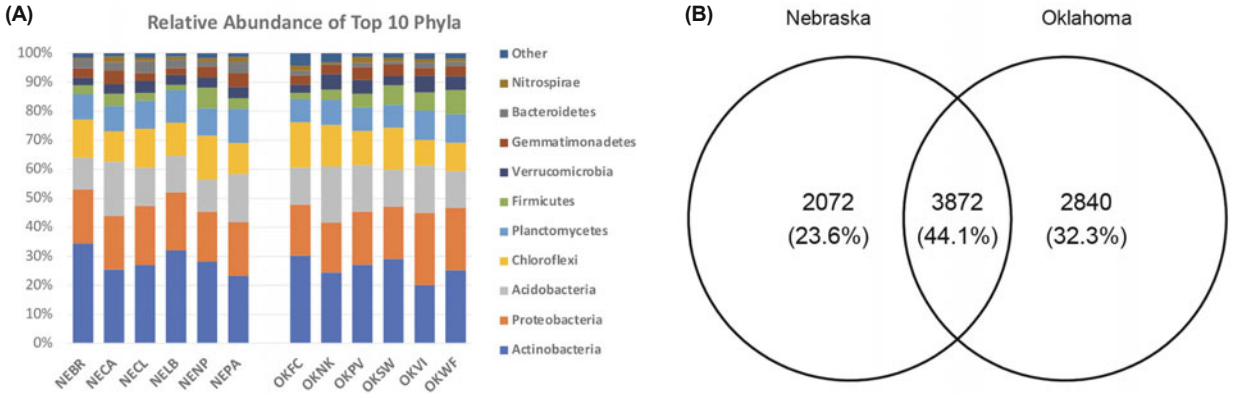

Fig. 3. Taxonomic composition of sampling sites in NE and OK. (A) Relative abundance of taxa at the phylum level (> 1% relative abundance, top 10 phyla). (B) Shared OTUs between NE and OK samples.

Fig. 4. Linear discriminant analysis (LDA) of NE and OK taxonomy. Each discriminating taxon represented in the plot is statistically significant with *P*-value < 0.05. Abbreviation: p, Phylum; c, Class; o, Order; f, Family; g, Genus; s, Species.

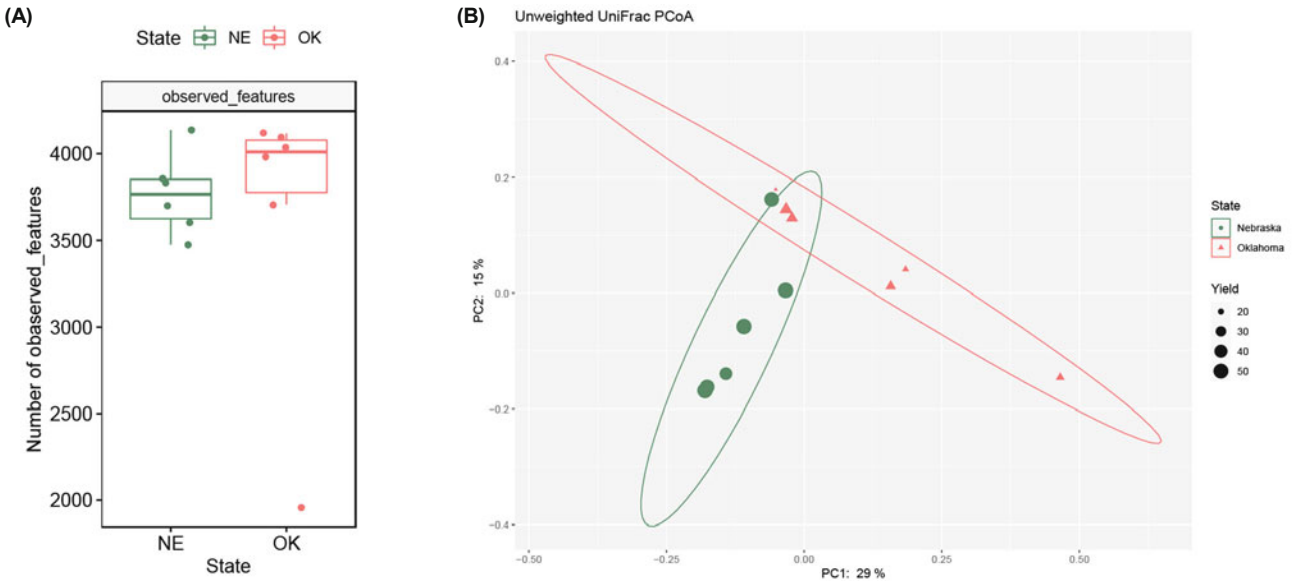

Fig. 5. Alpha and beta diversity analysis. (A) Alpha diversity of NE and OK samples showing the number of unique OTUs, and (B) beta diversity analysis in conjunction with soybean yield using Unweighted UniFrac distance matrix at the OTU level.

top 10 most abundant phyla (Fig. 3A). *Actinobacteria* (*Actinomycetota*) was the dominant phylum across all samples followed by *Proteobacteria* (*Pseudomonadota*), *Chloroflexi* (*Chloroflexota*), and *Acidobacteria* (*Acidobacteriota*) in both NE and OK soils. The difference between NE and OK samples was observed in *Fermicutes* and *Bacteroidetes* (*Bacteroidota*). The former was higher in OK (5.28 ± 0.02%) than NE (3.76 ± 0.02%) soils, whereas the letter was higher in NE (3.07 ± 0.01%) than OK (1.27 ± 0.01%) samples.

The linear discriminant analysis (LDA) of the bacterial taxonomy from NE and OK soil samples revealed that phylum *Bacteroidetes* (*Bacteroidota*) and class *Thermoleophilia* were the most abundant bacteria for NE and OK samples, respectively (Fig. 4). Interestingly, there were more diverse bacteria identified in NE compared to OK soils at the genus and species level (Fig. 4), although overall bacterial richness (i.e., alpha diversity) was greater in OK soils as described below (Fig. 5A). The difference could in part be attributed to the nature of the LDA analysis whose assumption requires normal distributions on samples.

