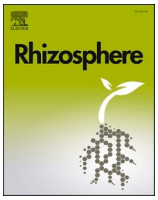
Rhizosphere 30 (2024) 100905

Contents lists available at ScienceDirect

Rhizosphere

journal homepage: www.elsevier.com/locate/rhisph



PGPg finder: A comprehensive and user-friendly pipeline for identifying plant growth-promoting genes in genomic and metagenomic data

Thierry Alexandre Pellegrinetti [a,*], Gabriel Gustavo Tavares Nunes Monteiro [a], Leandro Nascimento Lemos [b], Renato Augusto Corrêa dos Santos [c], Artur Gomes Barros [d], Lucas William Mendes [a]

[a] Cell and Molecular Biology Laboratory, Center for Nuclear Energy in Agriculture, University of São Paulo, Piracicaba, Brazil
[b] Ilum School of Science, Brazilian Center for Research in Energy and Materials (CNPEM), Campinas, São Paulo, Brazil
[c] Laboratory of Computational, Evolutionary, and Systems Biology, Center for Nuclear Energy in Agriculture, University of São Paulo, Piracicaba, São Paulo, Brazil
[d] Laboratory of Environmental Biotechnology and Genotoxicity, Federal University of Alfenas, Minas Gerais, Brazil

ARTICLE INFO

Keywords:
Plant-microbe interactions
Plant growth-promoting traits
Bioinformatics
Soil microbiome
Gene annotation

ABSTRACT

Identifying and comparing plant growth-promoting traits (PGPT) within whole-genome and metagenomic sequencing data can significantly advance agricultural research and promote sustainable crop production. This study introduces *PGPg finder*, a comprehensive pipeline designed to annotate and compare PGPT from both whole-genome and metagenome sequencing datasets. This pipeline utilizes direct sequence annotation alongside de novo assembly methods to accurately detect PGPT. By cross-referencing sequences from the PLaBAse database, it identifies and quantifies the presence of these genes within the original datasets, facilitating an intuitive comparison of the abundance and distribution of PGPT across various samples. We evaluated the performance of *PGPg finder* by analyzing genomes from five rhizobacterial strains: *Paenibacillus vini*, *Paenibacillus polymyxa*, *Fictibacillus* sp., *Brevibacillus agri,* and *Bacillus cereus*, and also metagenomic samples from bulk soils subjected to forest-to-pasture conversion in the Amazon rainforest. The genomic workflow revealed several genes associated with substrate utilization, abiotic stress neutralization, phosphate solubilization, and iron acquisition. It also identified genes unique to specific lineages, including those associated with colonization and plant-derived substrate usage in *P. polymyxa*, quorum sensing response and biofilm formation in *P. vini*, heavy metal detoxification and nitrogen acquisition in *B. agri*, and spore production and neutralizing biotic stress in *B. cereus*. The strain *Fictibacillus* sp. presented several unique genes related to surface attachment, stress response, xenobiotic degradation, phosphate solubilization, and phytohormone production. The use of *PGPg finder* highlights its potential to uncover novel inoculants and strains. The metagenomic workflow distinguished plant-growth promotion gene profiles between soils from the Amazon rainforest and pasture, with the latter showing a profile more aligned with simple carbohydrate consumption, abiotic stress tolerance, motility and chemotaxis, and phosphorus mineralization. Native forests exhibited a profile associated with the degradation of complex organic matter, oxidative stress tolerance, xenobiotic degradation, bactericidal activity, iron acquisition, and volatile pathways. These findings underscore the effectiveness and sensitivity of *PGPg finder* in accurately identifying and comparing PGPT genes, highlighting both commonalities and variations across samples. The application of this pipeline has the potential to significantly facilitate the identification of plant growth-promoting microbes.

