Host Plants, Not Soils, Shape the Phylogenetic Assembly of Culturable Legume-Associated Bacteria

María A. Pérez-Fernández

maperfer@upo.es

Universidad Pablo de Olavide

Nonkululeko Sithole

University of KwaZulu-Natal (Westville Campus)

Anathi Magadlela

University of KwaZulu-Natal (Westville Campus)

Article

Keywords: Legume microbiome, soil bacterial diversity, phylogenetic filtering, host–microbe interactions, nodule-associated bacteria

Posted Date: July 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10163834/v1

Additional Declarations: No competing interests reported.

Abstract

Understanding how environmental and host factors shape plant microbiomes is essential for predicting plant–microbe interactions. We investigated the influence of soil origin and legume host identity on bacterial communities associated with cowpea (*Vigna unguiculata*), groundnut (*Arachis hypogaea*), and mungbean (*Vigna radiata*) grown in five contrasting soils. Seventy-seven bacterial isolates were characterised by 16S rRNA gene sequencing and Maximum Likelihood phylogenetic analysis. Diversity was assessed using richness, Shannon diversity, rarefaction analyses, Bray–Curtis-based PCoA, and unweighted UniFrac distances. Soil-derived communities showed broad phylogenetic dispersion and high variability in composition and richness, reflecting their role as reservoirs of microbial diversity. In contrast, plant-associated communities displayed host-specific clustering and lower variability. Groundnut-associated communities were more distinct than those of cowpea and mungbean. Soil defines the available bacterial pool, while legumes selectively recruit and structure subsets of this diversity, highlighting host-driven filtering in microbiome assembly.

Introduction

Legume–microbe interactions are fundamental components of terrestrial ecosystems and global biogeochemical cycles. Leguminous plants are unique in their ability to establish symbiotic associations with diverse soil bacteria, many of which enhance plant growth through biological nitrogen fixation, phytohormone production, pathogen suppression, or nutrient mobilization [1, 2]. These interactions have traditionally been studied through the lens of the *Rhizobium*–legume symbiosis, yet it is increasingly clear that the broader community of plant-associated microbes also contributes significantly to plant health, ecological fitness, and nutrient dynamics [3, 4]. Understanding how legume hosts recruit and structure their associated microbiota is therefore essential for unravelling the ecological and evolutionary forces shaping these symbioses.

Soils act as the primary reservoir for plant-associated microbes, providing a diverse pool of potential colonizers whose taxonomic composition is strongly influenced by edaphic factors such as pH, organic matter, moisture, and mineralogy [5, 6]. Yet, despite the complexity and variability of soil microbial communities, plants do not associate with these communities randomly. Instead, a growing body of evidence indicates that plant hosts exert strong selective pressures that determine which microbial lineages become enriched in the rhizosphere, endosphere, and nodules [7, 8]. This selection is mediated through root exudates, plant immune responses, and specific recognition pathways, many of which display evolutionary conservation across plant lineages [9]. In legumes, for example, molecular mechanisms governing compatibility—such as Nod factor recognition and exopolysaccharide signalling—are highly specific and may constrain the range of microbial partners that a plant can support [10, 11].

An important unresolved question concerns the relative contribution of the soil versus the plant host in determining the phylogenetic structure of bacterial communities associated with legumes. Some studies argue that soil properties overwhelmingly shape community composition, with the plant host exerting

only modest influence [12, 13, 14]. Others, however, have demonstrated strong host-driven filtering, whereby plants grown in different soils converge toward similar bacterial assemblages due to shared physiological traits or conserved recruitment pathways [15, 16, 18]. These conflicting results highlight the need for experimental approaches that simultaneously consider both the soil of origin and the identity of the host plant.

Recent phylogenetic analyses have shown that host identity can lead to clustering of associated microbes into distinct evolutionary lineages, even when plants are grown in markedly different soils [19, 20]. Such patterns suggest that plants might recruit microbial clades with particular functional or genetic traits, potentially reflecting long-term coadaptation between hosts and their microbial partners. In legumes, compatibility filters are especially strong because symbiotic nitrogen fixation depends on precise and highly regulated molecular interactions. Consequently, legumes may favour specific bacterial groups that are able to activate the appropriate signalling cascades, colonize root tissues, or compete effectively within the rhizosphere [21, 22]. These processes collectively support the idea that the phylogenetic structure of legume-associated bacteria might reflect evolutionary constraints and selective filters imposed by the plant, rather than merely the composition of the surrounding soil.

Two conceptual frameworks are particularly relevant for interpreting plant–microbe evolutionary patterns. First, the reservoir–filter model proposes that the soil provides a diverse pool of microbial lineages, from which the plant selectively recruits taxa according to compatibility and functional needs [23]. Under this model, the soil defines the available diversity, but the plant determines which clades are enriched.

Second, the host-driven convergence model suggests that different soils may lead to different starting communities, but host specificity subsequently overrides these differences and drives communities toward similar phylogenetic compositions [24]. Both frameworks are consistent with scenarios in which host identity dominates the assembly of microbial communities associated with legumes. Because this study focuses on culturable isolates, these frameworks are evaluated within a selective but ecologically meaningful subset of the soil microbiota, rather than across the full microbial community.

Given these theoretical perspectives and the patterns observed in recent literature, understanding whether legumes recruit specific microbial clades from the soil—and whether this selection is phylogenetically structured—can provide insight into the evolutionary basis of host–microbe associations. By examining the phylogeny of culturable bacteria isolated from soils before planting and again after growing different legume hosts, we specifically evaluate whether the culturable fraction shows detectable host-driven phylogenetic filtering relative to the available edaphic reservoir, rather than aiming to characterize full microbial community assembly. In this study, we focus on two testable hypotheses emerging from these ecological and evolutionary frameworks, (i) within the culturable fraction of soil and nodule-associated bacteria, legume hosts may show detectable tendencies toward phylogenetic filtering relative to the available edaphic reservoir, rather than assembling random subsets of soil diversity; (ii) different legume species may exhibit distinct and reproducible patterns of culturable

bacterial recruitment across soils, consistent with host-specific compatibility filters acting on the culturable pool.