Alpha and beta diversity of microbial communities

We performed alpha diversity analysis of the OTUs at a rarefaction depth of 27,342 reads per sample (Supplementary data Fig. S3). Alpha diversity using observed features (i.e., observed OTUs) showed higher median values in OK samples compared to NE, indicating that the bacterial richness is higher in OK than NE soils (Fig. 5A). However, there seems to be an outlier (i.e., OKNK) in alpha rarefaction (Supplementary data Fig. S3). Although we cannot fully explain why it shows the lowest richness, this could be related with acidity (pH 5.1) which is the lowest pH among all samples (Supplementary data Fig. S2). We also observed most abundant *Acidobacteria* (*Acidobacteriota*) in OKNK among all the OK samples (Fig. 3A). Beta diversity was calculated using Unweighted UniFrac distance matrix (Fig. 5B). We also added the soybean yield trait in the PCoA plot to visualize the relationship between the sampling sites and soybean yield. PERMANOVA test showed a difference in beta diversity of microbial communities at the OTU level between NE and OK (F-test = 2.2, $P = 0.0024$), in which distance-matrix values are 0.42 ± 0.01 and 0.49 ± 0.03 for NE and OK, respectively. This result indicates that there are higher similarities among NE samples compared to OK samples. In addition, using bioenv function, we observed that pH showed the strongest correlation

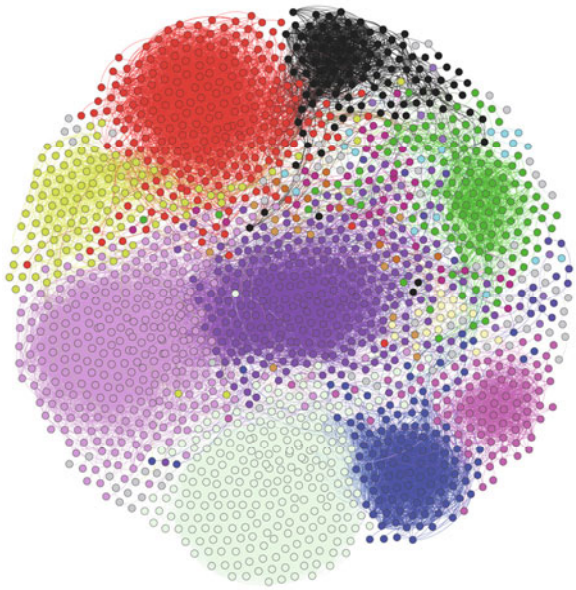

Fig. 6. Co-occurrence network of rhizosphere communities in NE and OK samples. Nodes represent OTUs and edges represent significant (P-value < 0.001) correlation between a pair of nodes. Networks are randomly colored by modules with at least 3 nodes.

Table 1. Topological features of rhizosphere microbiome network

Nodes	Edges	Modules (ME)	Modularity	Density	Diameter	Transitivity	Avg. path length
1,759	33,245	34	0.74	0.0215	12.267	0.754	4.295

(cor = 0.7277) with the beta diversity (data not shown), suggesting that pH is a key trait in shaping unique microbial communities between NE and OK soils. Furthermore, two-sided Mantel test also revealed strong correlation of sample distances with pH, also suggesting that pH is a strong predictor of the microbial community structure (Supplementary data Fig. S4A).

Our finding in alpha and beta diversity of the soybean rhizosphere microbiome in NE and OK was surprising since we expected that the higher soybean yield might be related with higher microbial diversity. In fact, we observed that the rhizosphere soil samples in the high yielding state NE showed lower alpha diversity, indicating low species richness within the samples. Thus, the higher soybean yield could be due to the combined effect of soil physicochemical properties and microbial communities, which inspired us to conduct the microbial co-occurrence network analysis in which all factors were considered rather than evaluating only microbial diversity.

Network statistics

After filtration of low abundant OTUs (i.e., < 0.1% abundance), we constructed a network containing 1,759 nodes and 33,245 connections for the microbial co-occurrence network analysis (Fig. 6). Topological properties of the network were calculated and clustered into separate modules (Table 1). The number of edges corresponding to positive correlation was 99.15% in the network, which mostly occurred among members of the same modules.

Modularity of the network

The network was clustered into 34 modules (ME1 through ME34) with modularity of 0.74, indicating the modular nature of the network as described by modularity value of > 0.4. (Newman and Girvan, 2004). Highly correlated modules (MEs with corr ≥ 0.75 and ≥ 3 nodes [members]) were selected and further analyzed, resulting in a total of 25 modules (Fig. 7A). These shortlisted 25 modules in the network contained 99.43% of total nodes (1,749 out of 1,759) and 99.97% of total edges (33,235 out of 33,245), suggesting that there was no significant loss in the number of nodes and their interactions after module filtering. Sample distances based on inferred microbial co-occurrence network modules were strongly correlated (rho = 0.681, $P = 0.0001$) with the Unweighted UniFrac distance (Supplementary data Fig. S4B), which could validate non-random co-occurrence of the microbial community.

Furthermore, we performed Spearman correlation analysis between sample traits and module eigengenes of each module (Fig. 7A). Modules ME1 ($P = 0.01$), ME9 ($P = 0.03$), ME13 ($P = 0.04$), ME23 ($P = 0.03$) and ME24 ($P = 0.01$) were significantly positively correlated with the soybean yield (Fig. 7A and Supplementary data Fig. S5). For instance, in addition to the soybean yield, ME1 (i.e., a box highlighted in Fig. 7A) showed positive correlation with pH ($P = 0.004$), potassium ($P = 0.05$) and sodium ($P = 0.03$), whereas negative correlation with Tm ($P = 0.04$). Interestingly, all the 5 MEs also revealed a trend of correlation with high potassium content, indicating that potassium is a key element for soybean growth in NE. This result is consistent with our PCA analysis

