Brazilian Journal of Biology®

ISSN 1519-6984 (Print)
ISSN 1678-4375 (Online)

Original Article

Bacterial diversity of lowland soils under cocoa cultivation in Amazon

Diversidade bacteriana dos solos de várzea sob cultivo de cacau na Amazônia

L. D. S. Ferreira[a*], J. F. B. R. Silva[b], M. P. S. P. Vilhena[c], O. C. Alegria[a], R. T. J. Ramos[a] and M. P. A. Sousa[d]

[a]Universidade Federal do Pará – UFPA, Belém, PA, Brasil
[b]Museu Paraense Emílio Goeldi – MPEG, Coordenação de Ciências Exatas e Ecologia, Belém, PA, Brasil
[c]Universidade Federal Rural da Amazônia, Belém, PA, Brasil
[d]Museu Paraense Emílio Goeldi – MPEG, Serviço de Tecnologia da Informação – STI, Belém, PA, Brasil

Abstract

This study investigated bacterial diversity in soils from six cacao-producing islands in Mocajuba, Pará, Brazil. Using next generation sequencing shotgun metagenomic DNA, we characterized the microbial composition and ecological structure of floodplain soils cultivated with *Theobroma cacao*. Taxonomic classification revealed a rich bacterial community encompassing 21 phyla, 54 classes, 121 orders, 240 families, 604 genera, and 2,289 species. The dominant phyla, Actinomycetota and Pseudomonadota, are known for their ecological roles in organic matter decomposition, antibiotic production, nitrogen cycling, and plant growth promotion. Alpha diversity metrics varied among samples, with P3 showing the highest species richness and P5 exhibiting the highest Shannon, Simpson, and evenness indices, suggesting a more balanced community. Beta diversity analysis based on Bray–Curtis dissimilarity under Total Sum Scaling (TSS) normalization revealed ecological gradients ranging from 0.228 to 0.527. Spatial ordination and hierarchical clustering indicated gradual shifts in community composition, supporting the concept of a compositional continuum shaped by environmental gradients. Functionally, *Burkholderia lata* was dominant in P1, reflecting its role in potassium solubilization, while Streptomyces species—detected in five of the six samples—contribute to biogeochemical cycling and pathogen suppression. *Bradyrhizobium* and *Paraburkholderia*, identified in P3, P5, and P6, are associated with nitrogen fixation and plant hormone regulation. These findings reveal the ecological complexity and functional potential of cacao soil microbiomes, providing insights for sustainable management of Amazonian floodplain agroecosystems.

Keywords: shotgun metagenomics, soil microbiome, floodplain soils.

Resumo

Este estudo investigou a diversidade bacteriana em solos de seis ilhas produtoras de cacau em Mocajuba, Pará, Brasil. Utilizando sequenciamento metagenômico shotgun de DNA, caracterizamos a composição microbiana e a estrutura ecológica de solos de várzea cultivados com *Theobroma cacao*. A classificação taxonômica revelou uma comunidade bacteriana rica, abrangendo 21 filos, 54 classes, 121 ordens, 240 famílias, 604 gêneros e 2.289 espécies. Os filos dominantes, Actinomycetota e Pseudomonadota, são reconhecidos por seus papéis ecológicos na decomposição da matéria orgânica, produção de antibióticos, ciclagem de nitrogênio e promoção do crescimento vegetal. As métricas de diversidade alfa variaram entre as amostras, com P3 apresentando a maior riqueza de espécies e P5 exibindo os maiores índices de Shannon, Simpson e equabilidade, indicando uma comunidade mais equilibrada. A análise de diversidade beta, baseada na dissimilaridade de Bray–Curtis sob normalização por Total Sum Scaling (TSS), revelou gradientes ecológicos variando de 0,228 a 0,527. A ordenação espacial e o agrupamento hierárquico indicaram mudanças graduais na composição das comunidades, sustentando o conceito de um continuum composicional moldado por gradientes ambientais. Funcionalmente, *Burkholderia lata* foi dominante em P1, refletindo seu papel na solubilização de potássio, enquanto espécies de Streptomyces, detectadas em cinco das seis amostras, contribuem para a ciclagem biogeoquímica e supressão de patógenos. *Bradyrhizobium* e *Paraburkholderia,* identificados em P3, P5 e P6, estão associados à fixação biológica de nitrogênio e à regulação hormonal vegetal. Esses resultados revelam a complexidade ecológica e o potencial funcional do microbioma do solo cacaueiro, oferecendo subsídios para o manejo sustentável de agroecossistemas de várzea amazônicos.

Palavra-chave: metagenômica shotgun, microbioma do solo, solos de várzea.

*e-mail: luan.ferreirabio@gmail.com
Received: August 4, 2025 – Accepted: January 5, 2026
Editor: Takako Matsumura Tundisi

1. Introduction

Cocoa is cultivated in most tropical regions throughout the world as an economically important crop for smallholder farmers (Holmes et al., 2004). Historically, cacao has been integral to the cultural identity of indigenous communities in the Amazon, influencing social structures and economic practices since colonial times (Chambouleyron and Arenz, 2021).

The best soil for cocoa production is the forest soil rich in humus, which should be well-drained and free-flowing to allow easy penetration of roots capable of retaining moisture during summer, and those that allow circulation of air and moisture (Adejumo and Adejoro, 2015). The beneficial roles of the cocoa microbial community include organic matter decomposition, mineralization of nutrients, biological degradation and as bio-filters for cleaning up soil and improvement of soil structure (Pierson and Smoot, 2001).

The floodplains, because they have soils of higher fertility, are intensively used for agricultural purposes in the period when they are not flooded. The productivity of crops in this ecosystem is higher than in upland (Albuquerque et al., 2023). Until recently, even though soil is an important component of global ecosystems, little was known about biodiversity patterns, the interactions in the belowground section of terrestrial ecosystems, and how they function in nature (Bardgett and van der Putten, 2014; Andriuzzi and Wall, 2017). Moreover, most soil microbial taxa remain undescribed (Ramirez et al., 2014; Oliverio et al., 2020).

The diversity of microorganisms in the soil environment depends on the factors that influence the composition and properties of the microbiome. Soil is affected by numerous environmental factors, both natural and anthropogenic, that is, man-made (Furtak and Galazka, 2019). Soil quality is a topic of concern for researchers worldwide, and the Amazon, as one of the main biomes due to its biodiversity, has been the subject of research on this matter. This biome houses around 40% of the world's remaining forest, playing an essential role in biodiversity conservation, climate regulation, and biogeochemical cycles (Rodrigues et al., 2009).

Currently, through metagenomic tools, it is possible to acquire information about a large group of bacteria called viable but non-cultivable, which generally comprises 97-99% of bacteria in the soil (Hernandez et al., 2010; Shamim et al., 2017).

The phyla Actinomycetota and Pseudomonadota are recognized as the most abundant in various environments, showcasing their ecological significance and versatility. Their prevalence is evident in diverse habitats, this abundance is attributed to their unique metabolic capabilities and adaptability to diversity of conditions (Kaale et al., 2023). Actinomycetota and Pseudomonadota are prevalent in cacao soils, contributing to the degradation of organic matter and nutrient availability (Benaud et al., 2022).

Therefore, it is important to analyze the microbiological diversity in floodplain soils under cocoa cultivation, in order to understand the dynamics that govern the processes that lead these soils to produce high-quality almonds.

2. Materials and Methods

2.1. Study area

The municipality of Mocajuba is located downstream of the Tucuruí Hydroelectric Power Plant in the Lower Tocantins region, comprising more than 10 municipalities. The study was conducted in six islands with native cacao cultivation in the Municipality of Mocajuba, State of Pará, Brazil: 1 - Santana Island, 2 - Santaninha Island, 3 - Angapijó Island, 4 - Conceição Island, 5 - São Joaquim Island, 6 - Tauaré Island (Figure 1). According to the Köppen classification, the climate in this region is of the Ami type, which is tropical and humid. The average annual temperature is 26.5 ºC, with a minimum of 22 ºC and a maximum of 31 ºC. The annual average relative humidity is 85%, and the annual average rainfall is 2375 mm, with the wettest season occurring from January to May and the drier season from June to December (Dubreuil et al., 2018).

The islands in this study are primarily located along the Tocantins River. In this stretch of the river, there is a semidiurnal tidal influence, and during the flood period, the river level rises, especially in March, and during the ebb, it decreases in September and October. The annual water level variation in this region is approximately 9 meters. In the ebb phase, the river water appears light green, while during the flood, the water becomes turbid due to the high presence of sediments from erosion along the banks of the Upper Tocantins (Pelicice et al., 2025). During the flood period, the cacao trees on the islands become inundated, and this period coincides with the cacao harvest.