To evaluate these hypotheses within the culturable fraction, we isolated bacteria from soils before planting and from root nodules after growing three legume species across five contrasting soils. This design allows us to assess host-associated phylogenetic tendencies relative to the available edaphic reservoir, while acknowledging that the study focuses on a selective subset of the soil microbiota.

Methods

Soil Collection and Characterization

Soils were collected from five rainfed, fallow, small-scale sugarcane plantations farms with nutrient poor soils, located along the northern and southern coasts of KwaZulu-Natal (KZN), South Africa. The northern sites comprised Mvutshini (28°50'52.9"S, 32°00'09.0"E), Mpembeni (28°51'25.7"S, 31°58'19.4"E), and Gingindlovu (28°56'29.3"S, 31°34'58.9"E), whereas the southern sites included Hibberdene (30°35'37.1"S, 30°31'30.3"E) and Mzinto (30°16'41.0"S, 30°39'47.9"E). All sites are situated in subtropical coastal zones characterized by warm temperatures, high humidity, and annual precipitation ranging from approximately 650–1,000 mm [25]. At each location, forty composite soil samples were collected using a screw-type Dutch auger to a depth of 20 cm, following a zig-zag transect design with sampling points spaced at 1 m intervals. Soils were transported to the laboratory in sealed sterile bags, homogenized, and stored at 4°C until use. All soils were classified as sandy, with 80–90% sand and 7–13% clay, and exhibited acidic pH values between 4.0 and 6.0.

Plant Material, Germination, and Greenhouse Experiment

Commercially sourced seeds of Cowpea (*Vigna unguiculata*), Groundnut (*Arachis hypogaea*), and Mungbean (*Vigna radiata*) were obtained from AGT Foods, Marji Mizuri Farm, Ingomankulu (KZN). Seeds were sterilized by soaking in warm water overnight to promote germination, transferred to sterile Petri dishes lined with moist filter paper, and incubated in darkness for four days. Germinated seedlings were transplanted individually into 15-cm pots containing homogenized soil from each of the five sites. The experiment followed a completely randomized design with 15 replicate pots per soil (n = 75 pots per legume species). Plants were grown in a controlled greenhouse at the School of Life Sciences, University of KwaZulu-Natal (Westville campus), under the following conditions: daytime temperature 30–33°C, night time temperature 11–13°C, relative humidity 70–80%, and irradiance equivalent to 35% of full sunlight (approx. 415 µmol m^{-2} s^{-1}). Plants were irrigated every two days using distilled water.

Isolation of Nodule-Associated Bacteria

Harvests were conducted at 90 days after seedling emergence. Nodules were excised manually. Nodules were surface-sterilized by sequential immersion in 70% ethanol (30 s) and 3.5% sodium hypochlorite (3 min), followed by ten rinses in sterile distilled water. Sterilized nodules were crushed in sterile 15% glycerol. The resulting suspension was streaked onto yeast mannitol agar (YMA). YMA medium consisted of 0.5 g L^{-1} yeast extract, 10 g L^{-1} mannitol, 0.5 g L^{-1} K_2HPO_4, 0.2 g L^{-1} $MgSO_4 \cdot 7H_2O$, 0.1 g L^{-1} NaCl, and 15 g L^{-1} agar. Plates were incubated at 28°C and re-streaked until pure cultures were obtained.

Isolation of Free-Living Soil Bacteria

Pre- and post-experiment soils from each site were subjected to standard serial dilution plating. Three functional groups were targeted using selective media: Phosphate-solubilizing bacteria were extracted from Pikovskaya's agar containing tricalcium phosphate (TCP) as the insoluble P source. Nitrogen-cycling bacteria grew in Simmons citrate agar, containing citrate as the sole carbon source and inorganic ammonium salts as the nitrogen source. Nitrogen-fixing bacteria grew in Jensen's nitrogen-free agar. Plates were incubated at 30°C for 4–6 days. Distinct colonies were sub-cultured onto fresh medium until pure isolates were obtained.

DNA Extraction, 16S rRNA Gene Amplification, and Sequencing

Pure bacterial isolates from nodules and soil were subjected to colony PCR targeting the 16S rRNA gene using primers 63F (5'-CAGGCCTAACACATGCAAGTC-3') and 1387R (5'-GGGCGGTGTGTACAAGGC-3'). PCR reactions (25 µL) contained 12.5 µL Emerald Amp GT Master Mix (Separations, South Africa), 0.5 µL of each primer, 3 µL of bacterial suspension, and 8.5 µL sterile Milli-Q water. Amplifications were performed on a T100 Thermal Cycler (Bio-Rad) under the following conditions: initial denaturation at 94°C for 2 min; 30 cycles of 92°C for 30 s, 56°C for 45 s, and 75°C for 45 s; and a final extension at 75°C for 10 min. PCR products were visualized on 1% agarose gels stained with 6× Orange DNA Loading Dye. Amplicons of the expected size (~ 1,324 bp) were excised and sequenced at Inqaba Biotech (Pretoria, South Africa).

Sequences were quality-checked and compared against the NCBI GenBank database using BLASTn (accessed 20 August 2023) to determine taxonomic identities. All 16S rRNA gene sequences generated in this study have been deposited in GenBank (National Center for Biotechnology Information, NCBI) under the accession numbers listed on tables S1 and S2.

Taxonomic assignment

The taxonomic identity of each isolate was determined based on 16S rRNA gene sequence similarity using BLASTn against the NCBI GenBank database. Taxonomic assignments were made using commonly applied sequence identity thresholds, with ≥ 99% identity considered indicative of putative species-level affiliation, ≥ 97% for genus-level assignment, and ≥ 95% for family-level classification.