1. Introduction

The relationship between plants and soil microorganisms is ancient and can be traced back to the terrestrialization process of ancestral plants and has accompanied its evolution and diversification (De Vries and Archibald, 2018; Heckman et al., 2001; Vandenkoornhuyse et al., 2015; Delaux and Schornack., 2021). Plants can transfer up to 40% of the carbon fixed through photosynthesis into the soil directly next to the roots (i.e., the rhizosphere), making this small zone one of the most resource-rich habitats of the soil (Bais et al., 2006; Pieterse et al., 2014;

* Corresponding author.
E-mail address: tpellegrinetti@usp.br (T.A. Pellegrinetti).

https://doi.org/10.1016/j.rhisph.2024.100905
Received 4 March 2024; Received in revised form 20 May 2024; Accepted 20 May 2024
Available online 23 May 2024

Gherardi and Sala, 2020). In this environment, several genera of microorganisms are not only able to consume the molecules exuded by the roots. Still, they can also provide essential services and resources to the plants, including enhanced water and mineral uptake, nitrogen fixation, and protection from pathogens (Lugtenberg and Kamilova, 2009). As a result, these rhizospheric microbes can alleviate a wide range of (a)biotic stresses, directly contributing to increased plant growth, commonly referred to as plant growth–promoting rhizobacteria (PGPR). Over the last few decades, the inoculation of different PGPRs has gained acceptance and prominence in agriculture as a biotechnological resource (Waltz, 2017; Santos et al., 2019). However, research has advanced from culture-dependent studies to culture-independent and high-throughput molecular approaches, such as amplicon sequencing (Langille et al., 2013), whole-genome sequencing (Jones and Good, 2016), and shotgun metagenomic and metatranscriptomic sequencing (Mendes et al., 2018). These techniques have shifted the view of the plant microbiome towards a more complex process with multiple pathways and genes involved (Wang et al., 2024). These tools are now widely used to characterize the genetic potential of the rhizosphere microbial communities (Kwak et al., 2018) and have changed the view of plant growth promotion into a more complex trait that is often the net result of changes in the whole rhizosphere microbial community structure and function (Carrión et al., 2019).

Whole-genome and metagenome sequencing are sophisticated molecular approaches able to unravel the structure, composition, and function of complex microbial communities through the identification of the potential pathways and genes of the rhizosphere microorganisms related to plant-growth promotion (Hu et al., 2022; Sun et al., 2021). Likewise, metatranscriptomic data can reveal processes occurring in the rhizosphere beyond the genetic potential, accurately reporting the active pathways and expressed genes (Mendes et al., 2018). Different studies have predicted rhizosphere microbial community functions by annotating predicted genes against several well-established databases. However, these functional databases are not dedicated to plant-growth promotion and lack a concise and curated list of genes related to plant-growth promotion.

Several databases have been established for other purposes, targeting specific genes associated with nitrogen, phosphorus, sulfur cycling, and methane (Qian et al., 2022; Tu et al., 2019; Yu et al., 2021; Zeng et al., 2022). Others can provide complete whole-genome annotation, including Prokka (Seemann, 2014), RASTtk (Brettin et al., 2015), and DRAM (Shaffer et al., 2020). Recently, the PLant-associated BActeria web resource (PLaBAse) has developed a vast database for genomic annotation and prediction of plant-associated microbial genes (Patz et al., 2021, 2024). However, despite being a valuable resource for predicting genes that promote plant growth, researchers frequently encounter the need to engage in time-consuming and intricate processes for assembly, annotation, and data visualization.

Here, we developed the software named *PGPg_finder*, which contains a comprehensive and easy-to-use pipeline to infer and compare plant-growth-promoting genes and pathways in whole-genome, metagenome, and metatranscriptome data. *PGPg_finder* greatly assists in annotating user data, summarizing and curating results, and directly using the PLaBAse database to identify known microbial-mediated plant growth promotion processes. As a modular pipeline, users can annotate genomes and metagenomes in three different workflows, and the output results consist of a summary of more than 800 relevant PGPTs. We also applied *PGPg_finder* to real-world metagenomic and whole-genome data, demonstrating this tool's actual and potential application for future studies. The broad-scale application and validation of *PGPg_finder* could accelerate future agricultural research and lead to sustainable practices.

2. Material and methods

2.1. The main procedure of PGPg_finder

PGPg_finder was implemented in Python 3.7.12 and Bash 5.1.16 and uses the bioinformatic tools in Table 1. The GitHub repository describes all functions and prerequisites for running the pipeline (https://github.com/tpellegrinetti/PGPg_finder). The main script "*PGPg_finder.py*" must be executed to initiate the analysis, serving as a critical link between the researcher and the desired analytical outcomes. By executing the command python "*PGPg_finder.py -h",* one can access the comprehensive functionalities and workflows integrated into the pipeline, along with instructions for its application. Initially, it is necessary to provide parameters such as the workflow (-w), the reads or assembly folder (-i), the output folder (-o), the number of threads (-t), and the mode of sensitive analysis (-m). The *PGPg_finder* pipeline encompasses three distinct and efficient workflows: *genome_wf*, *metafast_wf*, and *meta_wf*. The workflows were explicitly developed to address various research objectives, as illustrated in Fig. 1.