2.2. Soil metagenomic analysis

Soil samples for metagenomic analysis were collected from six islands, each within a flat and homogeneous area under *Theobroma cacao* (cacao) cultivation, at a depth of up to 20 cm from the topsoil. A single sampling site was designated on each island, where ten equidistant subsamples (spaced 10 meters apart) were randomly collected during the drier period (without flooding), specifically on August 18 and 19, 2023. Sampling was performed using a Dutch auger, which was sterilized between each collection to prevent cross-contamination. The subsamples were pooled and homogenized in 20 kg plastic bags, from which 100 g aliquots were weighed, packaged in sealed bags, and stored in 1 L Styrofoam containers. All procedures were conducted using gloves and face masks to minimize the risk of sample contamination.

Immediately after soil collection, samples were sent to the IPEC Guarapuava laboratory (Guarapuava, Paraná, Brazil) on August 21, 2023. The laboratory followed a protocol for genetic material extraction using the DNeasy PowerSoil Pro Kit (Qiagen). For the construction of metagenomic sequencing libraries via the shotgun approach, samples were processed according to the manufacturer's protocol, Illumina DNA Flex (also known as "Illumina DNA Prep"). Sequencing was carried out using the NovaSeq SP kit (300 cycles) with the Paired-End 2×150 bp method, generating paired reads of 150 bp on the Illumina NovaSeq 6000 platform. As a result, 8 GB of data were generated per sample (6 samples × 8 GB = 48 GB) in FASTQ format. The

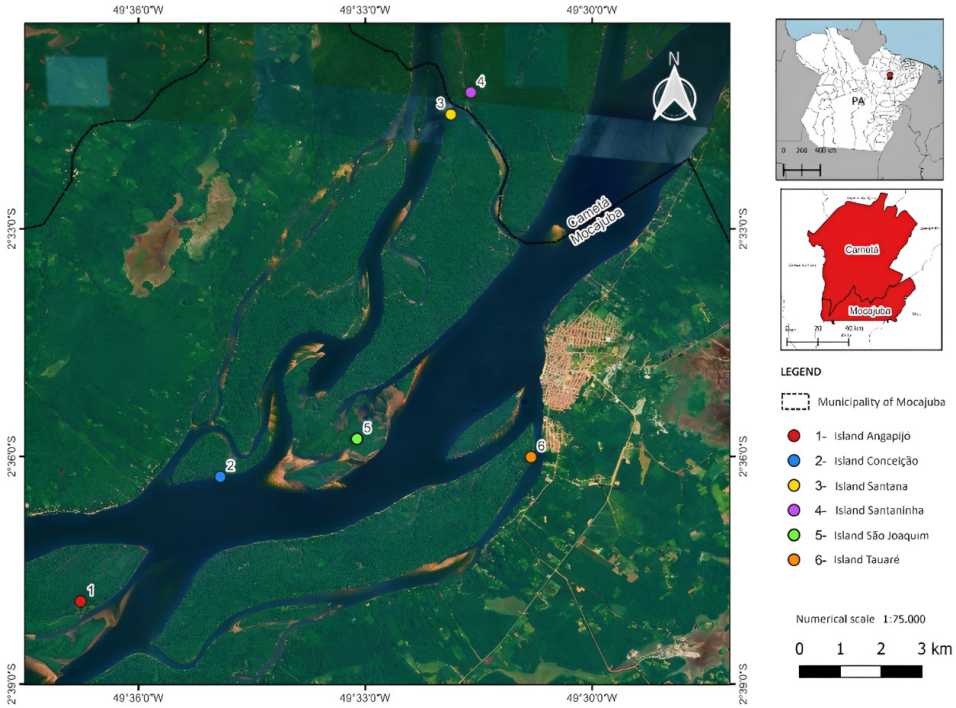

Figure 1. Map of Mocajuba Municipality Location and Study Islands. *The colored-highlighted circles represent the collection points.

adapter sequence used was "CTGTCTCTTATACACATCT," and reads were limited to 151 bp.

Shotgun metagenomic sequencing was performed across six sampling points, generating approximately 50 million raw reads per sample, totaling over 300 million reads. Accessing the data required creating an account and navigating to the "Projects" tab, where the datasets were hosted. For metagenomic community analysis, read quality was assessed and visualized using FastQC (version 0.11.9) (Mishra et al., 2022). Read filtering was performed with Fastp (version 0.23.2). A summary of high-quality reads retained after filtering is provided in Supplementary Table S1.

Initially, technical control sequences from PhiX (NC_001422.1) were removed, followed by filtering against human contaminants (GRCh38) and plant-specific sequences, such as *Theobroma cacao* (GCF000208741.1). Subsequent steps included the removal of ribosomal RNA sequences (16S, 18S, 23S, and 28S) using the SILVA database, as well as filtering against chloroplast genomes (NC_001879.2) retrieved from GenBank. All reference databases were pre-indexed using Bowtie2, and alignments were executed in paired-end mode to ensure precise removal of unwanted sequences. This approach was designed to maximize data quality for downstream taxonomic and functional analyses by minimizing interference from common environmental metagenome contaminants. These include total read counts, base composition, read length distributions, and quality score thresholds (Luandanbio, 2025). A detailed summary of the decontamination results is provided in Supplementary Tables S2.

Taxonomic classification was performed using Kraken2 (v2.1.6) based on the Kraken2 PlusPF database (version 20240715) (Bioinformatics Workbook, 2025). This database was constructed using 100-base pair k-mers and 35-base pair minimizers, parameters that directly influence classification specificity and performance. The confidence threshold was set to 0.1, requiring at least 10% of k-mers to support a taxonomic assignment. Paired-end reads were used, and additional filtering parameters included a minimum base quality of 20 and a minimum of two hit groups per read. This analysis enabled the identification of taxa within the Bacteria domain. To enhance the accuracy of abundance estimation, Bracken (v2.6.0) was applied to the Kraken2 classification outputs (Lu et al., 2017). The full analysis is available at Galaxy Europe.

Six independent analyses were conducted, each corresponding to a distinct taxonomic rank: phylum (P), class (C), order (O), family (F), genus (G), and species (S). The read length parameter was set to 150 bp, in accordance with the sequencing protocol. For diversity analysis, the R package *phyloseq* (McMurdie, 2022) was used to compute alpha and beta diversity metrics.

All classification reports, abundance tables, and intermediate files were generated using the most recent versions of the tools and are available in the project repository to ensure transparency and reproducibility (Luandanbio, 2025). The dataset has been deposited in the NCBI database under BioProject accession PRJNA1224407, with biosamples identified as follows: SRS24151884 (P1), SRS24151885 (P2), SRS24151886 (P3), SRS24151887 (P4), SRS24151888 (P5), and SRS24151889 (P6).

Alpha diversity analyses were conducted using the following ecological indices: richness (number of species with abundance > 0), Shannon index, Simpson index, and evenness (calculated as Shannon / log(Richness)). All calculations were performed in a Python 3.12 environment using the libraries pandas (v2.1.1), numpy (v1.26.0), scipy (v1.11.3), and matplotlib (v3.8.0). Graphs were generated with dual Y-axes to separately represent richness and diversity indices, with color-coded bars for each metric.

Rarefaction curves were constructed through random subsampling of the abundance matrix, simulating different sequencing depths. Subsampling intervals ranged from 100 to 10,000 reads, with increments of 100, and interpolation was applied to smooth the curves. This approach enabled the assessment of sample sufficiency and taxonomic coverage within each community.

To assess bacterial community structure across samples, we employed a beta-diversity framework comparing two normalization strategies: Total Sum Scaling (TSS) and Hellinger transformation. The abundance matrix (species × samples) was preprocessed to exclude samples with zero total counts, ensuring statistical robustness.

TSS normalization was applied to convert raw counts into relative abundances per sample. Subsequently, the Hellinger transformation was performed by taking the square root of these proportions, reducing the influence of dominant taxa and improving suitability for Euclidean-based analyses.

Sample dissimilarity was calculated using the Bray-Curtis index, implemented via the pdist function from the scipy.spatial.distance module. This metric was selected for its sensitivity to both species presence and abundance. The resulting dissimilarity matrix was visualized through hierarchical clustering dendrograms using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) to illustrate relationships among samples.

Community structure was further explored using Principal Coordinates Analysis (PCoA) based on the Bray-Curtis dissimilarity matrix. Dimensionality reduction was performed using the PCA algorithm from the scikit-learn library (v1.3.0), retaining the first two principal components. Plots were styled with individual ellipses around each sample to emphasize ecological groupings and enhance interpretability.