Isolates showing < 99% identity to described species were reported as Genus sp., while those with < 97% identity were assigned to the corresponding higher taxonomic level. In cases where BLAST results yielded similar identity values across multiple closely related taxa, assignments were made to the lowest taxonomic rank supported unambiguously by the sequence data.

All taxonomic assignments and corresponding GenBank accession numbers are provided in Tables S1 and S2.

Phylogenetic analyses

The taxonomic identity of each isolate was determined based on 16S rRNA gene sequence similarity using BLASTn against the NCBI GenBank database. Taxonomic assignments were made using commonly applied sequence identity thresholds, with ≥ 99% identity considered indicative of putative species-level affiliation, ≥ 97% for genus-level assignment, and ≥ 95% for family-level classification.

Isolates showing < 99% identity to described species were reported as Genus sp., while those with < 97% identity were assigned to the corresponding higher taxonomic level. In cases where BLAST results yielded similar identity values across multiple closely related taxa, assignments were made to the lowest taxonomic rank supported unambiguously by the sequence data. Unweighted UniFrac distances were selected due to the use of culturable isolates and the associated biases in abundance estimation

Data analyses

To evaluate whether the distribution of culturable bacterial taxa differed among soils, host plants and compartments, contingency tables were constructed and analyzed using Fisher's exact tests due to the presence of low expected frequencies in several cells. For larger contingency tables, Monte Carlo simulation was used to approximate exact p-values. Effect sizes were not calculated due to the limited sample size and sparse distribution of counts across categories.

Taxonomic diversity was assessed using richness (number of distinct culturable genera) and Shannon's diversity index for each soil and host plant. These indices were used descriptively to compare diversity patterns among treatments, without applying inferential statistical tests due to limited sample sizes. To account for differences in sampling effort, individual-based rarefaction curves were generated based on random permutations, and mean expected richness was calculated along with associated variability.

Differences in community composition were evaluated using Bray–Curtis dissimilarity matrices calculated from the relative abundance of bacterial taxa. Principal Coordinates Analysis (PCoA) was used to visualize patterns of compositional variation among soils and host plant species. Given the absence of true biological replicates within treatments, multivariate statistical tests were not applied, and observed patterns were interpreted descriptively.

Phylogenetic beta diversity was assessed using unweighted UniFrac distances derived from the Maximum Likelihood phylogenetic trees. UniFrac distances were calculated based on presence–absence data to minimize biases associated with culturable isolates and differences in isolation frequency. These analyses were used to compare the phylogenetic structure of bacterial communities associated with different soils and host plants.

In addition, phylogenetic community structure was quantified using mean pairwise distance (MPD) and mean nearest taxon distance (MNTD). Standardized effect sizes (SES) were calculated using null models based on random reshuffling of taxa labels, allowing the detection of phylogenetic clustering or overdispersion within communities.

All statistical and diversity analyses followed standard ecological approaches, and visualizations were generated in R (version 4.x) and complementary graphical tools.

Results

Composition and phylogenetic structure of soil-derived culturable bacteria

A Maximum-likelihood phylogenetic tree was constructed from 16S rRNA gene sequences of 77 culturable bacterial isolates obtained from the five soils sampled prior to legume cultivation (Mvuts, Gingi, Mpemb, Mzint and Hibbe). The tree resolved several well-supported clades corresponding to major bacterial lineages recovered from the soils, including *Burkholderia sensu lato* (comprising *Burkholderia*, *Paraburkholderia* and *Caballeronia*), *Pseudomonas*, *Paenibacillus*, *Bacillus*, and members of the *Sphingomonadaceae* (Fig. 1). Bootstrap support for internal nodes was generally high, indicating robust phylogenetic structure among the dominant culturable taxa.

When isolates were examined according to soil origin, representatives from all sites were distributed across multiple phylogenetic clades rather than forming soil-specific lineages. Each soil contributed isolates to several of the major bacterial groups, and no monophyletic clusters composed exclusively of isolates from a single soil were observed.

Phylogenetic alpha diversity metrics revealed quantitative differences among soils (Table 2). Observed mean pairwise distance (MPD) values ranged from 0.198 in Mzinto to 0.412 in Mpemb, while standardized effect sizes (SES-MPD) varied from − 2.16 to 0.34 across soils. Observed mean nearest taxon distance (MNTD) ranged from 0.012 in Gingi to 0.052 in Mpemb, with SES-MNTD values spanning from − 3.31 to − 0.41. Two soils, Gingi and Mvuts, exhibited SES-MNTD values below − 1.96, indicating significant phylogenetic clustering.

Table 1
Richness and Shannon diversity indices across host plant species (cowpea, groundnut, mungbean) and soils from different locations.

Host		Rhichness	Shannon's diveristy
	Cowpea	**4**	**1.332**
	Groundnut	6	1.696
	Mungbean	4	1.213
Soil	Gingi	2	0.693
	Hibbe	4	1.332
	Mpemb	3	1.011
	Mvuts	5	1.609
	Mzinto	4	1.386

Table 2
Net phylogenetic alpha diversity metrics and standardized effect sizes (SES) calculated across the five consolidated soil reservoirs. Taxonomic community structure was evaluated using Mean Pairwise Distance (MPD) to detect deep phylogenetic clustering and Mean Nearest Taxon Distance (MNTD) to evaluate terminal branch clustering at the tips of the tree. Negative SES values indicate a significantly clustered community architecture under localized environmental filtering. Statistically significant deviations from random evolutionary expectations ($p < 0.05$) are designated with asterisks (*).