2.2. Genomic annotation of plant growth promotion traits (genome_wf)

The *genome_wf* workflow was engineered to analyze plant growth-promoting genes (PGPG) within genomic datasets through a structured and sequential process. This process starts by identifying genomic files in various formats, including *fasta*, *fa*, or *fna*, sourced from a specified directory. Following this initial step, the pipeline employs Prodigal for gene prediction (Hyatt et al., 2010). Then, DIAMOND v.2.1.8.162 is used as an alignment tool for sequence annotation through its *blastx* function, ensuring high efficiency (Buchfink et al., 2021). The command allows for the optimization of the workflow speed through the selection of multiple threads. DIAMOND facilitates rapid and accurate alignments against the genome database from PLaBAse – PGPT-db, accessible via the PLaBAse website (https://plabase.cs.uni-tuebingen.de/pb/download.php) (Patz et al., 2021, 2024). The outcome of the *genome_wf* workflow is quantified as the count of gene hits, indicating correspondences between the sequences in PGPT-db and the genomic sequences analyzed.

The ensuing stage involves processing and compilation of the results into comprehensive tables and heatmaps for enhanced interpretability. The results are organized into folders containing tables of both

Table 1
List of tools used in the PGPg_finder pipeline.

Tool	Task	Repository
Python v.3.10.12	Script Structure	
DIAMOND v.2.1.8	Sequence Annotation	https://github.com/bbuchfink/diamond
Prodigal v.2.6.3	Gene Prediction	https://github.com/hyattpd/Prodigal
Trimmomatic v.0.39	Quality Control	https://github.com/usadellab/Trimmomatic
PEAR v.0.9.6	Sequence Assembly	https://github.com/tseemann/PEAR
MEGAHIT v.1.2.9	Metagenome Assembly	https://github.com/voutcn/megahit
Bowtie v.2 2.5.1	Read Aligner	https://github.com/BenLangmead/bowtie2
BBMap v.39.01	Coverage Calculation	https://sourceforge.net/projects/bbmap/
biom-format v.2.1.15	Data Manipulation	https://github.com/biocore/biom-format
Pandas v.2.0.3	Data Manipulation	https://github.com/pandas-dev/pandas
Numpy v.1.25.0	Data Manipulation	https://github.com/numpy/numpy
Matplotlib v.3.7.1	Data Manipulation	https://github.com/matplotlib/matplotlib
Seaborn v.0.12.2	Data Visualization	https://github.com/mwaskom/seaborn