3. Results

3.1. Bacterial community analysis

Taxonomic classification revealed that the proportion of unclassified reads ranged from 93.32% to 95.23%, while classified reads accounted for 4.77% to 6.68% of the total. Among the classified sequences, bacterial reads represented between 1.51% and 2.18% of the total reads per sample. These values reflect the predominance of uncharacterized microbial diversity in the sampled environments and highlight the limited taxonomic resolution achievable with current reference databases. Detailed read distribution per sample is presented in Table 1.

Bacterial rarefaction curves indicate that most samples reached a plateau in species richness, suggesting that the sequencing effort was sufficient to capture the existing diversity. Samples with higher curves exhibit greater diversity, whereas those that stabilize earlier reflect less diverse communities (Figure 2). This pattern supports the reliability of the data for ecological comparisons across samples.

The bacterial community analysis across six metagenomic samples revealed substantial taxonomic diversity. A total of 21 phyla, 54 classes, 121 orders, 240 families, 604 genera, and 2,289 species were identified. These values reflect a rich and complex microbial landscape across the sampled environments. Among the phyla, Actinomycetota and Pseudomonadota were the most abundant, dominating the overall bacterial composition (Figure 3).

The Bray-Curtis dissimilarity matrix among bacterial samples revealed moderate variation in microbial community composition across sampling points. Dissimilarity values ranged from 0.215 to 0.517, indicating that while there is overlap in taxonomic profiles, ecologically meaningful differences exist between locations. The lowest dissimilarity was observed between samples P1 and P2 (0.215), suggesting high similarity likely associated with shared environmental conditions or spatial proximity. In contrast, the highest dissimilarity was recorded between samples P2 and P3 (0.517), reflecting a substantial shift in bacterial community structure Figure 4.

The analysis of ecological indices revealed notable variations in the structure of bacterial communities across samples P1 to P6 (Figure 5). Species richness showed the highest absolute values in samples P3 (1,486 species) and P4 (1,354 species), indicating elevated raw taxonomic diversity. In contrast, sample P6 exhibited the lowest richness (1,056 species), suggesting a less diverse community in terms of species count. However, diversity and evenness indices revealed additional nuances. Sample P5, despite not having the highest richness, displayed the highest Shannon index (6.06) and Simpson index (0.995), indicating a highly diverse

Table 1 - Read Classification and Bacterial Assignment Statistics by Sample.

Sample	Total Reads	% Unclassified Reads	% Classified Reads	Reads Assigned to Bacteria	% Bacterial Reads
P1	54,070,354	94.29%	5.71%	925,544	1.71%
P2	50,553,278	93.70%	6.30%	956,468	1.89%
P3	59,460,840	93.32%	6.68%	1,294,298	2.18%
P4	55,082,046	94.23%	5.77%	1,057,262	1.92%
P5	53,564,386	94.70%	5.30%	997,870	1.86%
P6	49,933,046	95.23%	4.77%	756,070	1.51%

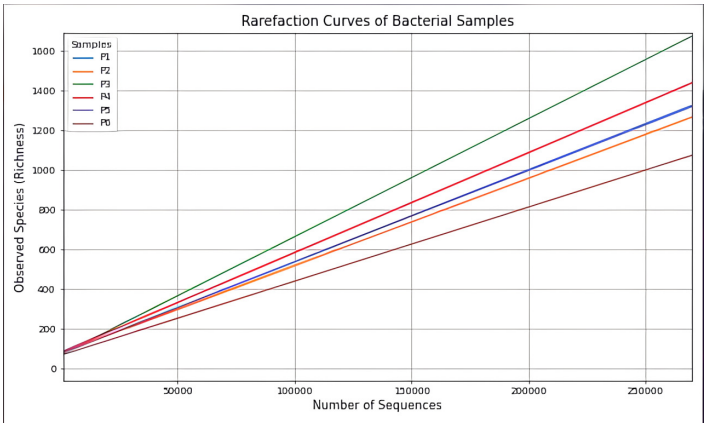

Figure 2. Rarefaction Curves of Bacterial Communities Across Samples P1–P6.

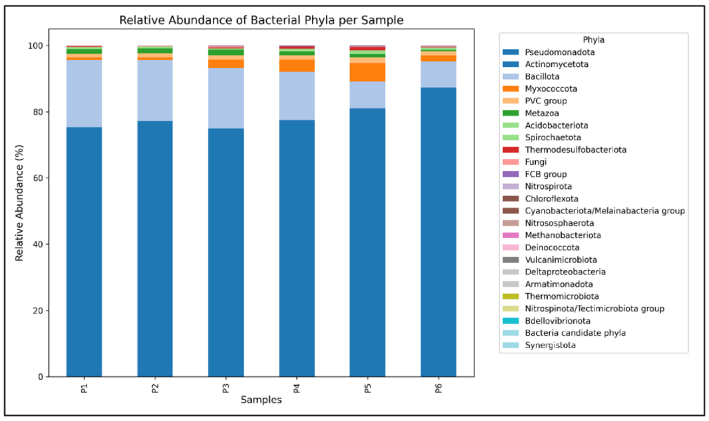

Figura 3. Relative Abundance of Dominant Bacterial Phyla in Metagenomic Soil Samples.

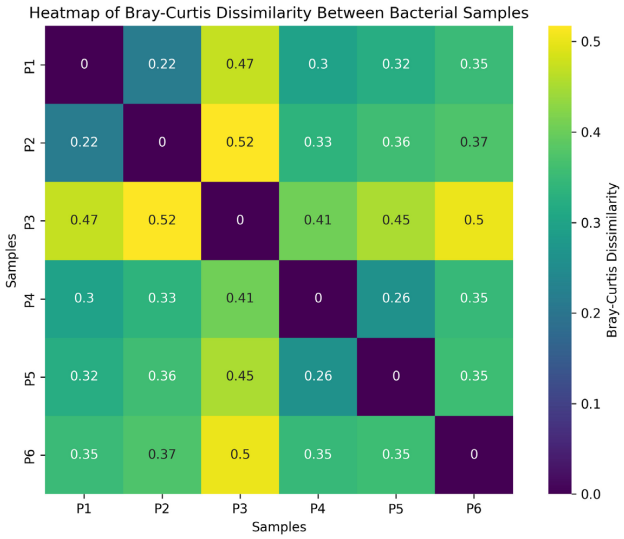

Figure 4. Heatmap of Bray–Curtis Dissimilarity Among Bacterial Samples Reveals Ecological Gradient.

and balanced community with a uniform distribution of taxa abundances. This is further supported by its evenness value of 0.848, the highest among all samples. Conversely, sample P3, although exhibiting the highest richness, showed

relatively lower values for Shannon (5.60), Simpson (0.986), and evenness (0.766), suggesting the presence of dominant species that reduce effective diversity and community uniformity. Sample P6 also stood out for presenting the lowest values across all diversity and evenness indices, which may indicate a less complex bacterial community dominated by a few highly abundant taxa. These findings underscore that richness alone is insufficient to characterize ecological diversity. A robust assessment of microbial community structure requires simultaneous consideration of Shannon, Simpson, and evenness indices.

The Principal Coordinates Analysis (PCoA) plots generated from the Bray-Curtis dissimilarity matrices revealed clear patterns in ecological variation among bacterial communities. For the TSS-normalized data, Axis PCoA1 explained 49.3% of the total variance, while Axis PCoA2 accounted for 32.9%, totaling 82.2% of the variance explained (Figure 6A). In contrast, the Hellinger-transformed data yielded 32.7% and 25.2% for PCoA1 and PCoA2 respectively, summing to 57.9% of the total variance (Figure 6B). These results suggest that TSS normalization preserved more ecological variation in the ordination space, allowing for clearer separation of sample profiles.

To further assess sensitivity to sample variation, we compared the mean Bray-Curtis dissimilarity across samples for each normalization method. The TSS-normalized matrix yielded a higher mean dissimilarity (0.3879) than the Hellinger-transformed matrix (0.2875), indicating that TSS was more sensitive to differences in community composition. This enhanced sensitivity is reflected in both the ordination plots and the clustering structure, where sample separation is more pronounced under TSS normalization.