Soil origin	Nº of strains	Observed MPD	SES (MPD	Observed MNTD	SES (MNTD
Gingi	12	0.245	-1.42	0.012	-1.98*
Hibbe	7	0.389	0.21	0.045	-0.64
Mpemb	7	0.412	0.34	0.052	-0.41
Mvuts	18	0.315	-1.18	0.018	-2.05*
Mzinto	14	0.198	-2.16	0.014	-3.31

Note: MPD = Mean Pairwise Distance; MNTD = Mean Nearest Taxon Distance; SES = Standardized Effect Size. Negative SES values indicate phylogenetic clustering, while positive values indicate overdispersion. Asterisks () denote statistically significant deviations from the null model distribution ($p < 0.05$)*.

Pairwise phylogenetic distances among soils further indicated moderate differentiation between communities (Table 3), ranging from 0.284 to 0.421. The smallest distance was observed between Mpemb and Mvuts (0.284), whereas the greatest divergence occurred between Gingi and Hibbe (0.421).

Table 3
Pairwise mean phylogenetic distance matrix calculated among the five consolidated soil reservoirs. Values populate a symmetric lower-triangular matrix where each cell represents the net mean phylogenetic separation computed across all pairwise operational taxonomic unit (OTU) combinations between respective edaphic reservoirs. Lower matrix coefficients indicate a higher degree of shared evolutionary history and taxonomic similarity between the culturable bacterial assemblages.

Origin of Soil	Gingi	Hibbe	Minzo	Mvuts	Mzinto
Gingi	0.000				
Hibbe	0.421	0.000			
Mpemb	0.342	0.411	0.000		
Mvuts	0.311	0.387	0.284	0.000	
Mzinto	0.354	0.301	0.378	0.341	0.000

Host-associated bacterial communities recovered from nodules

A Maximum Likelihood phylogeny was reconstructed from 16S rRNA gene sequences of bacterial isolates obtained from nodules of cowpea, groundnut, and mungbean. The resulting tree resolved several well-supported clades corresponding to dominant culturable genera, including Burkholderia sensu lato (*Burkholderia*, *Paraburkholderia* and *Caballeronia*), *Pseudomonas*, *Paenibacillus*, and members of the *Sphingomonadaceae* (Fig. 2). Bootstrap support for major internal nodes was consistently high, indicating clear phylogenetic structure among nodule-associated isolates.

When isolates were examined according to host plant, nodule-associated bacteria from each legume species were distributed across multiple phylogenetic clades rather than restricted to a single lineage. Each host contributed isolates to several of the dominant bacterial groups, although clustering of isolates within particular clades was evident for individual host species. In contrast to the broadly dispersed pattern observed in soils, nodule isolates showed a tendency to group within specific regions of the phylogeny.

Alpha and beta diversity patterns across hosts and soils

Phylogenetic alpha diversity metrics differed among host plants (Table 4). Observed mean pairwise distance (MPD) ranged from 0.102 in mungbean to 0.442 in cowpea, with standardized effect sizes (SES-MPD) ranging from − 2.41 to 0.48. Mean nearest taxon distance (MNTD) values ranged from 0.008 in mungbean to 0.185 in cowpea, while SES-MNTD varied between − 2.85 and − 0.52. Mungbean exhibited strongly negative SES values for both MPD and MNTD (below − 1.96), indicating significant phylogenetic clustering among its associated isolates.

Table 4
Net phylogenetic alpha diversity metrics and standardized effect sizes (SES) calculated across distinct soil reservoirs for the second isolated bacterial cluster (N = 21).

Nodules origin	Nº of strains	Observed MPD	SES (MPD	Observed MNTD	SES (MNTD
Groundnut	10	0.395	0.14	0.054	-1.35
Cowpea	5	0.442	0.48	0.185	-0.52
Mungbean	6	0.102	-2.41*	0.008	-2.85*

Note: Taxonomic community structure was evaluated using Mean Pairwise Distance (MPD) to detect deep phylogenetic clustering and Mean Nearest Taxon Distance (MNTD) to evaluate terminal branch clustering at the tips of the Maximum Likelihood tree. Metric deviations were tested against a null model distribution derived from 999 randomizations using the 'taxa.labels' algorithm in the R package picante. Negative SES values indicate a significantly clustered community architecture under localized environmental filtering, while positive values represent overdispersion. Statistically significant deviations from random evolutionary expectations ($p < 0.05$) are designated with asterisks.

Rarefaction curves revealed substantial variability in culturable richness across soils (Fig. 3A), reflecting the heterogeneous nature of the edaphic reservoir. In contrast, rarefied richness among host plants showed a narrower range and a tendency toward convergence (Fig. 3B), consistent with nodules harbouring a filtered subset of the available soil diversity.

Differences in community composition were evident in the Bray–Curtis PCoA, with soils distributed broadly across ordination space (Fig. 4A) and host-associated communities showing clearer clustering patterns, particularly for groundnut (Fig. 4B).

Consistent with the broad phylogenetic dispersion observed in soil isolates (Fig. 1), the plant-associated dataset also revealed non-random phylogenetic relationships among taxa. The pairwise cophenetic distances (Table 5) showed that bacterial assemblages from Cowpea and Mungbean were more closely related (0.3) than those associated with Groundnut, which displayed higher divergence values (0.5–0.7). This pattern suggests that plant identity may exert a selective filter shaping the phylogenetic structure of culturable symbionts, even when soil communities themselves do not exhibit clustering by origin.

Table 5
Pairwise mean phylogenetic distance matrix calculated among the three plant-associated bacterial taxa derived from the second isolated bacterial dataset.

Origin of nodules	Groundnut	Cowpea	Mungbean
Groundnut	0		
Cowpea	0.5	0	
Mungbean	0.7	0.3	0

Note: Values populate a symmetric distance matrix where each cell represents the net mean phylogenetic separation (cumulative branch length divergence) computed across all pairwise operational taxonomic unit (OTU) combinations between respective bacterial taxa associated with Groundnut, Cowpea, and Mungbean. Calculations were executed using cophenetic distance paths derived directly from the unrooted Maximum Likelihood phylogenetic tree. Lower matrix coefficients indicate a higher degree of shared evolutionary history and taxonomic similarity between the culturable bacterial assemblages associated with these plant species.