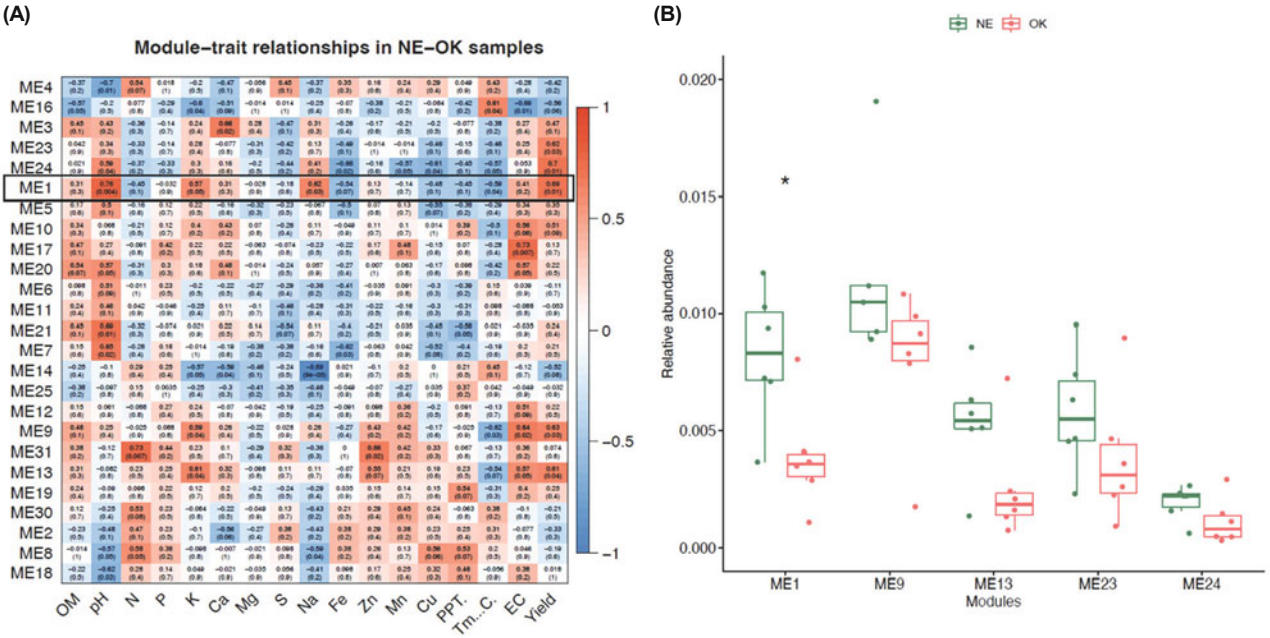

Fig. 7. Module-trait relationships in NE and OK samples. (A) Spearman's correlation of modules in the network to the sample traits. A box highlights all traits in ME1. (B) Relative abundances of soybean yield associated module members in NE and OK samples (* indicates *P*-value < 0.05, Wilcox test).

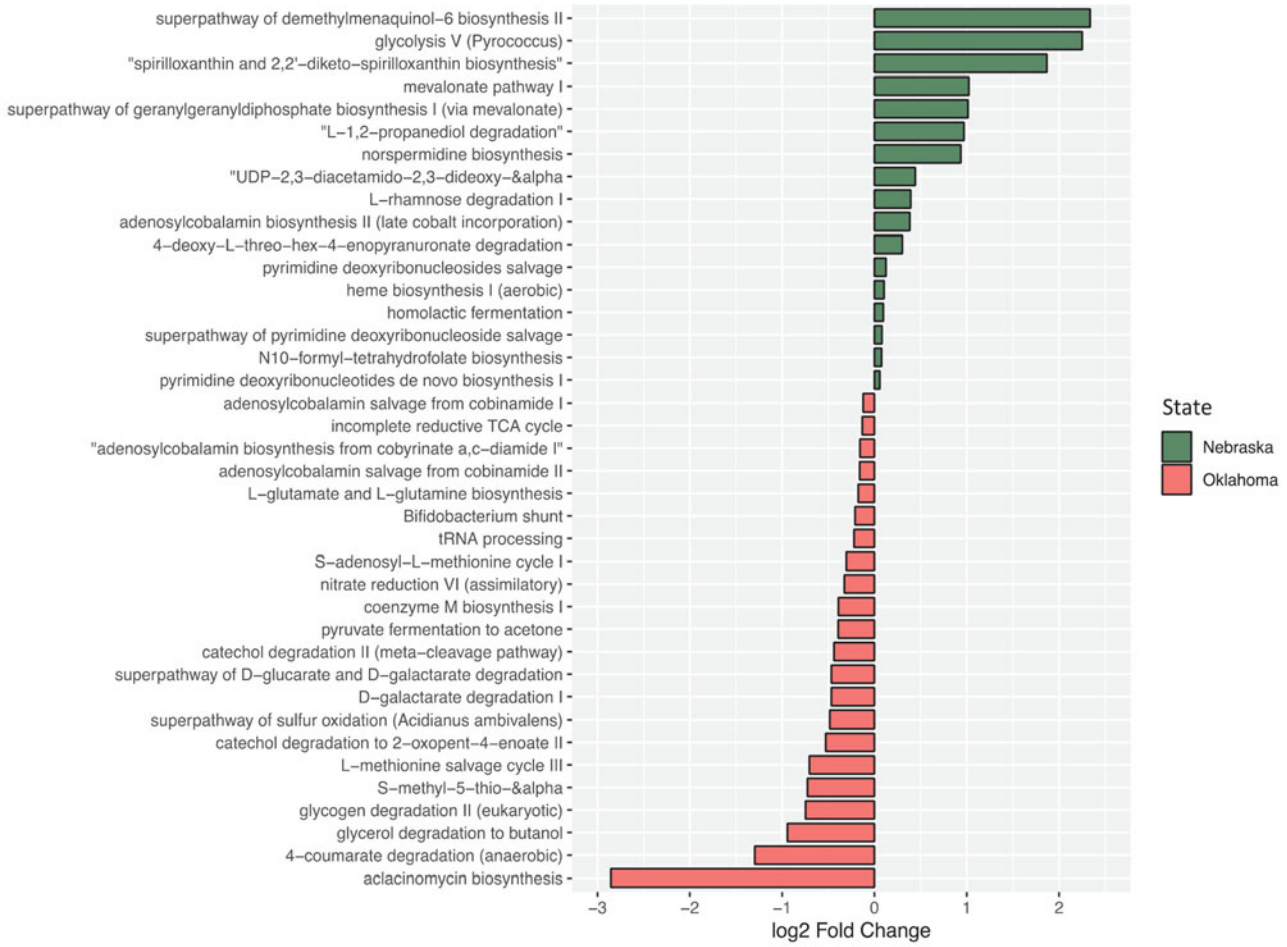

Fig. 8. Differential analysis of predicted METACYC pathways using PICRUSt2 predicted function analysis. Differential functions are statistically significant ($P < 0.05$) with LDA score > 2.

showing that potassium is the more important nutrient than nitrogen and phosphorus for the higher soybean yield in NE compared to OK (Fig. 2).