The hierarchical clustering dendrogram, constructed using Bray-Curtis dissimilarity and the UPGMA (Unweighted Pair Group Method with Arithmetic Mean) algorithm, revealed structured ecological relationships among samples P1 to P6 (Figure 7). Under TSS normalization, samples P1 and P2 clustered

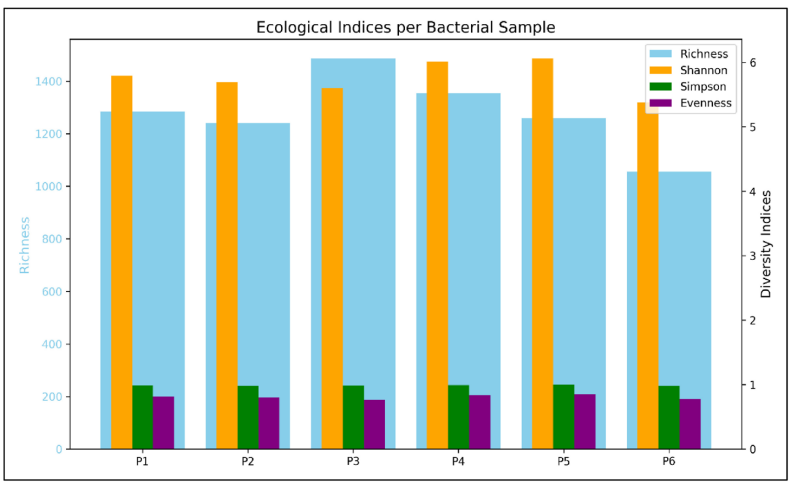

Figure 5. Multimetric Assessment of Bacterial Community Structure Highlights Ecological Complexity Beyond Species Richness.

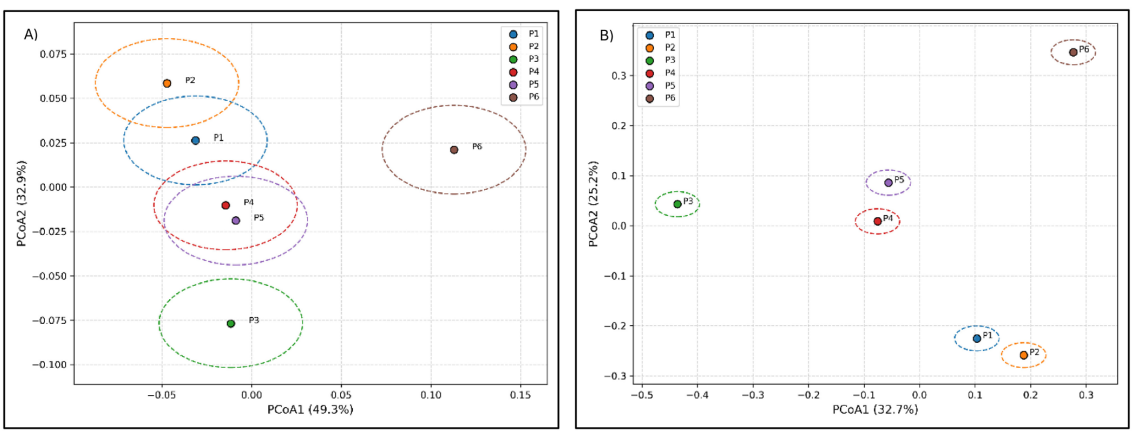

Figure 6. Comparison of Sample Dissimilarity: TSS vs. Hellinger Normalization. Principal Coordinates Analysis (PCoA) based on Bray-Curtis dissimilarity matrices, highlighting patterns of ecological variation among bacterial communities. (A) TSS-normalized data, with PCoA1 and PCoA2 axes explaining 49.3% and 32.9% of the total variance, respectively (82.2% combined). (B) Hellinger-transformed data, with PCoA1 and PCoA2 explaining 32.7% and 25.2% of the variance (57.9% combined).

closely together (dissimilarity = 0.228), suggesting high similarity in community structure. Samples P4 and P5 also formed a coherent cluster (dissimilarity = 0.283), indicating moderate similarity. In contrast, sample P3 was positioned more distantly from all others, particularly from P2 (dissimilarity = 0.527), reflecting its distinct microbial profile. Sample P6, while not the most dissimilar, exhibited consistently elevated dissimilarity values (range: 0.368-0.391) and did not cluster tightly with any other sample, reinforcing its ecological uniqueness.

In samples P2 through P6, a uniform distribution of sequencing reads was observed across all identified species, each registering 996 reads. This homogeneity may indicate a stable microbial community or reflect limitations in the resolution of the sequencing method, such as signal saturation or automated data normalization.

Sample P1 exhibited a distinct pattern: although four of the five species also recorded 996 reads, *Burkholderia lata* stood out with 9,925 reads. Given that all samples underwent prior decontamination, this discrepancy cannot be attributed to contamination. Therefore, the high abundance of *Burkholderia lata* likely reflects a natural dominance of this species within the sample, potentially associated with specific environmental conditions or its competitive advantage within the analyzed niche.

Moreover, species belonging to the genus *Streptomyces* were recurrent across multiple samples (P1, P2, P3, P4, and P5), suggesting broad distribution and potential ecological relevance. Less frequent taxa, such as *Gemmata obscuriglobus* and *Lignipirellula cremea*, were also identified, indicating substantial phylogenetic diversity among the detected microorganisms (Figure 8).

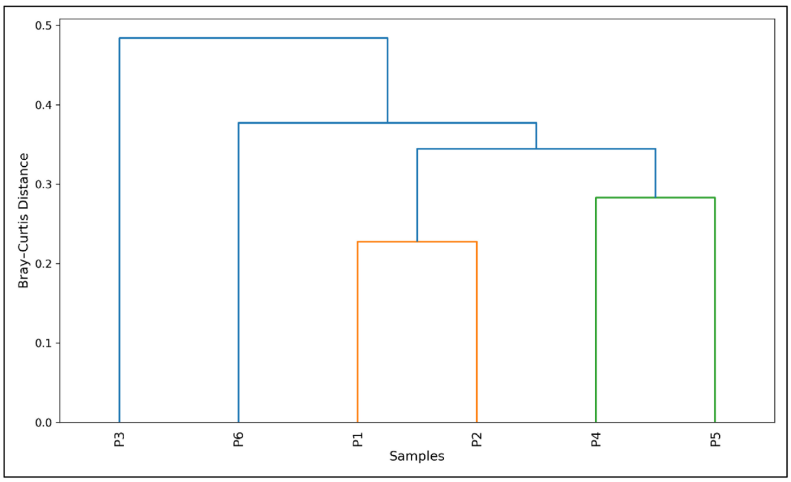

Figure 7. Hierarchical Clustering Dendrogram of Bacterial Communities Based on Bray–Curtis Dissimilarity.

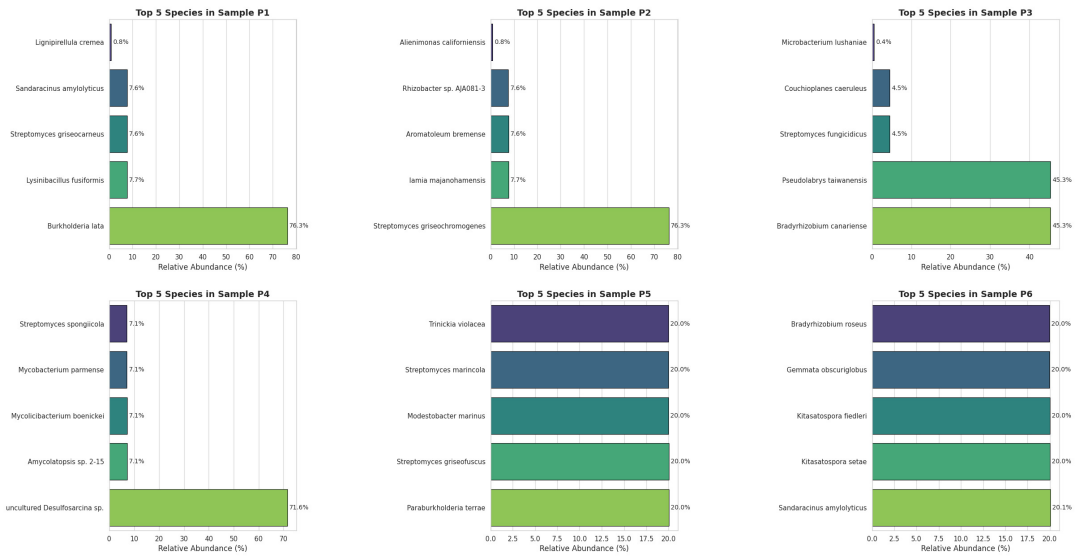

Figure 8. Relative Abundance of Microbial Species Identified in Samples P1 to P6.