Discussion

This study provides an integrated evaluation of how soil origin and legume host identity jointly structure the diversity, composition, and phylogenetic organization of culturable bacterial communities. By combining phylogenetic reconstruction, diversity metrics, and compositional analyses, our results reveal a hierarchical pattern of microbial assembly in which soils act as reservoirs of phylogenetically diverse taxa, while plant hosts selectively recruit and structure subsets of this diversity within nodules. This framework is consistent with current ecological models of plant microbiome assembly, which emphasize the interplay between environmental filtering and host-driven selection [7, 23].

Phylogenetic reconstruction of soil-derived isolates demonstrated that bacterial lineages were broadly distributed across major clades, with no evidence of soil-specific phylogenetic clustering. These findings support the well-established view that soils harbour highly diverse and taxonomically widespread microbial communities shaped primarily by abiotic factors such as pH, nutrient availability, and texture [5, 6, 26, 27].

Consistent with this interpretation, rarefaction analyses indicated substantial variability in richness among soils, further reflecting the heterogeneous nature of edaphic environments. At the same time, compositional analyses based on Bray–Curtis dissimilarity revealed clear differences among soils, indicating that while phylogenetic lineages are broadly shared, their relative representation varies across sites.

In contrast, host-associated nodule communities displayed more structured patterns. Although isolates from each legume species were distributed across multiple phylogenetic clades, both compositional (Bray–Curtis) and phylogenetic (UniFrac) analyses indicated that plant-associated communities were more constrained than soil communities. Rarefaction curves further showed convergence in richness among host plants, suggesting that nodules harbour a subset of the available soil diversity. Together,

these results indicate that plant hosts do not randomly assemble their associated microbiota, but rather select from the surrounding microbial pool.

Differences among host plants were also evident at both compositional and phylogenetic levels. Ordination analyses showed that groundnut-associated communities were distinct from those of cowpea and mungbean, while phylogenetic diversity metrics revealed stronger clustering in certain hosts, particularly mungbean. These patterns are consistent with previous studies demonstrating that plant species can selectively recruit distinct microbial assemblages through a combination of biochemical signalling, immune interactions, and compatibility filters [2, 18, 20, 21]. In legumes, these processes are particularly relevant due to the specificity of symbiotic interactions and the requirement for coordinated molecular communication between host and bacteria [10, 11].

The phylogenetic distances observed among nodule-associated isolates further support the idea that plant identity acts as a selective filter shaping the evolutionary structure of culturable symbionts [7, 9]. The closer relatedness between Cowpea- and Mungbean-associated taxa, compared with the more divergent Groundnut assemblage, suggests that host traits may influence which bacterial lineages are preferentially recruited or retained. This pattern aligns with previous work showing that legume hosts can impose phylogenetically structured filters on their microbial partners, even when the surrounding soil community is taxonomically broad and unstructured [3, 8].

Importantly, the integration of alpha, compositional, and phylogenetic approaches provides a coherent picture of community assembly. Soil communities exhibited greater variability in both composition and phylogenetic structure, as reflected by wide UniFrac distances and dispersed patterns in ordination space. In contrast, plant-associated communities showed reduced compositional variation and lower phylogenetic divergence, supporting the idea that host plants act as ecological filters that constrain community assembly. This pattern aligns with the reservoir–filter framework, in which soils define the available diversity, while plants determine which taxa are able to establish and persist within host tissues [23, 24].

Overall, our results indicate that plant hosts play a central role in structuring nodule-associated bacterial communities, not by generating entirely novel assemblages but by filtering and reorganizing the phylogenetic diversity present in soils. This selective process results in communities that are both compositionally and phylogenetically more constrained than their soil counterparts. These findings contribute to a growing body of literature demonstrating that plant–microbe interactions are governed by both environmental context and host-specific selection, and highlight the importance of considering multiple dimensions of diversity when studying microbiome assembly in legumes.

Conclusions

Our results demonstrate that culturable bacterial isolates from contrasting soils are phylogenetically widespread and do not form soil-specific clusters, indicating that soil origin alone does not predict the evolutionary placement of recovered taxa. Rarefaction analyses further show that soils harbour

comparable levels of culturable diversity despite differences in taxonomic composition. In contrast, plant-associated isolates exhibited structured phylogenetic relationships, with Cowpea and Mungbean sharing more closely related bacterial assemblages than Groundnut. These findings highlight that while soils act as broad reservoirs of phylogenetically diverse bacteria, plant identity may impose selective filters that shape the evolutionary composition of associated culturable taxa.

Declarations

The authors have no relevant financial or non-financial interests to disclose.

The authors report generative AI was not used in their research or preparation of this manuscript

Author Contribution

All authors contributed equally to the work. All authors participated in data interpretation, manuscript writing and revision, and each contributed substantially to different components of the study.

Data Availability

The datasets generated during and/or analysed during the current study have been deposited in GenBank (National Center for Biotechnology Information, NCBI) under accession numbers PX910258 to PX910294 and PX916018 to PX916089.

Financial statements

This work was supported by the National Research Foundation under Grant agreement number (Grant UID 138091). Nonkululeko Sithole was supported by the National Research Foundation (Grant UID MND200712543053 and PMDS230710131135).

References

1. Hartmann, M., Frey, B., Mayer, J., Mäder, P. & Widmer, F. Distinct soil microbial diversity under long-term organic and conventional farming. *ISME J.* **9**, 1177–1194. 10.1038/ismej.2014.210 (2015).
2. Fañanás-Pueyo, I., Carrera-Castaño, G., Pernas, M. & Oñate-Sánchez, L. Signalling and regulation of plant development by carbon/nitrogen balance. *Physiol. Plant.* **177**, e70228. 10.1111/ppl.70228 (2025).
3. Zhou, Y. et al. Superiority of native soil core microbiomes in supporting plant growth. *Nat. Commun.* **15**, 6599. 10.1038/s41467-024-50685-3 (2024).