Microbial composition in the network

Overall, the network nodes were dominated by the following 10 phyla: *Actinobacteria* (19.81%), *Proteobacteria* (17.36%), *Acidobacteria* (16.68%), *Chloroflexi* (15.03%), *Planctomycetes* (13.03%), *Verrucomicrobia* (4.55%), *Gemmatimonadetes* (4.21%), *Firmicutes* (3.24%), *Bacteroidetes* (2.79%), and *Nitrospirae* (1.31%). We examined the relative abundance of soybean yield-associated module members in NE and OK samples. There was a trend for higher abundance of members of all 5 modules (ME1, ME9, ME13, ME23, and ME24) in NE compared to OK samples, although the statistically significant difference was observed only in ME1 ($P < 0.05$) (Fig. 7B).

To identify the most influential variable (e.g., bacteria) at all taxonomic levels, we revealed relationships of module members (i.e., nodes) in each module (Supplementary data Fig. S5). For instance, in the most complex module ME9, several key influencers such as class *Anaerolineae*, family *Micromonosporaceae*, genus *Planctomyces*, and genus *Nitrospira* were identified. Although little is known about an agricultural impact of *Anaerolneae* on crop production, genera *Planctomyces* and *Nitrospira* are known to be involved in the nitrogen cycle via anammox and nitrification (Strous *et al.*, 1999; Mehrani *et al.*, 2020). It has been also reported that application of *Micromonospora chalcea* UAE1 enhanced the growth of dwarf glasswort under gnotobiotic and greenhouse conditions (El-Tarabily *et al.*, 2019). Interestingly, genus *Rhizobium* is the key bacteria in ME23, indicating that nitrogen fixation derived by rhizobia would induce higher soybean yield.

Predicted functional profiles of microbial communities

We assessed potential functional differences in microbial communities between NE and OK rhizosphere soils using PICRUSt 2.0 (Fig. 8). The differential analysis of predicted METACYC pathways revealed, for NE soils, a high occurrence of pathways relating to cell membrane production (demethylmenaquinol-6 biosynthesis and mevalonate pathway), photosynthetic related pathways (spirilloxanthin, 2,2′-diketo-spirilloxanthin, and geranylgeranyl diphosphate biosynthesis),

metabolism (glycolysis V, UDP, and L-1,2-propanediol degradation), and biofilm formation (norspermidine biosynthesis).

The most dominant pathway being utilized in the OK soils involves the biosynthesis of aclacinomycin, an antibiotic produced by *Streptomyces*. Other pathways highlighted in the OK soils involve harvesting carbon sources (glycerol, glycogen II, D-galactarate, and D-glucarate degradation), aromatics degradation (catechol degradation), and degradation of toxic or inhibitory compounds (4-coumarate, S-methyl-5-thioadenosine, and L-methionine salvage cycle III). A sulfur oxidation pathway was also identified in the OK soils.

Discussion

Rhizosphere microbial communities are known to have complex interactions governed by soil physicochemical properties, host-associated chemicals, and other factors at multiple domains. In this study, a comparative analysis was performed to identify spatial variations in community structure at taxonomic and predicted functional levels in the soybean rhizosphere that could potentially define differences in soybean yield. Additionally, we performed correlation network-based analysis using bacterial OTUs to explore the co-occurrence patterns of microorganisms and their association with soybean yield.

Assembly of rhizosphere microbiota occurs in a series of events beginning from shaping microbial communities by physicochemical properties of soil and biogeochemical processes, followed by the exposure of plant roots to the indigenous microbiota and finally the selection of microbes in response to root exudates (Philippot *et al.*, 2013; Sugiyama, 2019). Our PCA analysis showed spatial differences in terms of their soil physicochemical properties and environmental factors (Fig. 2). Among all environmental variables tested, potassium, calcium, sodium, electrical conductivity, atmospheric temperature, and soybean yield have shown statistically significant difference between NE and OK soils (Supplementary data Fig. S2). Interestingly, among three major macronutrients (N-P-K), potassium is the more crucial nutrient than nitrogen and phosphorus for the higher soybean yield in NE soils. The physicochemical properties of the soil is known to alter rhizosphere microbiomes (Sugiyama, 2019). To identify those differences, we used UniFrac distance matrix, which can capture the differences in communities based on their evolutionary history (Lozupone *et al.*, 2011). Significant difference (P-value < 0.03) in beta diversity at spatial scale (Fig. 5B) indicates the presence of phylogenetically divergent communities in the rhizosphere in these two localities, where OK samples contain more phylogenetically diverse communities as estimated by Faiths-pd index (Supplementary data Fig. S3B). Higher phylogenetic diversity in OK samples may indicate its complementary nature to adapt under less nutrient and more stressful environmental conditions (Lankau *et al.*, 2022). In addition, the diversity could also be attributed to the differences in soil physicochemical properties, nutrient availability (Lozupone *et al.*, 2007), and colonization of selected microbial communities in the rhizosphere from near bulk soil in response to host root exudates (Sugiyama, 2019). Beta diversity of these samples is strongly correlated with pH, which is known to be the most important driver of the microbial community in soil (Zhang *et al.*, 2018). Although differences in pH are not statistically significant between the two sampling groups (Supplementary data Fig. S2), it is found to be the most important variable to shape the microbial community structure.