4. Discussion

Despite the substantial sequencing depth-approximately 50 million reads per sample, the taxonomic classification rate remained low, with bacterial sequences accounting for only a small fraction of the total reads. These findings underscore both the richness of microbial life in tropical soils and the limitations of current reference databases, which fail to encompass the full extent of environmental microbial diversity. As noted by Uffelen et al. (2024), taxonomic classification of environmental metagenomes tends to be poor unless classifiers and databases are specifically tailored. James et al. (2022) further emphasize that while low classification rates in shotgun environmental datasets can be partially mitigated through the development of soil-focused reference sets and classifier optimization, significant challenges to accurate identification persist.

The fact that most of our samples reached a clear plateau consistent with findings from well-established studies in the literature (Caporaso et al., 2012), indicates that the sequencing depth employed was sufficient to capture the dominant bacterial diversity (Thompson et al., 2017; Weiss et al., 2017). Exceptions were observed in samples P3 and P6, which exhibited an upward trend in their rarefaction curves, suggesting a higher inherent ecological complexity in these environments, a phenomenon also reported in hyperdiverse microbial communities.

The observation that Actinomycetota and Pseudomonadota (formerly Actinobacteria and Proteobacteria) were the dominant bacterial phyla aligns with extensive literature describing these groups as ubiquitous and functionally critical components of soil ecosystems. Their dominance is not random but reflects their specialized ecological roles (Fierer et al., 2007; Lladó et al., 2017). As reviewed by Bulgarelli et al. (2013), Actinomycetota play a key role in the decomposition of recalcitrant organic matter and the production of antibiotics, while Pseudomonadota, a metabolically versatile phylum, include numerous members involved in nitrogen cycling, plant growth promotion, and root interactions. The consistency between our taxonomic profile and these reference studies supports the validity of our sampling strategy and bioinformatic analysis. However, as our results demonstrate, the identity of dominant taxa tells only part of the story. The contrast between samples P5 (a balanced community) and P6 (a simplified, low-diversity community) illustrates how distinct environmental conditions can profoundly shape community structure, even when derived from a similar taxonomic pool. This ecological interpretation—linking microbial identity to function and community structure, is a central principle of modern microbial ecology, as articulated in foundational reviews of the field (Bulgarelli et al., 2013).

The high similarity between samples P1 and P2, as indicated by their low Bray-Curtis dissimilarity value (0.228), suggests that these sites are subject to analogous environmental conditions, such as pH, moisture, or soil type, that act as filters in selecting a shared microbial assemblage. This pattern supports the idea that geographically proximate locations tend to cluster together, reflecting similarities in bacterial communities due to spatial proximity and potentially shared ecological drivers (Proctor et al., 2020; Horst et al., 2023). Similarly, samples P4 and P5 formed a coherent cluster (dissimilarity = 0.283), indicating moderate similarity in community structure. In contrast, sample P3 exhibited the highest dissimilarity values across comparisons (up to 0.527), pointing to the presence of distinct ecological factors or divergent selective pressures. Sample P6, although not the most dissimilar, remained ecologically isolated, with consistently elevated dissimilarity values (0.368-0.391) and no tight clustering, reinforcing its unique microbial profile (Zhu et al., 2025).

The ordination results further support these patterns. Principal Coordinates Analysis (PCoA) based on Bray–Curtis distances explained 82.2% of the total variance under TSS normalization (PCoA1 = 49.3%, PCoA2 = 32.9%), and 57.9% under Hellinger transformation (PCoA1 = 32.7%, PCoA2 = 25.2%). The higher variance explained by TSS reflects stronger ecological structuring and greater sensitivity to compositional variation, as confirmed by the mean dissimilarity values (TSS = 0.3879, Hellinger = 0.2875). These results suggest that TSS normalization better captures ecological gradients and enhances sample discrimination. It is imperative to compute these elements for your datasets to ensure accurate interpretations of ecological patterns, since the Bray-Curtis index tends to exaggerate the magnitude component (Greenacre, 2017).

The dendrogram, which displays continuous variation in taxonomic composition with samples clustering based on gradual similarity, reinforces the notion that bacterial community assembly is not an abrupt process. This pattern aligns with the concept of a "community continuum," in which microbial composition shifts smoothly along environmental gradients. Such gradients are often governed by a combination of factors including substrate type, land use, and microclimatic conditions, which vary spatially in a continuous manner (Dini-Andreote et al., 2015; Idbella et al., 2025). The positioning of sample P6 as an ecological outlier does not contradict this model; rather, it supports it by indicating that this sample lies at one end of the environmental gradient, exposed to markedly different conditions than the others.

Thus, the hierarchical clustering validates that community assembly within this agroecosystem is a filtering and gradual process, rather than a stochastic or categorical one. Single metrics are insufficient to capture the complexity of interactions and assembly processes in understudied environments (Lu et al., 2023). The combined application of high-resolution sequencing, robust normalization strategies, and a suite of ecological metrics was crucial for uncovering gradient patterns and subtle heterogeneity that might otherwise remain undetected (Wu et al., 2024).

Actinobacteria play an important role in the decomposition of organic matter. Their abundance and distribution can reflect a good level of soil fertility as well as biological activity (Mitra et al., 2022). The microbes associated with the plant rhizosphere are termed as rhizospheric microbes and rhizospheric actinobacteria are most dominant in nature. The study of different microbial diversity in form of plant microbiomes it can be concluded the members of phylum actinobacteria has been reported from different genera such as *Streptomyces*, *Sanguibacter*, *Rhodococcus*, *Pseudonocardia*, *Propionibacterium*, *Nocardia*, *Mycobacterium*, *Micrococcus*,

Microbacterium, *Frankia*, *Corynebacterium*, *Clavibacter*, *Cellulomonas*, *Bifidobacterium*, *Arthrobacter*, *Actinomyces*, and *Acidimicrobium* (Yadav et al., 2018).

Actinobacteria have been representing a large portion of soil microbiomes in the plant root systems (Ali et al., 2016). The Actinobacteria has been isolated from diverse study sources e.g. chickpea (*Cicer arietinum*) (Elsayed et al., 2024), Sugarcane (*Saccharum officinarum*) (Chen et al., 2024), Sunflower (*Helianthus annuus*) (Mudassar et al., 2024) and wheat (*Triticum aestivum*) (Budi et al., 2024). Actinomycetes represent important microbial communities in soils and inhabit a very high proportion of soil microbial biomass that has the capacity to produce a wide range of high-value antibiotics, organic acids, phytohormones, extracellular enzymes, bioactive compounds, and non-antimicrobial secondary metabolites (Singh et al., 2018).

The influence of pseudomycota soil cultures, particularly through the application of beneficial microorganisms such as *Pseudomonas* and arbuscular mycorrhizal fungi, significantly enhances the growth and yield of cocoa plants. These microorganisms improve soil fertility, promote nutrient uptake, and enhance plant resilience against pathogens, leading to improved overall plant health and productivity. Thus, the presence of pseudomonads caused an increase in dry mass and leaf area by 18.3% and 37.8%, respectively, indicating improved plant vigor (Negrín et al., 2022).

Analysis of microbial communities across samples P1 to P6 revealed distinct patterns of taxonomic abundance. The high abundance of *Burkholderia lata* in sample P1 (9,925 reads) may reflect genuine ecological dominance, potentially linked to its strong rhizospheric competence, root colonization capacity, and production of antifungal metabolites. Species of *Burkholderia* are known to promote plant growth and increase leaf polyphenol content by enhancing potassium availability in the soil. This occurs through the solubilization of insoluble potassium compounds, which are subsequently absorbed by roots and translocated to leaves, resulting in improved growth and elevated concentrations of beneficial compounds (Zhang et al., 2022).

The recurrence of *Streptomyces* species in samples P1, P2, P3, P4, and P5 underscores their broad distribution and ecological relevance in the studied environments. The population size and activity of *Streptomyces* tend to increase in soil compared to other bacterial groups due to their ability to degrade plant and animal residues, as well as materials such as cotton, plastics, rubber, paper, cellulose, lignocellulose, chitin, and various organic compounds. Remarkably, they can survive and recover even from the C horizon of the soil (Hasani et al., 2014). Through these functions, *Streptomyces* influence their environment by modulating biogeochemical cycles, shaping soil community structure, and interacting with other microorganisms (Bontemps et al., 2013; Donald et al., 2022).