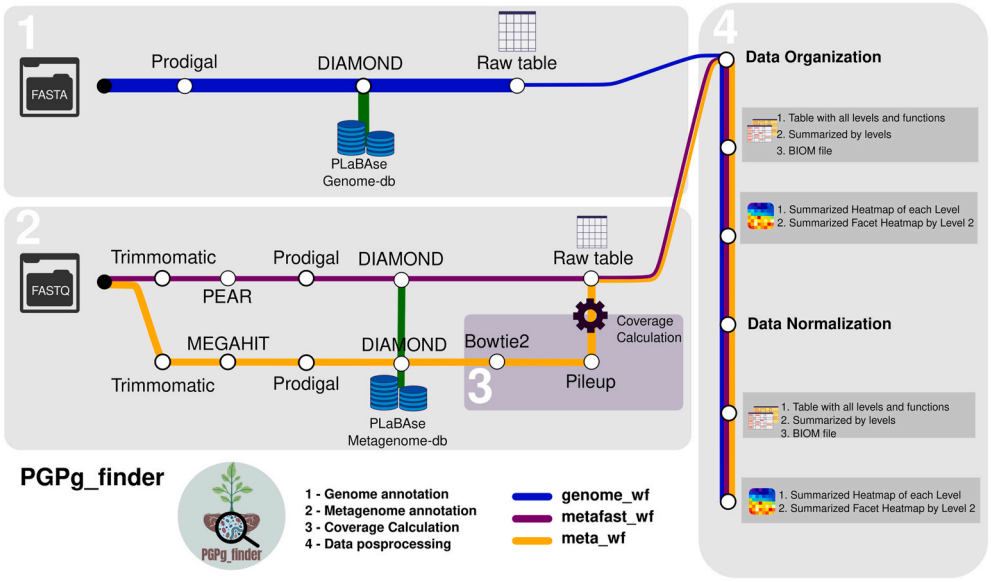


Fig. 1. A flowchart of the major steps involved in running PGP_finder. The genome_wf (A) uses genomes or metagenome-assembled genomes in the fasta format provided by the user to search for genes associated with plant-growth promotion trait (PGPT) genes using the PLaBAse database. The meta_wf and metafast_wf (B) involve the metagenomic assembly using MEGAHIT or unassembled reads to detect PGPT genes. After this step, PGPT genes will be organized in a summary table and plotted by using heatmaps with the seaborn Python library (C).

normalized and non-normalized matches. Normalization occurs through the ratio of PGPG counts to the total hit counts, providing a measure of gene relative abundance. This structured presentation of data facilitates straightforward interpretation and comparison, ensuring accessibility for further analytical endeavors.

2.3. Metagenomic annotation of plant growth promotion traits (metafast_wf and meta_wf)

Two distinct strategies have been formulated for sequence analysis within metagenomes in these workflows. The *metafast_wf* workflow provides a rapid approach to sequence annotation without requiring the assembly of metagenomic reads. In contrast, the *meta_wf* workflow adopts a more thorough but time-intensive strategy, encompassing read assembly and coverage assessment.

The *metafast_wf* workflow is designed for quick annotation of unassembled reads, eliminating the necessity for de novo assembly or coverage analysis. This expedited approach is suitable primarily for preliminary exploratory investigations. It accommodates various file formats, including *fastq*, *fq*, *fq.gz*, and *fastq.gz* for paired-end reads. The initial step starts by quality filtering the sequences with Trimmomatic v.0.39 (Bolger et al., 2014), following the merge of forward and reverse reads using PEAR v.0.9.6 (Zhang et al., 2013), resulting in unified sequences for subsequent annotation via the DIAMOND tool using the *blastx* function. Annotation within this workflow targets the *mgPGPT-db* database, featuring an extensive collection of genes from metagenome-assembled genomes and the AnnoTree database. This process yields a *gene_counts.txt* file, which is then utilized in Python scripts to compile and normalize data into tables for further analysis.

The *meta_wf* workflow focuses on precision and annotating reads assembled into contigs. Firstly, sequences underwent quality control with Trimmomatic. After this, MEGAHIT v1.2.9 is employed for the de novo assembly of metagenomic reads, offering an efficient and precise assembly process (Li et al., 2016). Users can input their assemblies using the -a parameter for flexibility when choosing assembly tools. Following assembly, sequences are predicted for genes using Prodigal and annotated with DIAMOND against the *mgPGPT-db* database using the *blastp* option. Additionally, nucleotide sequences are indexed with Bowtie2 and mapped back to the assembly, with gene abundance calculated via BBMap's pileup.sh script (Bushnell, 2014; Langmead and Salzberg, 2012). Integrating alignment outputs with gene abundance data is facilitated through Python scripts, culminating in a consolidated file for each gene's alignment and abundance within the sample.

Both workflows, *metafast_wf* and *meta_wf*, generate a preliminary gene_counts.txt file. This data undergoes further refinement and analysis, employing similar processing techniques to ensure comprehensive and interpretable results.

2.4. The use of PGPg_finder in rhizosphere-isolated genomes

The performance of the *PGPg_finder* was assessed by annotating PGPG in five strains extracted from the rhizosphere of common beans cultivated in Amazonian Dark Earth, as documented by (Pellegrinetti et al., 2023; Mendes et al., 2019). These strains, earmarked for potential bioinoculant use, exhibit numerous genes linked to plant growth promotion and pathogen resistance, notably against *Fusarium oxysporum*. The objective was to delineate the PGPG profile of each strain to gauge their suitability for bioinoculant applications. The genomic sequences include *Paenibacillus vini* (CENA-BCM001), *Paenibacillus polymyxa* (CENA-BCM002), *Fictibacillus* sp. (CENA-BCM004), *Brevibacillus agri* (CENA-BCM005), and *Bacillus cereus* (CENA-BCM007), hosted on the NCBI database under the bioproject PRJNA988090. This analysis was conducted on a server equipped with an Ubuntu 22.04 system, utilizing 30 threads, powered by an Intel(R) Xeon(R) CPU E5-2696 v4 @ 2.20 GHz with 44 CPUs and 128 GB of RAM. The comprehensive processing duration was meticulously logged by *PGPg_finder*, facilitating future comparative studies.