Taxonomic classification of OTUs showed common bacterial phyla of the soybean rhizosphere (e.g., *Actinobacteria* [*Actinomycetota*], *Proteobacteria* [*Pseudomonadota*], *Chloroflexi* [*Chloroflexota*], *Acidobacteria* [*Acidobacteriota*], *Planctomycetes* [*Planctomycetota*], *Gemmatimonadetes* [*Gemmatimonadota*], *Firmicutes* [*Bacillota*], *Verrucomicrobia* [*Verrucomicrobiota*], *Nitrospirae* [*Nitrospirota*], and *Bacteroidetes* [*Bacteroidota*]) as shown in Fig. 3A (Mendes *et al.*, 2014; Tao *et al.*, 2018; Zhang *et al.*, 2018). Relative abundances of OTUs were used for the taxonomic classification, since the total abundances of OTUs at the phylum varied among the samples, ranging from 41,732 to 98,352 OTUs with the average value of 63,678 ± 6,537 OTUs. For the total bacterial abundances per each sample, it would be valuable to quantify the absolute bacterial abundance using more precise methods such as quantitative PCR or digital PCR. LEFSe (LDA effect size) analysis resulted in a number of discriminating biomarkers in NE and OK rhizosphere samples. Among those, phylum *Bacteroidetes* (*Bacteroidota*) is discriminately abundant in NE samples (Fig. 4). Members of the phylum are anaerobes (Woese, 1987) and tend to be more dominant in soil with higher carbon availability (Liang *et al.*, 2014). Indeed, there is a trend of higher organic matter in NE soils compared to the counterpart (Supplementary data Fig. S2). In addition, we observed higher abundance of beneficial plant growth-promoting bacteria in NE soils, such as *Agromyces, Microlunatus, Flavobacterium*, which indicates their importance in crop productivity besides the soybean endosymbiont *Bradyrhizobium* species (Bokukhatem *et al.*, 2022). *Flavobacterium* is a well-known plant growth-promoting bacterium that degrades complex organic matter and have antagonistic effect with the plant pathogens (Sang and Kim, 2012; Kraut-Cohen *et al.*, 2021). Similarly, *Lysobacter* species are found to promote plant growth by producing metabolites against bacterial and fungal pathogens in the plant root (Expósito *et al.*, 2015). *Agromyces* is found to enhance plant growth by mobilizing Zn and Cd (Kuffner *et al.*, 2008). Interestingly, discriminately abundant taxa in NE samples such as f_*Pirellulaceae*, f_*Chitonophagaceae*, o_*Bacteroidales* and g_*Afifella* (Fig. 4) are also found in modules that are significantly positively correlated with soybean yield (Supplementary data Fig. S5), suggesting that these taxa may contribute to higher soybean yield either directly or indirectly. In addition to plant growth-promoting bacteria, abundance of N_2O reducing bacteria such as *Gemmatimonadaceae* (Park *et al.*, 2017) in NE soil indicates part of biogeochemical nitrogen cycle for mitigating greenhouse gas emissions (Fig. 4).

Previous studies have found that spatial variability causes significant effects on the composition of microbial communities in the terrestrial ecosystem (Fierer and Jackson, 2006) and also in co-occurrence patterns in the soybean rhizosphere (Zhang *et al.*, 2018). Co-occurring microbial communities can be clustered into densely connected smaller subunits defined as modules that can then be visualized as niche differentiation

and habitat filtering (Röttjers and Faust, 2018). These small-word networks favor a quick response to perturbations and information and resource sharing (Faust and Raes, 2012) in addition to niche filtering (Röttjers and Faust, 2018). Module-based network analyses have the potential to minimize the effect of inter-kingdom interactions, which is more profound in the rhizosphere, thereby providing precise effects of environmental factors on the bacterial communities. For example, the network analysis by de Menezes *et al.* (2015) showed that the bacterial and fungal modules correlated independently with the edaphic factors. In addition to being robust to aforementioned effects, modules preserve community differences based on individual species level as observed by beta diversity correlation in our study (Supplementary data Fig. S4B). This validates our module-based approach in order to identify key drivers of sample traits including the soybean yield at sub-network levels. Modules in the global network showed significant correlations with the soybean yield and edaphic factors such as potassium and pH. The members of positively correlated modules with the soybean yield showed higher abundance in NE samples, suggesting their agricultural impact on increased soybean production.

For prediction of functional abundances in the NE soils, we found an increased occurrence of pathways related to bacterial growth and cellular membrane production via isoprenoid quinone production which have been previously identified in other rhizosphere studies (Collins and Jones, 1981; Nowicka and Kruk, 2010; Jin *et al.*, 2016; Pinda *et al.*, 2016; Jiménez *et al.*, 2020). Carotenoids are also produced in these soils to help facilitate the function of photosystems and combat oxidative stress (Liu *et al.*, 2013; Chi *et al.*, 2015; Ding *et al.*, 2019). In addition to the pathways focused on the production of cellular membranes, other pathways were also identified that are involved in metabolisms and carbon utilization for growth and function (Naik *et al.*, 2009; Chowdhury *et al.*, 2014; Staib and Fuchs, 2015; Tao *et al.*, 2021). Finally, biofilm synthesis or inhibition was emphasized in these soils, which could play an important role in controlling the bacterial communities near the rhizosphere region of the plant roots (Hobley *et al.*, 2014; Wotanis *et al.*, 2017; Becerra-Rivera and Dunn, 2019; Kiliny and Nehela, 2020).

For the OK soils, there is a trend for lower availability of nutritional resources as compared to the NE soils (Supplementary data Fig. S2). Microorganisms in the OK rhizosphere have to compete more vigorously with one another to gain a foothold. The predominant method of staving off the competition involves the production of antibiotics and toxin/inhibitor degradation. The pathway related to the antibiotic aclacinomycin was emphasized in all the OK soils (Fig. 8). This antibiotic is typically produced by *Streptomyces*, which was found uniquely in these soils, and could be effective in eliminating competitive microbial populations in the rhizosphere (Kim *et al.*, 1996; Wang *et al.*, 2019). Pathways involved in the degradation of toxins created by plants, protein degradation, and destruction of protein degradation inhibitors were also emphasized in the OK soils (Sekowska *et al.*, 2004; Parveen and Cornell, 2011; Timm *et al.*, 2015). Overall, the microbial populations in the NE rhizosphere have more nutrients available and live within a moderate pH and temperature range, allowing them to focus more on metabolic pathways that have to deal with growth and proliferation, while the populations in the OK soils, on the other hand, are living within more hostile conditions, which may force them to spend their energy for the purpose of survival such as combatting against toxins or antibiotics.