Other genera identified include *Bradyrhizobium*, present in samples P3 (*Bradyrhizobium canariense*) and P6 (*Bradyrhizobium roseus*), and *Paraburkholderia*, found in sample P5 (*Paraburkholderia terrae*). Both are directly involved in biological nitrogen fixation. The genus *Bradyrhizobium* employs multiple nitrogen-fixing mechanisms in symbiosis with host plants, primarily through the formation of root nodules and the activity of nitrogenase enzymes. This symbiotic relationship is initiated by the exchange of signaling molecules between the bacterium and the plant, triggering nodule formation and the differentiation of bacteria into nitrogen-fixing bacteroids (Likic et al., 2024; Mahapatra et al., 2022). The genus *Paraburkholderia* also holds promise for sustainable agricultural applications, including nitrogen fixation, regulation of plant hormones, and detoxification of xenobiotics. These traits suggest that related species may contribute to plant growth and beneficial plant–microbe interactions. However, further functional studies are required to confirm the safety and efficacy of these strains in agricultural contexts (Herpell et al., 2020).

Less frequent species such as *Gemmata obscuriglobus* (P6) and *Lignipirellula cremea* (P1) indicate significant phylogenetic diversity. The presence of *Gemmataceae* in both environments suggests that this bacterial group is ubiquitous and resilient. Previous studies have identified *Gemmataceae* in a variety of habitats, including wetlands, wastewater treatment plants, peat bogs, and acidic marshes (Rakitin et al., 2021). These findings, combined with our results, suggest that the genus *Gemmata* may possess resistance mechanisms that enable survival under diverse environmental stressors. Interestingly, *Gemmataceae* also thrive in nutrient-rich soils, which may explain their higher abundance in uncontaminated environments (Camargo et al., 2022; Peprah et al., 2025).

5. Conclusion

The predominance of the phyla Actinomycetota and Pseudomonadota confirms ecological patterns widely documented in the literature, reinforcing the functional roles of these groups in organic matter decomposition, antibiotic production, plant growth promotion, and nutrient cycling. The structure of bacterial communities varied significantly across samples, with notable contrasts between P5 (a balanced community) and P6 (a simplified and isolated community), highlighting the influence of local environmental conditions on microbial community assembly. The similarity observed between samples P1 and P2, as well as between P4 and P5, suggests that factors such as pH, moisture, and soil type act as ecological filters, promoting similar microbial assemblages in geographically proximate locations. The application of robust ecological metrics, including Bray-Curtis dissimilarity and ordination analyses (PCoA), enabled the detection of subtle environmental gradients and continuous patterns of community assembly, consistent with the concept of a "community continuum." Sample P6, positioned as an ecological outlier, reinforces this interpretation by representing one end of the environmental gradient. Ultimately, the data demonstrate that the cacao soil microbiota comprises functionally relevant taxa, such as actinobacteria and pseudomonads, which directly contribute to soil fertility, pathogen suppression, and plant vigor. These findings not only validate the sampling and analytical strategy employed but also provide a foundation for the development of sustainable agricultural practices based on soil microbiota management.

Acknowledgements

To the Foundation of Amazon for the Support of Studies and Research - FAPESPA, for financing the Mocajuba Islands Cocoa project; to the Mocajuba Islands Cocoa Project, for the financial support provided for this work; to the Coordination for the Improvement of Higher Education Personnel – CAPES, for granting a doctoral scholarship to the first author; to the Emílio Goeldi Museum of Pará - MPEG/PA, for providing the structural support necessary for conducting the analyses.

Data Availability Statement

Data is available at: https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA1224407.

References

ADEJUMO, T.O. and ADEJORO, D.O., 2015. Assessment of bacteria, fungi and protozoa in three theobroma cacao soils in Ondo State, Nigeria. *International Journal of Sciences*, vol. 7, pp. 28-33. https://doi.org/10.18483/ijSci.768.

ALBUQUERQUE, K.C., CORRÊA, G.R. and CAMPOS, P.V., 2023. Ethnoknowledge in the Amazon: traditional agriculture in the middle amazon. *Revista Verde Grande: Geografia e Interdisciplinaridade*, vol. 5, no. 2, pp. 487-508. https://doi.org/10.46551/rvg26752395202324875O8.

ALI, A., BASHIR, U., AKHTAR, N., AKHTAR, N. and HAIDER, M.S., 2016 [viewed 4 August 2025]. Characterization of growth promoting rhizobacteria of leguminous plants. *Pakistan Journal of Phytopathology* [online], vol. 28, no. 1, pp. 57-60. Available from: https: //pjp.pakps.com/index.php/PJP/ article/view/ 201/125

ANDRIUZZI, W.S. and WALL, D.H., 2017. Responses of belowground communities to large aboveground herbivores: meta-analysis reveals biome-dependent patterns and critical research gaps. *Global Change Biology*, vol. 23, no. 9, pp. 3857-3868. https://doi.org/10.1111/gcb.13675. PMid:28245090.

BARDGETT, R. and VAN DER PUTTEN, W., 2014. Belowground biodiversity and ecosystem functioning. *Nature*, vol. 515, no. 7528, pp. 505-511. https://doi.org/10.1038/nature13855. PMid:25428498.

BENAUD, N., CHELLIAH, D.S., WONG, S.Y. and FERRARI, B.C., 2022. Soil substrate culturing approaches recover diverse members of Actinomycetota from desert soils of Herring Island, East Antarctica. *Extremophiles : Life Under Extreme Conditions*, vol. 26, no. 2, pp. 24. https://doi.org/10.1007/s00792-022-01271-2. PMid:35829965.

BONTEMPS, C., TOUSSAINT, M., REVOL, P.-V., HOTEL, L., JEANBILLE, M., UROZ, S., TURPAULT, M.-P., BLAUDEZ, D. and LEBLOND, P., 2013. Diversidade taxonômica e funcional de *Streptomyces* em um solo florestal. *FEMS Microbiology Letters*, vol. 342, pp. 157-167. https://doi.org/10.1111/1574-6968.12126. PMid:23489323.

BUDI, S., LAILIYAH, W.N., PRIHATININGRUM, A.E. and SAMIDJO, G.S., 2024. Quality coefficient on gene differentiation and phenotype: clone assessment of *Saccharum officinarum* LINN. *Planta Tropika*, vol. 21, no. 2, pp. 170-185. https://doi.org/10.18196/pt.v12i2.22232.

BULGARELLI, D., SCHLAEPPI, K., SPAEPEN, S., VAN THEMAAT, E.V.L. and SCHULZE-LEFERT, P., 2013. Structure and functions of the bacterial microbiota of plants. *Annual Review of Plant Biology*, vol. 64, pp. 807-838. https://doi.org/10.1146/annurev-arplant-050312-120106. PMid:23373698.

CAMARGO, A.P., DE SOUZA, R.S.C., JOSE, J., GERHARDT, I.R., DANTE, R.A., MUKHERJEE, S., HUNTEMANN, M., KYRPIDES, N.C., CARAZZOLLE, M.F. and ARRUDA, P., 2022. Plant microbiomes harbor potential to promote nutrient turnover in impoverished substrates of a Brazilian biodiversity hotspot. *The ISME Journal*, vol. 17, no. 3, pp. 354-370. https://doi.org/10.1038/s41396-022-01345-1. PMid:36536072.

CAPORASO, J.G., LAUBER, C.L., WALTERS, W.A., BERG-LYONS, D., HUNTLEY, J., FIERER, N., OWENS, S.M., BETLEY, J., FRASER, L., BAUER, M., GORMLEY, N., GILBERT, J.A., SMITH, G. and KNIGHT, R., 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal*, vol. 6, no. 8, pp. 1621-1624. https://doi.org/10.1038/ismej.2012.8. PMid:22402401.

CHAMBOULEYRON, R. and ARENZ, K.H., 2021. Atlântico Amazônico: cacau, expansão colonial e trabalho indígena na região Amazônica portuguesa (séculos XVII e XVIII). *Journal of Latin American Studies*, vol. 53, no. 2, pp. 221-244. https://doi.org/10.1017/S0022216X21000213.

CHEN, S., FENG, X., ZHANG, Q., HUA, X., ZHANG, Q., CHEN, C., LI, J., LIU, X., WENG, C., CHEN, B., ZHANG, M., YAO, W., TANG, H., MING, R. and ZHANG, J., 2024. ScDB: A comprehensive database dedicated to *Saccharum*, facilitating functional genomics and molecular biology studies in sugarcane. *Plant Biotechnology Journal*, vol. 22, no. 12, pp. 3386-3388. https://doi.org/10.1111/pbi.14457. PMid:39213302.

DINI-ANDREOTE, F., STEGEN, J.C., VAN ELSAS, J.D. and SALLES, J.F., 2015. Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 11, pp. E1326-E1332. https://doi.org/10.1073/pnas.1414261112. PMid:25733885.