2.5. The use of PGPg_finder in Amazon rainforest and pasture metagenomes

To test our metagenome annotation pipeline, we evaluated the PGPG profile in eight metagenomes available in MG-RAST under the code *mgp83361*, which represents a study of Amazon rainforest under forest-to-pasture conversion (Pedrinho et al., 2019, 2020). We selected four samples of primary forest (F1 to F4) and four samples of pasture (F9 to F12). The native forest (NF) is located at 2°51′23.9″S, 54°57′28.4″W and is recognized as a well-maintained native forest, showing no evidence of

fire, logging, or other forms of disturbance. In contrast, the pasture (P) is located at 3°07′52.9″S, 54°57′28.1″W representing an area that was a pasture established over 21 years ago. This area was developed following the clearing of native vegetation through slash-and-burn methods. It was subsequently replanted with non-native grass species (*Urochloa* spp.), now serving as a grazing ground for extensive livestock farming.

Metagenome annotation's efficacy was tested by utilizing *metafast_wf* and *meta_wf* for comparison time purposes. The compressed metagenomic data accounted for 13.7 Gb, an average of 218M reads. The computational analysis was executed on the same server as described above. The total processing time was recorded by the log file generated by *PGPg_finder* upon the final run of the workflow.

2.6. Downstream analysis of PGPG metagenomic and genomic results

Data derived from metagenomic and genomic datasets were analyzed using R and RStudio alongside the SHAMAN pipeline for in-depth exploration (Volant et al., 2020). A particular focus was placed on visualizing the distribution of PGPG within isolated bacterial genomes through heatmaps generated on the SHAMAN platform, considering the quantity of PGPG annotated per genome as an indicator of functional richness. This metric was visually represented in bar plots for straightforward comparison. Furthermore, the investigation into orthologous clusters across bacterial genomes was facilitated by OrthoVenn3, highlighting the unique and common genes among the samples (Sun et al., 2023). Genes specific to each genome underwent a secondary annotation process via *PGPg_finder's genome_wf* workflow, with the results depicted in pie charts through ggplot2 for an intuitive understanding of genomic composition (Wickham, 2011).

In the realm of metagenomic analysis, PGPG annotations were scrutinized using both SHAMAN and R. The analysis extended to comparing PGPG abundance between primary forest and pasture samples, employing Principal Coordinate Analysis grounded in Bray-Curtis distance and substantiated by PERMANOVA to ascertain statistical significance at a threshold of $p < 0.05$. Additionally, the Wilcoxon test, applied at the same significance level, helped identify statistical discrepancies between the two environments, with outcomes illustrated in bar charts emphasizing log2 fold changes. PGPG diversity within the samples was quantified by the observed number of PGPG functions, indicative of functional richness, and through the Shannon Index to assess functional diversity. Differential abundance analysis of PGPG at the function level was conducted in STAMP and subsequently visualized in R, offering a nuanced perspective on the functional landscape presented by the PGPG annotations (Parks et al., 2014).

3. Results and discussion

3.1. General information of PGPg_finder

Here, we present the *PGPg_finder*, an innovative, easy-to-use pipeline developed to infer and contrast the prevalence of genes and pathways related to Plant-Growth Promotion Genes (PGPG) in both genomic and metagenomic datasets. As we grapple with global issues such as food security and climate change, the role of microorganisms in fostering plant growth is increasingly recognized in scientific research (Fadiji et al., 2022, 2022). PGPTs hold the potential to enhance crop yields and ameliorate soil fertility under stressful conditions (Fadiji et al., 2022). Thus, the need for efficient computational tools to identify and categorize PGPGs in genomes and metagenomes is growing, and the introduction of this pipeline serves as a crucial step toward fulfilling this need. The impact of *PGPg_finder* on sustainable agricultural practices could be profound.