Conclusion

In summary, our study investigated differences in soybean rhizosphere microbiomes between a high-yield (NE) and a low-yield state (OK). The results showed that co-occurring patterns of microbial communities differed between NE and OK rhizospheres. Spatial variation in community composition was observed at physicochemical, taxonomic, and predicted functional levels. Potassium is a key nutrient for the higher soybean yield in NE compared to OK soils, while pH is the most important edaphic factor for shaping the microbial community structure. Furthermore, not only do our results expand our limited knowledge on interactions of microbial communities in the soybean rhizosphere, but they also provide a basis to formulate a hypothesis for the development of a bioinoculant (e.g., microbial consortium) to enhance soybean production in the U.S. as well as worldwide.

Acknowledgements

We thank the Genomic Sequencing and Analysis Facility (GSAF) at the University of Texas at Austin for sequencing.

Conflict of Interest

The authors have no conflict of interest to report.

Data Availability

The complete 16S rRNA gene sequencing data in this study are available in the National Center for Biotechnology Information (NCBI) database under the BioProject accession number PRJNA873129 associated with the accession numbers of 12 BioSamples (SAMN30489248 - SAMN30489259).

References

Barberán, A., Bates, S.T., Casamayor, E.O., and Fierer, N. 2012. Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J.* **6**, 343–351.

Bastian, M., Heymann, S., and Jacomy, M. 2009. Gephi : an open source software for exploring and manipulating networks. *Proc. Int. AAAI Conf. Web Soc. Media* **3**, 361–362. Retrieved from https://ojs.aaai.org/index.php/ICWSM/article/view/13937.

Becerra-Rivera, V.A. and Dunn, M.F. 2019. Polyamine biosynthesis and biological roles in rhizobia. *FEMS Microbiol. Rev.* **366**, fnz084.

Berry, D. and Widder, S. 2014. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front. Microbiol.* **5**, 219.

Bokulich, N.A., Kaehler, B.D., Rideout, J.R., Dillon, M., Bolyen, E.,

Knight, R., Huttley, G.A., and Caporaso, J.G. 2018. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* **6**, 90.

Bolyen, E., Rideout, J.M., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al Ghalith, A.G., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., *et al.* 2018. QIIME 2 : reproducible , interactive , scalable, and extensible microbiome data science. *Nat. Biotechnol.* **37**, 852–857.

Boukhatem, Z.F., Merabet, C., and Tsaki, H. 2022. Plant growht promoting actinobacteria, the most promising candidates as bio-inoculants? *Front. Agron*. **4**, 849911.

Cao Y. 2022. Package 'microbiomeMarker': microbiome biomarker analysis toolkit. R package version 1.2.2. https://github.com/yiluheihei/microbiomeMarker.

Chi, S.C., Mothersole, D.J., Dilbeck, P., Niedzwiedzki, D.M., Zhang, H., Qian, P., Vasilev, C., Grayson, K.J., Jackson, P.J., Martin, E.C., *et al.* 2015. Assembly of functional photosystem complexes in *Rhodobacter sphaeroides* incorporating carotenoids from the spirilloxanthin pathway. *Biochim. Biophys. Acta* **1847**, 189–201.

Chowdhury, C., Sinha, S., Chun, S., Yeates, T.O., and Bobik, T.A. 2014. Diverse bacterial microcompartment organelles. *Microbiol. Mol. Biol. Rev*. **78**, 438–468.

Clarke, K.R. and Ainsworth, M. 1993. A method of linking multivariate community structure to environmental variables. *Mar. Ecol. Prog. Ser.* **92**, 205–219.

Collins, M.D. and Jones, D. 1981. Distribution of isoprenoid quinone structural types in bacteria and their taxonomic implications. *Microbiol. Rev*. **45**, 316–354.

Csárdi, G. and Nepusz, T. 2006. The igraph software package for complex network research. *InterJ. Complex Syst.* **1695**, 1–9.

Cram, J.A., Xia L.C., Needham D.M., Sachdeva R., Sun F., and Fuhrman J.A. 2015. Cross-depth analysis of marine bacterial networks suggests downward propagation of temporal changes. *ISME J.* **9**, 2573–2586.

de Menezes, A.B., Prendergast-Miller, M.T., Richardson, A.E., Toscas, P., Farrell, M., Macdonald, L.M., Baker, G., Wark, T., and Thrall, P.H. 2015. Network analysis reveals that bacteria and fungi form modules that correlate independently with soil parameters. *Environ. Microbiol.* **17**, 2677–2689.

Ding, B., Niu, J., Shang, F., Yang, L., Chang, T., and Wang, J. 2019. Characterization of the geranylgeranyl diphosphate synthase gene in *Acyrthosiphon pisum* (Hemiptera: Aphididae) and its association with carotenoid biosynthesis. *Front. Physiol.* **10**, 1398.

Douglas, G.M., Maffei, V.J., Zaneveld, J., Yurgel, S.N., Brown, J.R., Taylor, C.M., Huttenhower, C., Langille, M.G.I. 2020. PICRUSt2: an improved and extensible approach for metagenome inference. *Nat. Biotechnol.* **38**, 685–688.

El-Tarabily, K., AlKhajeh, A.S., Ayyash, M.M., Alnuaimi, L.H., Sham, A., ElBaghdady, K.Z., Tariq, S., and AbuQamar, S.F. 2019. Growth promotion of *Salicornia bigelovii* by *Micromonospora chalcea* UAE1, and endophytic 1-aminocyclopropane-1-carboxylic acid deaminase-producing actinobacterial isolate. *Front. Microbiol.* **10**, 1694.

Expósito, R.G., Postma, J., Raaijmakers, J.M., and Bruijn, I.D. 2015. Diversity and activity of *Lysobacter* species from disease suppressive soils. *Front. Microbiol.* **6**, 1243.