4. Yusuf, A. et al. Harnessing plant–microbe interactions: strategies for enhancing resilience and nutrient acquisition for sustainable agriculture. *Front. Plant. Sci.* **16**, 1503730. 10.3389/fpls.2025.1503730 (2025).
5. Fierer, N. & Jackson, R. B. The diversity and biogeography of soil bacterial communities. *PNAS* **103**, 626–631. https://doi.org/10.1073/pnas.0507535103 (2006).
6. Bahram, M. et al. Structure and function of the soil microbiome underlying N2O emissions from global wetlands. *Nat. Commun.* **13**, 1430. 10.1038/s41467-022-29161-3 (2022).
7. Philippot, L. Going back to the roots: The microbial ecology of the rhizosphere. *Nat. Rev. Microbiol.* **11**, 789–799. 10.1038/nrmicro3109 (2013).
8. Edwards, J. et al. Structure, variation, and assembly of the root-associated microbiomes of rice. *Proc. Natl. Acad. Sc. U S A*. n, 112, E911-20. (2015). 10.1073/pnas.1414592112
9. Sasse, J., Martinoia, E. & Northen, T. Feed your friends: do plant exudates shape the root microbiome? *Trends Plant. Sci.* **23**, 25–41. 10.1016/j.tplants.2017.09.003 (2018).
10. Doyle, J. J. Phylogenetic perspectives on the origins of nodulation. *Plant. Physiol.* **156**, 1250–1256. https://doi.org/10.1094/MPMI-05-11-0114 (2011).
11. Sprent, J. I. Evolving ideas of legume evolution and diversity: a taxonomic perspective on the occurrence of nodulation. *New. Phytol*. **174**, 11–18. 10.1111/j.1469-8137.2007.02015.x (2007).
12. Xue, P. P., Carrillo, Y., Pino, V., Minasmy, B. & McBratney, A. Soil Properties drive microbial community etructure in a large scale transect in South-Eastern Australia. *Sci. Rep.* **8**, 11725. 10.1038/s41598-018-30005-8 (2018).
13. Louisson, Z. et al. Soil conditions are a more important determinant of microbial community composition and functional potential than neighbouring plant diversity *iSci*. **27**, 110056 (2024). https://doi.org/10.1016/j.isci.2024.110056
14. Gupta, V. V. S. R. & Tiedje, J. M. Ranking environmental and edaphic attributes driving soil microbial community structure and activity with special attention to spatial and temporal scales. *mLife* 3, 21–41. 10.1002/mlf2.12116
15. Berendsen, R. L., Pieterse, C. M. J. & Bakker, P. .M. The rhizosphere microbiome and plant health. *Trends Plant. Sci.* **17**, 478–486. 10.1016/j.tplants.2012.04.001 (2012).
16. Berendsen, R. L., Pieterse, C. M. J. & Bakker, P. A. H.M. Disease-induced assemblage of a plant-beneficial bacterial consortium. *ISME J.* **12**, 1496–1507. 10.1038/s41396-018-0093-1 (2018).
17. Mendes, R., Garbeva, P. & Raaijmakers, J. M. The rhizosphere microbiome: Significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. *Sci. Adv.* **5**, eaaw9287. 10.1111/1574-6976.12028 (2019).
18. Hacquard, S. et al. Microbiota and Host Nutrition across Plant and Animal Kingdoms. *Cell. Host Microbe*. **17**, 603–616. 10.1016/j.chom.2015.04.009 (2015).
19. Zgadzaj, R., Garrido-Oter, R., Jensen, D. B. & Radutoiu, S. Root nodule symbiosis and the evolution of plant–microbe interactions. *ISME J.* **10**, 60–73. https://doi.org/10.1073/pnas.1616564113 (2016).

20. Masson-Boivin, C. & Sachs, J. L. Symbiotic nitrogen fixation by rhizobia—the roots of a success story. *Curr. Opin. Plant. Biol.* **44**, 7–15. https://doi.org/10.1016/j.pbi.2017.12.001 (2018).
21. Ruan, M., Hu, Z., Zhu, Q., Li, Y. & Nie, X. 16S rDNA Sequencing-Based Insights into the Bacterial Community Structure and Function in Co-Existing Soil and Coal Gangue. *Microorganisms* **11**, 2151. 10.3390/microorganisms11092151 (2023).
22. Fitzpatrick, C. R. et al. The Plant Microbiome: From Ecology to Reductionism and Beyond. *Annu. Rev. Microbiol.* **74**, 81–100. 10.1146/annurev-micro-022620-014327 (2020).
23. Müller, D. B., Vogel, C., Bai, Y. & Vorholt, J. A. The Plant Microbiota: Systems-Level Insights and Perspectives. Annu. Rev. Genet. 50, 211–234 (2016). 10.1146/annurev-genet-120215-034952. Ndlovu, M., Clulow, A.D., Savage, M.J., Nhamo, L., Magidi, J., Mabhaudhi, T. An assessment of the impacts of climate variability and change in KwaZulu-Natal Province, *S. Afri. Atmosph.* 12, 427 (2021). DOI:10.3390/atmos12040427.
24. Lauber, C. L., Hamady, M., Knight, R. & Fierer, N. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl. Environ. Microbiol.* **75**, 5111–5120. 10.1128/AEM.00335-09 (2009).
25. Delgado-Baquerizo, M. et al. Fierer, N. A global atlas of the dominant bacteria found in soil. *Science* **359**, 320–325. 10.1126/science.aap9516 (2018).