The *PGPg_finder* is based on three workflows: *genome_wf*, *metafast_wf,* and *meta_wf*. *Genome_wf* is explicitly designed to detect PGPG in assemblies, a process following protein prediction and gene annotation through DIAMOND alignment. In contrast, *metafast_wf* provides a rapid method to annotate PGPG in metagenomic reads without de-novo assembly, albeit with a trade-off in precision compared to *meta_wf* workflow. The third, *meta_wf*, is a comprehensive, precise workflow encompassing de novo assembly, protein prediction, gene annotation, and mapping, with an integrated coverage calculation. The de novo assembly stage increases the workflow precision due to generating sizable contigs, thus minimizing errors compared to unassembled reads. Collectively, these workflows prioritize both speed and efficiency, utilizing Prodigal for protein prediction (except in *metafast_wf*) and employing a dedicated database, PLaBAse, for executing DIAMOND annotations (Patz et al., 2021, 2024).

In a representative run, we processed five genomes in 4 min, showcasing different PGPG annotated numbers (Table 2). The genomes presented contigs number between 16 and 111 and total length between 4.96 and 5.71 million of base pair. A total varying between 1543 and 1711 PGPG were identified in the genomes. CENA-BCM004 was the genome with the most PGPG found, while CENA-BCM001 had the lowest number of genes annotated. Considering the metagenome workflow, eight metagenomes were processed, totaling approximately 13.7 Gb, representing Primary Forest (PF) and Pasture (PS). The *metafast_wf* took 267 min, while *meta_wf* took 478 min. More than 218 million sequences were annotated with 5463 and 5186 PGPG in PF and PS, respectively (Table 2).

3.2. Rhizosphere bacterial isolates present a wide range of PGPG

Genomic analysis has revealed a broad spectrum of PGPG in rhizosphere-isolated bacteria, demonstrating significant functional diversity. These rhizobacteria enhance plant growth, pathogen control, and abiotic stress resistance, contributing to plant health and resilience through mechanisms like phosphate solubilization and phytohormone production (Leontidou et al., 2020; Saeed et al., 2021). The classification of PGPG into specific categories revealed variability within bacterial genomes, with critical functions including "Colonization-Plant Derived Substrate Usage", "Abiotic Stress Neutralization", "Phosphate Solubilization'', and "Iron Acquisition'' (Fig. 2A). Additionally, capabilities such as "Plant Vitamin Production'', "Cell Envelope Remodeling", "Universal Stress Response", and "Xenobiotics Biodegradation'' emphasize the complex roles of rhizobacteria in the ecosystem. These functionalities are essential for rhizobacterial survival and plant nutrition, as they rely on plant exudates and contribute to plant growth by synthesizing metabolites and solubilizing nutrients, significantly impacting plant health and growth metrics (Majeed et al., 2015; Mendes et al., 2013).

Enumerating and detecting PGPG in metagenomes is challenging due to database limitations, computational limitations, and available pipelines. Here, the variability and number of detected PGPG among the genomes were shown as a valuable metric for assessing bacterial potential for plant growth promotion (Fig. 2B; Suppl. Table 1). The Orthovenn3 analysis identified 1211 universally shared PGPG orthologous clusters across the bacterial strains, indicating a core functionality, while each strain also presented between 56 and 163 unique clusters, highlighting their distinct genetic diversity (Fig. 2C). Analyzing core and exclusive functions within bacterial genomes provides a reliable method for predicting inoculation success in plant hosts and identifying key traits such as abiotic and biotic resistance or enhancement of plant nutrition. For instance, many unique genes linked to "Colonization-Plant Derived Substrate Usage" were identified in CENA-BCM002 (*Paenibacillus polymyxa*). At the same time, functions related to "CE-Quorum Sensing Response and Biofilm Formation'' were found in CENA-BCM001 (*Paenibacillus vini*), indicating a direct relationship between these bacteria and their plant hosts. Research has already highlighted the effectiveness of *P. polymyxa* in forming biofilms (Timmusk et al., 2005). It produces biofilm polysaccharides that can antagonize pathogens (Timmusk et al., 2019) and absorb heavy metals (Govarthanan

Table 2
Comprehensive overview of various samples analyzed in the study, detailing their format, sequence number, average length, PGPG annotation, and the time taken for the analysis. The table includes samples from different environments and methods, such as Native Forest and Pasture. Additionally, it includes multiple samples of genomes and their corresponding PGPG annotations. The table also indicates the time taken for running each analysis, marked with asterisks, and a footnote provides clarifications on the workflows used for each sample: *metafast_wf*, *meta_wf*, and *genome_wf*.