DONALD, L., PIPITE, A., SUBRAMANI, R., OWEN, J., KEYZERS, R.A. and TAUFA, T., 2022. *Streptomyces*: ainda o maior produtor de novos metabólitos secundários naturais, uma perspectiva atual. *Microbiology Research*, vol. 13, pp. 418-465. https://doi.org/10.3390/microbiolres13030031.

DUBREUIL, V., FANTE, K., PLANCHON, O. and NETO, J.L.S., 2018. Annual climate types in Brazil: an application of the Köppen classification from 1961 to 2015. *Confins*, vol. 37. https://doi.org/10.4000/confins.15738.

ELSAYED, H., ABDALLA, N. and AMER, S., 2024. Actinobacteria as alternative agents for sustainable agriculture. *Egyptina Journal of Pure and Applied Science*, vol. 62, no. 2, pp. 52-64. https://doi.org/10.21608/ejaps.2024.282093.1093.

FIERER, N., BRADFORD, M.A. and JACKSON, R.B., 2007. Toward an ecological classification of soil bacteria. *Ecology*, vol. 88, no. 6, pp. 1354-1364. https://doi.org/10.1890/05-1839. PMid:17601128.

FURTAK, K. and GALAZKA, A., 2019. Edaphic factors and their influence on the microbiological biodiversity of the soil environment. *Postępy Mikrobiologii*, vol. 58, no. 4, pp. 375-384. https://doi.org/10.21307/PM-2019.58.4.375.

GREENACRE, M., 2017. 'Size' and 'shape' in the measurement of multivariate proximity. *Methods in Ecology and Evolution*, vol. 8, no. 11, pp. 1415-1424. https://doi.org/10.1111/2041-210X.12776.

HASANI, A., KARIMINIK, A. and ISSAZADEH, K., 2014. Estreptomicetos: características e suas atividades antimicrobianas. *International Journal of Advanced Biological and Biomedical Research*, vol. 2, pp. 63-75.

HERNANDEZ, R., VELAZQUEZ, I., OROZCO, M. and SANTOYO, G., 2010. Soil Metagenomics: new challenges and biotechnological

opportunities. *Phyton*, vol. 79, no. 1, pp. 133-139. https://doi.org/10.32604/phyton.2010.79.133.

HERPELL, J.B., SCHINDLER, F., BEJTOVIĆ, M., FRAGNER, L., DIALLO, B., BELLAIRE, A., KUBLIK, S., FOESEL, B.U., GSCHWENDTNER, S., KEROU, M., SCHLOTER, M. and WECKWERTH, W., 2020. The potato yam phyllosphere ectosymbiont *Paraburkholderia* sp. Msb3 is a potent growth promotor in tomato. *Frontiers in Microbiology*, vol. 11, pp. 581. https://doi.org/10.3389/fmicb.2020.00581. PMid:32373084.

HOLMES, K.A., SCHROERS, H., THOMAS, S.E., EVANS, H.C. and SAMUELS, G.J., 2004. Taxonomy and biocontrol of a new species of Trichoderma from the Amazon basin of South America. *Mycological Progress*, vol. 3, no. 3, pp. 199-210. https://doi.org/10.1007/s11557-006-0090-z.

HORST, A.M.T., FUDYMA, J.D., SONES, J.L. and EMERSON, J.B., 2023. Dispersal, habitat filtering, and eco-evolutionary dynamics as drivers of local and global wetland viral biogeography. *The ISME Journal*, vol. 17, no. 11, pp. 2079-2089. https://doi.org/10.1038/s41396-023-01516-8. PMid:37735616.

IDBELLA, M., IACOMINO, G., ABD-ELGAWAD, A.M. and BONANOMI, G., 2025. Soil microbial co-occurrence networks across climate and land use gradient in Southern Italy. *Environmental Microbiology Reports*, vol. 17, no. 2, pp. e70093. https://doi.org/10.1111/1758-2229.70093. PMid:40210220.

JAMES, M.T., FARRISI, S.T., SHAH, S. and SHAH, V., 2022. Identificação dos principais organismos envolvidos no ecossistema nutricional no solo ácido da Pensilvânia, EUA. *Frontiers in Environmental Science*, vol. 10. https://doi.org/10.3389/fenvs.2022.766302.

KAALE, S., MACHANGU, R.S. and LYIMO, T.J., 2023. Molecular characterization and phylogenetic diversity of actinomycetota species isolated from sediments of Lake Natron in Arusha, Tanzania. *Microbiological Research*, vol. 278, pp. 127543. https://doi.org/10.1016/j.micres.2023.127543. PMid:37950928.

BIOINFORMATICS WORKBOOK, 2025 [viewed 9 November 2025]. *Metagenomics on nematodes* [online]. Kraken2. Available from: https://bioinformaticsworkbook.org/dataAnalysis/Metagenomics/Kraken.html

LIKIC, C., MAHER, E., RAE, A.E., ZHANG, R.Y. and MATHESIUS, U., 2024. Nodulation. In: T. BEECKMAN and A. ESHEL, eds. *Plant Roots*. Boca Raton: CRC Press. https://doi.org/10.1201/b23126-36.

LLADÓ, S., LÓPEZ-MONDÉJAR, R. and BALDRIAN, P., 2017. Forest soil bacteria: diversity, involvement in ecosystem processes, and response to global change. *Microbiology and Molecular Biology Reviews : MMBR*, vol. 81, no. 2, pp. e00063-e16. https://doi.org/10.1128/MMBR.00063-16. PMid:28404790.

LU, J., BREITWIESER, F.P., THIELEN, P. and SALZBERG, S.L., 2017. Bracken: estimating species abundance in metagenomics data. *PeerJ. Computer Science*, vol. 3, pp. e104. https://doi.org/10.7717/peerj-cs.104. PMid:40271438.

LU, Z., ZHOU, Y., LI, Y., LI, C., LU, M., SUN, X., LUO, Z., ZHAO, J. and FAN, M., 2023. Effects of partial substitution of chemical fertilizer with organic manure on the activity of enzyme and soil bacterial communities in the mountain red soil. *Frontiers in Microbiology*, vol. 14, pp. 1234904. https://doi.org/10.3389/fmicb.2023.1234904. PMid:37736094.

LUANDANBIO, 2025 [viewed 9 November 2025]. *luanferreirabio-oss/decontamination: v1.0 (v1.0)*. Zenodo. https://doi.org/10.5281/zenodo.17498295.

MAHAPATRA, R., MISHRA, P. and PATEL, Z.M., 2022. The molecular architecture of rhizobium–plant symbiosis in nitrogen fixation. In: V. SHARMA, R. SALWAN, E. MOLISZEWSKA, D. RUANO-ROSA and M. JĘDRYCZKA, eds. *The chemical dialogue between plants and beneficial microorganisms*. Cambridge, MA: Academic Press, pp. 137-144. https://doi.org/10.1016/B978-0-323-91734-6.00006-5.

MCMURDIE, P.J., 2022 [viewed 9 November 2025]. *phyloseq: an R package for reproducible interactive analysis of microbiome census data* [software]. GitHub. Available from: https://github.com/joey711/phyloseq

MISHRA, P., MAURYA, R., AVASHTHI, H., MITTAL, S., CHANDRA, M. and RAMTEKE, P.W., 2022. Genome assembly and annotation. In: D.B. SINGH and R.K. PATHAK, eds. *Bioinformatics*. Cambridge, MA: Academic Press, pp. 49-66. https://doi.org/10.1016/B978-0-323-89775-4.00013-4.

MITRA, B., MONDAL, R., KHOSHRU, B., SENAPATI, A., RADHA, T.K., MAHAKUR, B., UNIYAL, N., MYO, E.M., BOUTAJ, H., GUERRA SIERRA, B.E., PANNEERSELVAM, P., GANESHAMURTHY, A.N., ELKOVIĆ, S., VASIĆ, T., RANI, A., DUTTA, S. and DAS MOHAPATRA, P.K., 2022. Actinobacteria-enhanced plant growth, nutrient acquisition, and crop protection: advances in soil, plant, and microbial multifactorial interactions. *Pedosphere*, vol. 32, no. 1, pp. 149-170. https://doi.org/10.1016/S1002-0160(21)60042-5.

MUDASSAR, M., RIAZ, U., AWAN, H.I.U., MAQSOOD, U. and JUNAID, M.A., 2024. Sugarcane cultivation amidst climate change challenges: an in-depth review of mitigation strategies and environmental resilience. *Global Academic Journal of Agriculture and Biosciences*, vol. 6, no. 4, pp. 90-99. https://doi.org/10.36348/gajab.2024.v06i04.002.