Faust, K. and Raes, J. 2012. Microbial interactions: from networks to models. *Nat. Rev. Microbiol.* **10**, 538–550.

Fierer, N. and Jackson, R.B. 2006. The diversity and biogeography of soil bacterial communities. *Proc. Natl. Acad. Sci. USA* **103**, 626–631.

Hartmann, A., Rothballer, M., and Schmid, M. 2008. Lorenz Hiltner; a pioneer in rhizosphere microbial ecology & soil bacteriology research. *Plant Soil* **312**, 7–14.

Herridge, D.F., Peoples, M.B., and Boddey, R.M. 2008. Global inputs of biological nitrogen fixation in agricultural systems. *Plant Soil* **311**, 1–18.

Hobley, L., Kim, S.H., Maezato, Y., Wyllie, S., Fairlamb, A.H., Stanley-Wall, N.R., and Michael, A.J. 2014. Norspermidine is not a self-produced trigger for biofilm disassembly. *Cell* **156**, 844–854.

Jiménez, J.A., Novinscak, A., and Filion, M. 2020. Inoculation with plant-growth-promoting rhizobacterium *Pseudomonas fluorescens* LBUM677 impacts the rhizosphere microbiome of three oilseed crops. *Front. Microbiol.* **11**, 569366.

Jin, D., Kong, X., Li, H., Luo, L., Zhuang, X., Zhuang, G., Deng, Y., and Bai, Z. 2016. *Patulibacter brassicae* sp. nov., isolated from rhizosphere soil of chinese cabbage (*Brassica campestris*). *Int. J. Syst. Evol. Microbiol.* **66**, 5056–5060.

Jin, J., Wang, G.H., Liu, X.B., Liu, J.D., Chen, X.L., and Herbert, S.J. 2009. Temporal and spatial dynamics of bacterial community in the rhizosphere of soybean genotypes grown in a black soil. *Pedosphere* **19**, 808–816.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. 2002. MAFFT : a novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic Acids Res.* **30**, 3059–3066.

Killiny, N. and Nehela, Y. 2020. Citrus polyamines: structure, biosynthesis, and physiological functions. *Plants* **9**, 426.

Kim, H.S., Kim, Y.H., Yoo, O.J., and Lee, J.J. 1996. Aclacinomycin X, a novel anthracycline antibiotic produced by *Streptomyces galilaeus* ATCC 31133. *Biosci. Biotech. Biochem*. **60**, 906–908.

Kraut-Cohen, J., Shapiro, O.H., Dror, B., and Cytryn, E. 2021. Pectin induced colony expansion of soil-derived *Flavobacterium* strains. *Front. Microbiol.* **12**, 651891.

Kuffner, M., Puschenreiter, M., Wieshammer, G., Gorfer, M., and Sessitsch, A. 2008. Rhizosphere bacteria affect growth and metal uptake of heavy metal accumulating willows. *Plant Soil* **304**, 35–44.

Langfelder, P. and Horvath, S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559.

Langille, M.G.I., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A., Clemente, J.C., Burkepile, D.E., Thurber, R.L.V., Knight, R., *et al.* 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* **31**, 814–821.

Lankau, R.A., George, I., and Miao, M. 2022. Crop performance is predicted by soil microbial diversity across phylogenetic scales. *Ecosphere* **13**, e4029.

Liang, J., Sun, S., Ji, J., Wu, H., Meng, F., Zhang, M., Zheng, X., Wu, C., and Zhang, Z. 2014. Comparison of the rhizosphere bacterial communities of zigongdongdou soybean and a high-methionine transgenic line of this cultivar. *PLoS ONE* **9**, e103343.

Liu, C., Sun, Z., Shen, S., Lin, L., Li, T., Tian, B., and Hua, Y. 2013. Identification and characterization of the geranylgeranyl diphosphate synthase in *Deinococcus radiodurans*. *Lett. Appl. Microbiol.* **58**, 219–224.

Lozupone, C.A., Hamady, M., Kelley, S.T., and Knight, R. 2007. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Appl. Environ. Microbiol.* **73**, 1576–1585.

Lozupone, C., Lladser, M., Knights, D., Stombaugh, J., and Knight, R. 2011. UniFrac: an effective distance metric for microbial community comparison. *ISME J.* **5**, 169–172.

Luster, J., Göttlein, A., Nowack, B., and Sarret, G. 2009. Sampling, defining, characterising and modeling the rhizosphere-the soil science tool box. *Plant Soil* **321**, 457–482.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**, 209–220.

Mcdonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., Andersen, G.L., Knight, R., and Hugenholtz, P. 2012. An improved greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* **6**, 610–618.

Mehrani, M., Sobotka, D., Kowal, P., Ciesielski, S., and Makinia, J. 2020. The occurrence and role of *Nitrospiria* in nitrogen removal systems. *Bioresour. Technol.* **303**, 122936.

Mendes, L.W., Kuramae, E.E., Navarrete, A.A., van Veen, J.A., and Tsai, S.M. 2014. Taxonomical and functional microbial community selection in soybean rhizosphere. *ISME J.* **8**, 1577–1587.

Meng, J., Xu,Y., Li, S., Li, C., Zhang, X., Dong, D., and Chen, P. 2010. Soybean growth and soil microbial populations under conventional and conservational tillage systems. *J. Crop Improv.* **24**, 337–348.

Mizrahi-Man, O., Davenport, E.R., and Gilad, Y. 2013. Taxonomic classification of bacterial 16S rRNA genes using short sequencing reads: evaluation of effective study designs. *PLoS ONE* **8**, e53608.

Naik, D., Smith, E., and Cumming, J.R. 2009. Rhizosphere carbon deposition, oxidative stress and nutritional changes in two poplar species exposed to aluminum. *Tree Physiol.* **29**, 423–436.

Naim, M.S. 1965. Development of rhizosphere and rhizoplane microflora of *Aristida coerulescens* in the Libyan desert. *Archiv. Mikrobiol.* **50**, 321–325.