Sample	Format	Sequence Number	Total Length (M)	Average length (bp)	PGPG Annotated	Time running (min)
Native Forest	FASTQ	120590180		101	5463	267[a]
Pasture	FASTQ	97401088		101	5186	478[b]
CENA-BCM001	FASTA	55	5.71	103,858.2	1543	4[c]
CENA-BCM002	FASTA	57	5.63	98,863	1661	
CENA-BCM004	FASTA	16	4.96	310,476.7	1711	
CENA-BCM005	FASTA	111	5.28	47,576.3	1632	
CENA-BCM007	FASTA	22	5.57	253,152.7	1603	

[a] Running in metafast_wf.
[b] Running with meta_wf.
[c] Running in genome_wf.

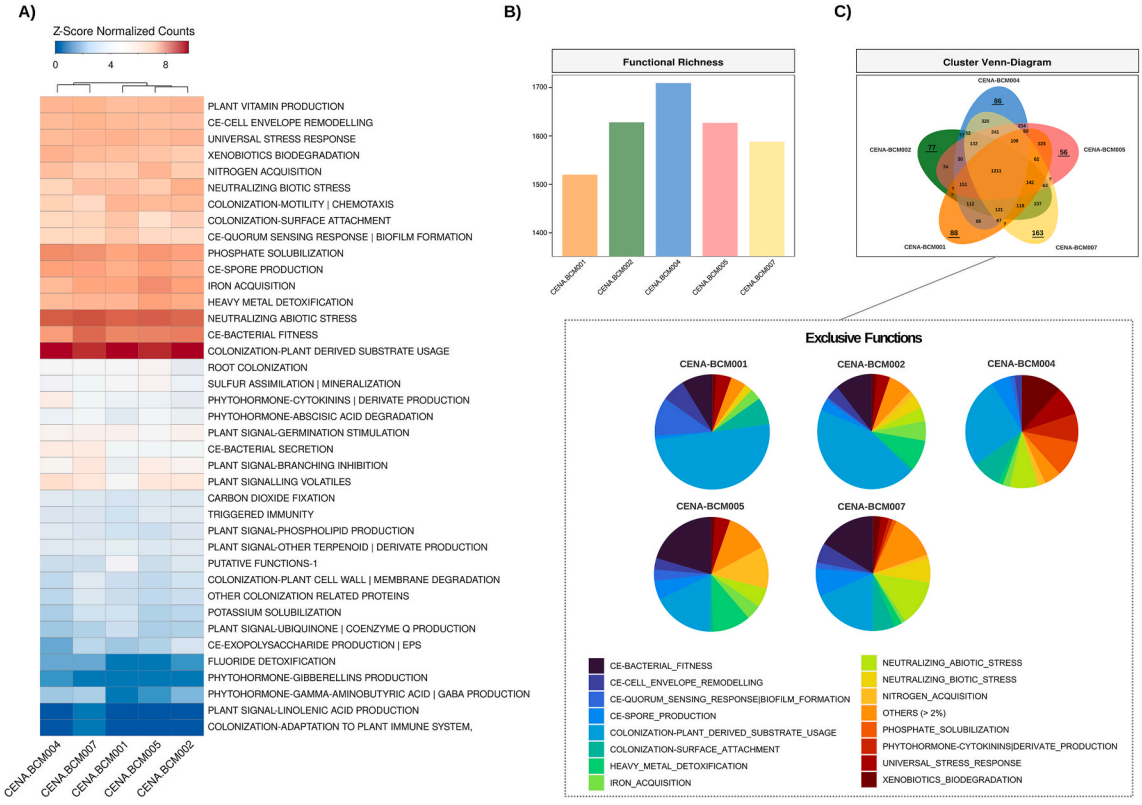


Fig. 2. Overview of Plant-Growth Promotion Gene (PGPG) Abundance in genomes of Strains isolated from the rhizosphere of common bean A) A Heatmap exhibiting a general view of PGPGs in the genomes. B) PGPG Richness among the genomes. C) A heatmap displaying the abundance of PGPG genes among the different strains.

et al., 2016). Moreover, a recent study demonstrated the effectiveness of *P. polymyxa* inoculation as a plant growth promoter by enhancing nitrogen and iron levels in common bean plants (da Cunha et al., 2024).