NEGRÍN, Y.F., GONZÁLEZ, N.O.G. and GÓMEZ, N.O., 2022. Pseudomona sp en la emergencia de semillas y el desarrollo de posturas de cacao. *Agronomia Costarricense*, vol. 46, no. 1. https://doi.org/10.15517/rac.v46i1.49873.

OLIVERIO, A.M., BISSETT, A., MCGUIRE, K.L., SALTONSTALL, K., TURNER, B.L. and FIERER, N., 2020. The role of phosphorus limitation in shaping soil bacterial communities and their metabolic capabilities. *mBio*, vol. 11, no. 5, pp. e01718-e01720. https://doi.org/10.1128/mBio.01718-20. PMid:33109755.

PELICICE, F.M., PEREIRA, H.R., CARVALHO, P., DAS CHAGAS, D.B. and LIMA-JUNIOR, D.P., 2025. Tocantins and Araguaia. In: M.A.S. GRAÇA, M. CALLISTO, F.T. MELLO, D. RODRÍGUEZ-OLARTE, eds. *Rivers of South America*. Amsterdam: Elsevier, pp. 407-435. https://doi.org/10.1016/B978-0-12-823429-7.00019-7.

PEPRAH, S., ADDO-FORDJOUR, P., FEI-BAFFOE, B., BOAMPONG, K., AVICOR, S.W. and DAMSERE-DERRY, J., 2025. Effects of pesticide application on soil bacteria community structure in a cabbage-based agroecosystem in Ghana. *PLoS One*, vol. 20, no. 5, pp. e0323936. https://doi.org/10.1371/journal.pone.0323936. PMid:40440308.

PIERSON, M.D. and SMOOT, L.M., 2001. Indicator microorganisms and microbiological criteria. In: MP DOYLE, L.P. BEUCHAT and T.J. MANIVILLE, eds. *Fundamentals of food microbiology*. 2nd ed. Washington, DC: American Society for Microbiology, pp. 71-78.

PROCTOR, D.M., SHELEF, K.M., GONZALEZ, A., DAVIS, C.L., DETHLEFSEN, L., BURNS, A.R., LOOMER, P.M., ARMITAGE, G.C., RYDER, M.I., MILLMAN, M.E., KNIGHT, R., HOLMES, S.P. and RELMAN, D.A., 2020. Microbial biogeography and ecology of the mouth and implications for periodontal diseases. *Periodontology 2000*, vol. 82, no. 1, pp. 26-41. https://doi.org/10.1111/prd.12268. PMid:31850642.

RAKITIN, A.L., NAUMOFF, D.G., BELETSKY, A.V., KULICHEVSKAYA, I.S., MARDANOV, A.V., RAVIN, N.V. and DEDYSH, S.N., 2021. Sequência completa do genoma da planctomiceta celulolítica *Telmatocola sphagniphila* SP2T e caracterização da primeira enzima celulolítica de planctomicetas. *Systematic and Applied Microbiology*, vol. 44, no. 6, pp. 126276. https://doi.org/10.1016/j.syapm.2021.126276. PMid:34735803.

RAMIREZ, K.S., LEFF, J.W., BARBERÁN, A., BATES, S.T., BETLEY, J., CROWTHER, T.W., KELLY, E.F. and OLDFIELD, E.E., 2014. Biogeographic patterns in belowground diversity in New York City's Central Park are similar to those observed globally. *Proceedings of the Royal Society B: Biological Sciences*, vol. 281, no. 1795, pp. 20141988. https://doi.org/10.1098/rspb.2014.1988.

RODRIGUES, R.R., LIMA, R.A., GANDOLFI, S. and NAVE, A.G., 2009. On the restoration of high diversity forests: 30 years of experience in the Brazilian Atlantic Forest. *Biological Conservation*, vol. 142, no. 6, pp. 1242-1251. https://doi.org/10.1016/j.biocon.2008.12.008.

SHAMIM, K., SHARMA, J. and DUBEY, S.K., 2017. Rapid and efficient method to extract metagenomic DNA from estuarine sediments. *3 Biotech*, vol. 7, no. 3, pp. 182. https://doi.org/10.1007/s13205-017-0846-y. PMid:28664369.

SINGH, S., PATIL, H., PRABHA, P., YANDIGERI, M. and PRASAD, S., 2018. Actinomycetes as potential plant growth-promoting microbial communities. In: R. PRASAD, S.S. GILL and N. TUTEJA, eds. *Crop improvement through microbial biotechnology*. Amsterdam: Elsevier, pp. 27-38. https://doi.org/10.1016/B978-0-444-63987-5.00002-5.

THOMPSON, L.R., SANDERS, J.G., MCDONALD, D., AMIR, A., LADAU, J., LOCEY, K.J., PRILL, R.J., TRIPATHI, A., GIBBONS, S.M., ACKERMANN, G., NAVAS-MOLINA, J.A., JANSSEN, S., KOPYLOVA, E., VÁZQUEZ-BAEZA, Y., GONZÁLEZ, A., MORTON, J.T., MIRARAB, S., ZECH XU, Z., JIANG, L., HAROON, M.F., KANBAR, J., ZHU, Q., JIN SONG, S., KOSCIOLEK, T., BOKULICH, N.A., LEFLER, J., BRISLAWN, C.J., HUMPHREY, G., OWENS, S.M., HAMPTON-MARCELL, J., BERG-LYONS, D., MCKENZIE, V., FIERER, N., FUHRMAN, J.A., CLAUSET, A., STEVENS, R.L., SHADE, A., POLLARD, K.S., GOODWIN, K.D., JANSSON, J.K., GILBERT, J.A., KNIGHT, R., and the EARTH MICROBIOME PROJECT CONSORTIUM, 2017. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*, vol. 551, no. 7681, pp. 457-463. https://doi.org/10.1038/nature24621. PMid:29088705.

UFFELEN, A.V., POSADAS, A., ROOSENS, N.H.C., MARCHAL, K., KEERSMAECKER, S.C.J. and VANNESTE, K., 2024. Benchmarking bacterial taxonomic classification using nanopore metagenomics data of several mock communities. *Scientific Data*, vol. 11, no. 1, pp. 864. https://doi.org/10.1038/s41597-024-03672-8. PMid:39127718.

WEISS, S., XU, Z.Z., PEDDADA, S., AMIR, A., BITTINGER, K., GONZALEZ, A., LOZUPONE, C., ZANEVELD, J.R., VÁZQUEZ-BAEZA, Y., BIRMINGHAM, A., HYDE, E.R. and KNIGHT, R., 2017. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*, vol. 5, no. 1, pp. 27. https://doi.org/10.1186/s40168-017-0237-y. PMid:28253908.

WU, W., HSIEH, C., LOGARES, R., LENNON, J.T. and LIU, H., 2024. Ecological processes shaping highly connected bacterial communities along strong environmental gradients. *FEMS Microbiology Ecology*, vol. 100, no. 12, pp. fiae146. https://doi.org/10.1093/femsec/fiae146. PMid:39479791.

YADAV, A.N., VERMA, P., KUMAR, S., KUMAR, V., SUGITHA, T.C.K., SINGH, B.P., SAXENA, A.K. and DHALIWAL, H.S., 2018. Actinobacteria from rhizosphere: molecular diversity, distributions and potential biotechnological applications. In: B.P. SINGH, V.K. GUPTA and A.K. PASSARI, eds. *New and future developments in microbial biotechnology and bioengineering*. Amsterdam: Elsevier, pp. 13–41. https://doi.org/10.1016/B978-0-444-63994-3.00002-3.

ZHANG, X., WANG, N., HOU, M., WU, H., JIANG, H., ZHOU, Z., CHANG, N., WANG, Q., WAN, X., JIANG, J., SHEN, Z. and LI, Y., 2022. Contribution of K solubilising bacteria (*Burkholderia* sp.) promotes tea plant growth (*Camellia sinensis*) and leaf polyphenols content by improving soil available K level. *Functional Plant Biology : FPB*, vol. 49, no. 3, pp. 283-294. https://doi.org/10.1071/FP21193. PMid:35101164.

ZHU, Z.M., ZHANG, Y., LI, W., GREENACRE, M., SAHA, S., SHI, Y. and ZHANG, L., 2025. Mathematical foundations of beta diversity: why common metrics fail in microbiome analysis. https://doi.org/10.1101/2025.08.14.670326.

Supplementary Material

Supplementary material accompanies this paper.
Legenda S1. Overview of read filtering results performed with Fastp.
Legenda S2. Distribution of decontamination results by sample.
This material is available as part of the online article from https://doi.org/10.1590/1519-6984.295836