Newman, M.E.J. and Girvan, M. 2004. Finding and evaluating community structure in networks. *Phys. Rev. E* **69**, 026113.

Niraula, S., Choi, Y.K., Payne, K., Muir, J.P., Kan, E., and Chang, W.S. 2021. Dairy effluent-saturated biochar alters microbial communities and enhances bermudagrass growth and soil fertility. *Agronomy* **11**, 1794.

Nowicka, B. and Kruk, J. 2010. Occurrence, biosynthesis and function of isoprenoid quinones. *Biochim. Biophys. Acta* **1797**, 1587–1605.

Park, D., Kim, H., and Yoon, S. 2017. Nitrous oxide reduction by an obligate aerobic bacterium, *Gemmatimonas aurantiaca* strain T-27. *Appl. Environ. Microbiol.* **83**, e00502-17.

Parveen, N. and Cornell, K.A. 2011. Methylthioadenosine/S-adenosylhomocystein nucleosidase, a critical enzyme for bacterial metabolism. *Mol. Microbiol.* **79**, 7–20.

Philippot, L., Raaijmakers, J.M., Lemanceau, P., and van der Putten, W.H. 2013. Going back to the roots: the microbial ecology of the rhizosphere. *Nat. Rev. Microbiol.* **11**, 789–799.

Pinda, E.S., Silva, D.B., Teixeira, S.P., Coppede, J.S., Furlan, M., França, S.C., Lopes, N.P., Pereira, A.M.S., and Lopes, A.A. 2016. Mevalonate-derived quinonemethide triterpenoid from *in vitro* roots of *Peritassa laevigata* and their localization in root tissue by MALDI imaging. *Sci. Rep.* **6**, 22627.

Price, M.N., Dehal, P.S., and Arkin, A.P. 2010. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490.

Rascovan, N., Carbonetto, B., Perrig, D., Díaz, M., Canciani, W., Abalo, M., Alloati, J., González-Anta, G., and Vazquez, M.P. 2016. Integrated analysis of root microbiomes of soybean and wheat from agricultural fields. *Sci. Rep.* **6**, 28084.

Röttjers, L. and Faust, K. 2018. From hairballs to hypotheses–biological insights from microbial networks. *FEMS Microbiol. Rev.* **42**, 761–780.

Salvagiotti, F., Cassman, K.G., Specht, J.E., Walters, D.T., Weiss, A., and Dobermann, A. 2008. Nitrogen uptake, fixation and response to fertilizer N in soybeans: a review. *Field Crops Res.* **108**, 1–13.

Sang, M.K. and Kim, K.D. 2012. The volatile-producing *Flavobacterium johnsoniae* strain GDE09 shows biocontrol activity against *Phytopthora capsici* in pepper. *J. Appl. Microbiol.* **113**, 383–398.

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., and Huttenhower, C. 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods* **9**, 811–814.

Sekowska, A., Dénervaud, V., Ashida, H., Michoud, K., Haas, D., Yokota, A., and Danchin, A. 2004. Bacterial variations on the methionine salvage pathway. *BMC Microbiol.* **4**, 9.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504.

Spearman, C. 1904. The proof and measurement of association between two things. *Am. J. Psychol.* **15**, 72–101.

Staib, L. and Fuchs, T.M. 2015. Regulation of fucose and 1,2-propanedioll utilization by *Salmonella enterica* serovar Typhimurium. *Front. Microbiol.* **6**, 1116.

Strous, M., Fuerst, J.A., Kramer, E.H.M., Logemann, S., Muyzer, G., van de Pas-Schoonen, K.T., Webb, R., Kuenen, J.G., and Jetten, M.S.M. 1999. Missing lithotroph identified as new planctomycete. *Nature* **400**, 446–449.

Sugiyama, A. 2019. The soybean rhizosphere: metabolites, microbes, and beyond–a review. *J. Adv. Res.* **19**, 67–73.

Sugiyama, A., Ueda, Y., Zushi, T., Takase, H., and Yazaki, K. 2014. Changes in the bacterial community of soybean rhizospheres during growth in the field. *PLoS ONE* **9**, e100709.

Tao, Y., Bu, C., Zou, L., Hu, Y., Zheng, Z., and Ouyang, J. 2021. A comprehensive review on microbial production of 1,2-proanediol: micro-organisms, metabolic pathways, and metabolic engineering. *Biotechnol. Biofuels* **14**, 216.

Tao, J., Meng, D., Qin, C., Liu, X., Liang, Y., Xiao, Y., Liu, Z., Gu, Y., Li, J., and Yin, H. 2018. Integrated network analysis reveals the importance of microbial interactions for maize growth. *Appl. Microbiol. Biotechnol.* **102**, 3805–3818.

Timm, C.M., Campbell, A.G., Utturkar, S.M., Jun, S., Parales, R.E., Tan, W.A., Robeson, M.S., Lu, T.S., Jawdy, S., Brown, S.D., *et al.* 2015. Metabolic functions of *Pseudomonas fluorescens* strains from *Populus deltoids* depend on rhizosphere or endosphere isolation compartment. *Front. Microbiol.* **6**, 1118.

Wang, R., Zhang, S., Ye, Y., Yu, Z., Qi, H., Zhang, H., Xue, Z., Wang, J., and Wu, M. 2019. Three new isoflavonoid glycosides from the mangrove-derived actinomycete *Micromonospora aurantiaca* 110B. *Mar. Drugs* **17**, 294.

Woese, C.R. 1987. Bacterial evolution. *Microbiol. Rev.* **51**, 221–271.

Wotanis, C.K., Brennan, W.P.3rd, Angotti, A.D., Villa, E.A., Zayner, J.P., Mozina, A.N., Rutkovsky, A.C., Sobe, R.C., Bond, W.G., and Karatan, E. 2017. Relative contributions of norspermidine synthesis and signaling pathways to the regulation of *Vibrio cholera* biofilm formation. *PLoS ONE* **12**, e0186291.

Zhang, B., Zhang, J., Liu, Y., Shi, P., and Wei, G. 2018. Co-occurrence patterns of soybean rhizosphere microbiome at a continental scale. *Soil Biol. Biochem.* **118**, 178–186.