Figures

Figure 1

Maximum-likelihood phylogenetic tree of 16S rRNA sequences from bacterial isolates collected across five soils (Mvuts, Gingi, Mpemb Mzint, Hibbe), color-coded by soil origin. Isolates from each soil are widely distributed across major bacterial lineages, including Burkholderia sensu lato, Pseudomonas, Paenibacillus, Bacillus, and Sphingomonadaceae, with no soil-specific clustering. This broad distribution

indicates that soils contain a diverse reservoir of phylogenetically widespread taxa and that soil origin does not predict the phylogenetic placement of culturable isolates.

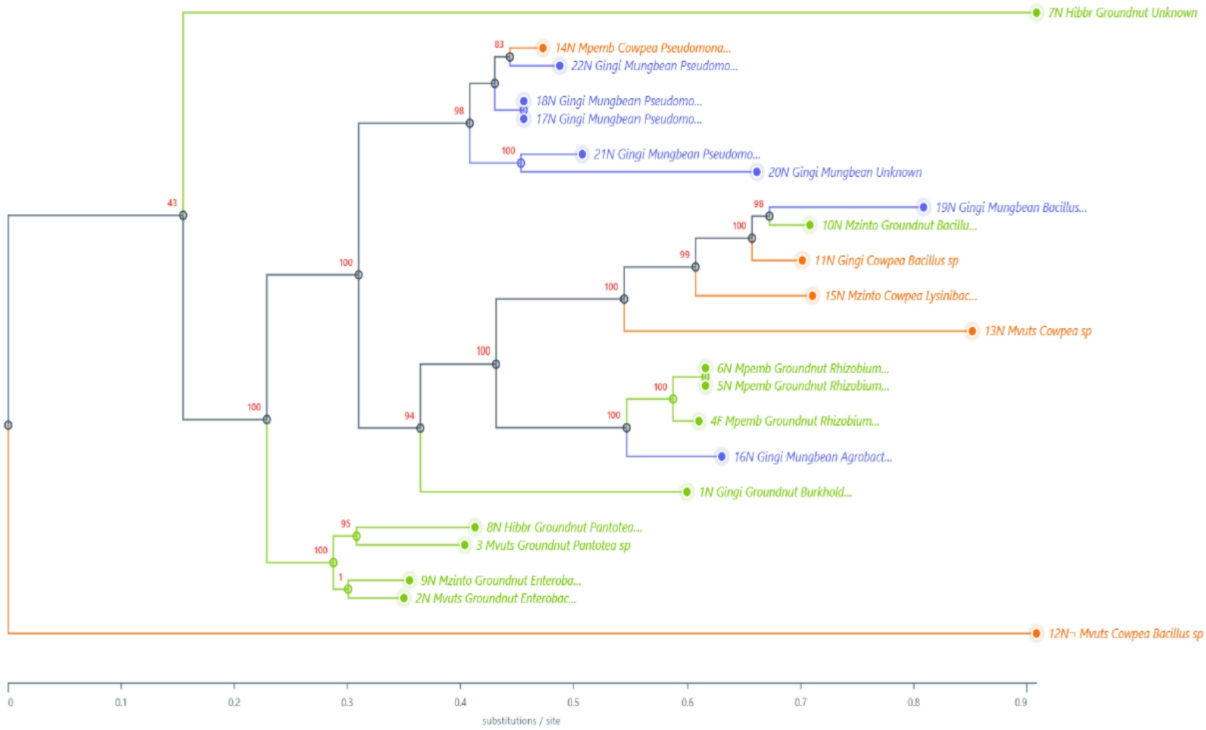

Figure 2

Maximum-likelihood phylogenetic tree of 16S rRNA sequences from bacterial isolates collected across five soils (Mvuts, Gingi, Mpemb Mzint, Hibbe), color-coded by soil origin. Isolates from each soil are widely distributed across major bacterial lineages, including Burkholderia sensu lato, Pseudomonas, Paenibacillus, Bacillus, and Sphingomonadaceae, with no soil-specific clustering. This broad distribution indicates that soils contain a diverse reservoir of phylogenetically widespread taxa and that soil origin does not predict the phylogenetic placement of culturable isolates.

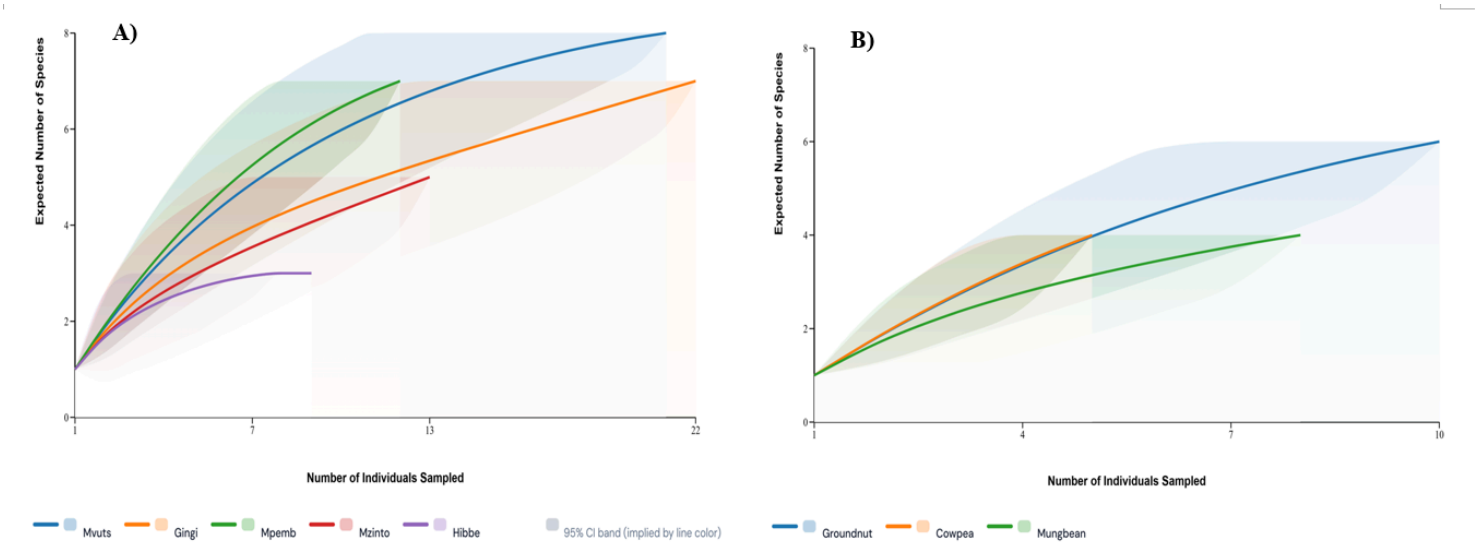

Figure 3

Individual-based rarefaction curves showing the accumulation of culturable bacterial genera recovered from different soil (A) and host pants (B). Curves were generated using 500 random permutations of the isolate dataset, and shaded areas represent ±1 standard deviation around the mean expected richness. Rarefied richness estimates were used to compare taxonomic diversity among treatments while accounting for differences in sampling effort and isolate abundance.

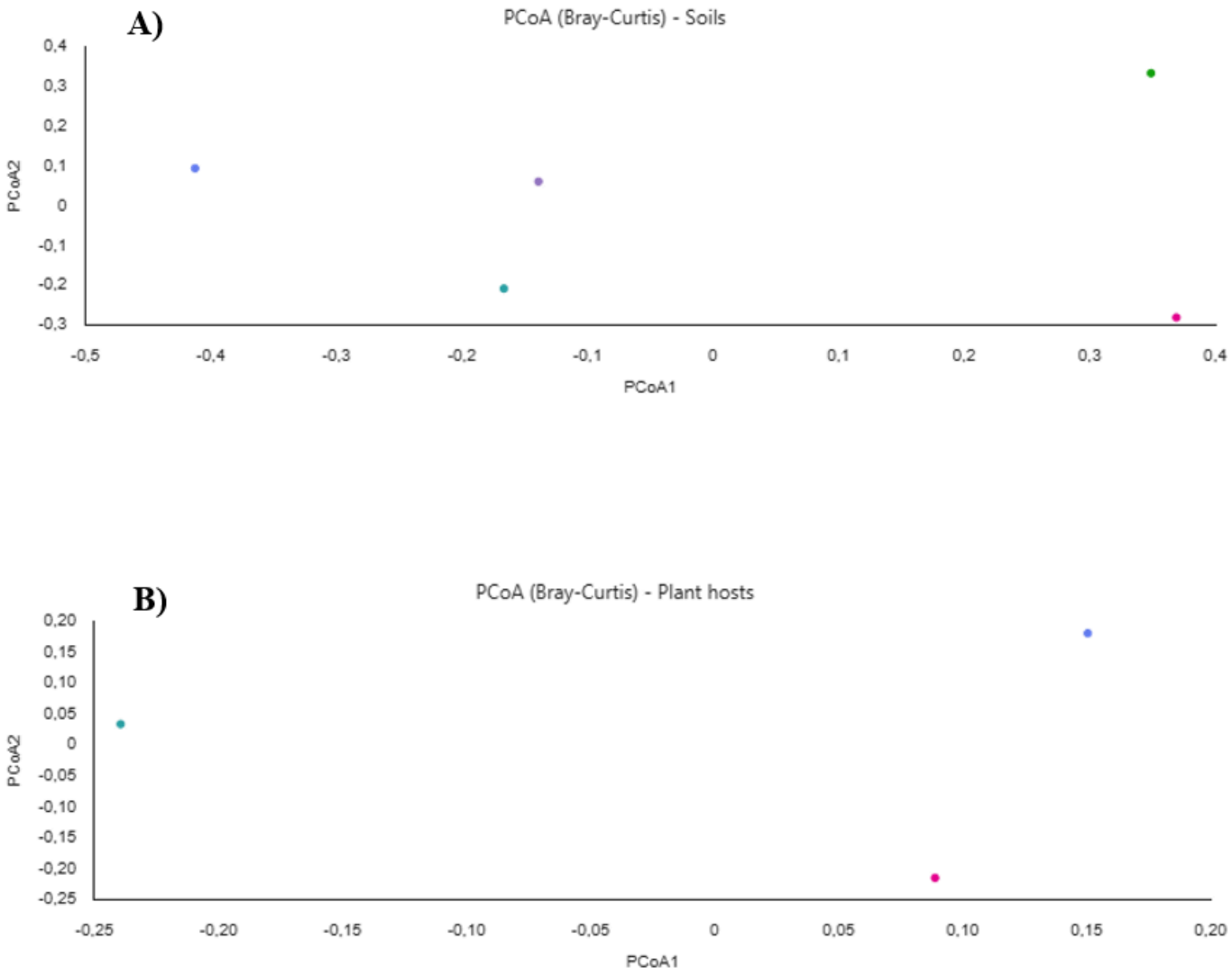

Figure 4

Principal Coordinates Analysis (PCoA) based on Bray–Curtis dissimilarity showing differences in bacterial community composition across (A) soils and (B) plant host species. Each point represents a composite sample based on the relative abundance of culturable bacterial taxa at the genus level. In panel A, soils are distributed across ordination space, indicating variation in community structure among sites. In panel B, plant-associated communities show distinct clustering patterns, with groundnut differing from cowpea and mungbean, suggesting a host-driven effect on bacterial assemblages. Due to the absence of biological replicates, patterns should be interpreted descriptively.

Supplementary Files

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

- Supplementarymaterial.docx