Another example was the CENA-BCM004 (*Fictibacillus* sp.), which stood out for its high number of unique PGPGs, including "Colonization Surface Attachment", "Universal Stress Response", "Xenobiotics Degradation", "Phosphate Solubilization'', and "Phytohormone-Cytokinins and Derivate Production". A recent study demonstrated the versatile metabolic and functional richness of the CENA-BCM004, which is classified as a new species from the *Fictibacillus* genus called *Fictibacillus terranigra* (Pellegrinetti et al., 2024). Despite limited studies on using these genera as bioinoculants, some studies have evaluated their potential as nematocidal control and arsenic resistance (Zheng et al., 2017). CENA-BCM05 (*Brevibacillus agri*) was notable for genes linked to "Heavy Metal Detoxification" and "Nitrogen Acquisition''. Interestingly, research with this genus corroborated our findings, illustrating the potential of *B. parabrevis* OZF5 in the remotion of metals such as Cr (VI) and Zn in experiments of bean growth in amended metal pots (Wani et al., 2023). Another study described the potential of *B. panacihumi* strain ZB1 to be used in biological treatment, where this strain was shown to remove nitrogen and heavy metals, including copper, selenium, and cadmium (Er et al., 2018). Lastly, CENA-BCM007 (*Bacillus cereus*) was characterized by unique PGPG associated with "Neutralizing Abiotic Stress", "CE-Spore Production", and "Neutralizing Biotic Stress", demonstrating the distinct functional capabilities and ecological niches these strains may occupy within the rhizosphere. This aligns with research describing the thermotolerance effect of *B. cereus* SA1 in soybeans and the production of antioxidant enzymes essential to the plant (Khan et al., 2020). Another study described the effectiveness of *B. cereus* YN917 used as a biocontrol agent since this strain presents biosynthetic gene cluster associated with plant promotion and antifungal compounds, such as IAA, tryptophan, siderophores, and

phenazine (Zhou et al., 2021).

3.3. Contrasting plant-growth promotion genes in different land uses

In metagenomic analysis, numerous methodologies and approaches for data analysis are available. However, the vast data size and the need for advanced bioinformatics skills and substantial computational resources can pose significant challenges. In this work, we quantified the PGPG across soils from forest-to-pasture conversion to explore their differential profiles. Our findings reveal substantial differences between native forests and pasture samples, with pastures exhibiting greater PGPG diversity and richness. This variation in PGPG is followed by shifts in the soil microbial communities, where forest-to-pasture conversion leads to altered abundance and composition of critical microbial phyla, impacting nutrient cycling and soil health (Mendes et al., 2015a, b; Navarrete et al., 2015; Pedrinho et al., 2019). The observed changes in PGPG and microbial diversity revealed the impact of deforestation on the Amazonian soil biodiversity and function, highlighting the need for sustainable land management to preserve biodiversity and ecosystem functions (Mendes et al., 2015a, b; Pedrinho et al., 2020; Venturini et al., 2022).

Forest transformation into pasture is the most prevalent form of land-use change in the Brazilian Amazon (Nascimento et al., 2019). Studies have shown that these changes adversely affect vegetation cover and biogeochemical cycles, impacting soil microbial communities (Mendes et al., 2015a, b; Pedrinho et al., 2023; Venturini et al., 2022). However, research on changes in PGPG during this transition remains limited and warrants further exploration. In native forest soils, a higher abundance of PGPG associated with crucial functions was found, including oxidative stress, xenobiotic degradation, bactericidal activity, root colonization, iron acquisition, and volatile pathways. These findings shed light on the hidden potential of previously unexplored taxa in Amazonian Forest soils as a promising candidate for biotechnological investigations. Research in Amazon soils has discovered new fungal isolates capable of degrading glyphosate herbicide and Benzo(a)pyrene and showing tolerance to polycyclic aromatic hydrocarbons (Correa et al., 2021; Souza et al., 2017). A study on the Amazon River microbiome revealed exclusive functions associated with the degradation of rainforest organic matter, xenobiotic biodegradation, and secondary metabolism (Santos-Júnior et al., 2020). These findings underscore the untapped potential of these underexplored ecosystems in terms of biotechnological applications.

At the functional level (Fig. 3B), some genes were more abundant in Amazon soils than in pastures. We could infer that some genes are related to organic matter decomposition, nutrient cycling, horizontal gene transfer, and genome plasticity. For instance, genes such as *cya*B, ABC transport, *livK*, *plc*, *and OAR1 highlight the metabolic versatility required for decomposing complex plant-derived organic compounds and* nutrient recycling in forest soils. To decompose and use the carbon deposited through litter deposition, the microbial communities of the

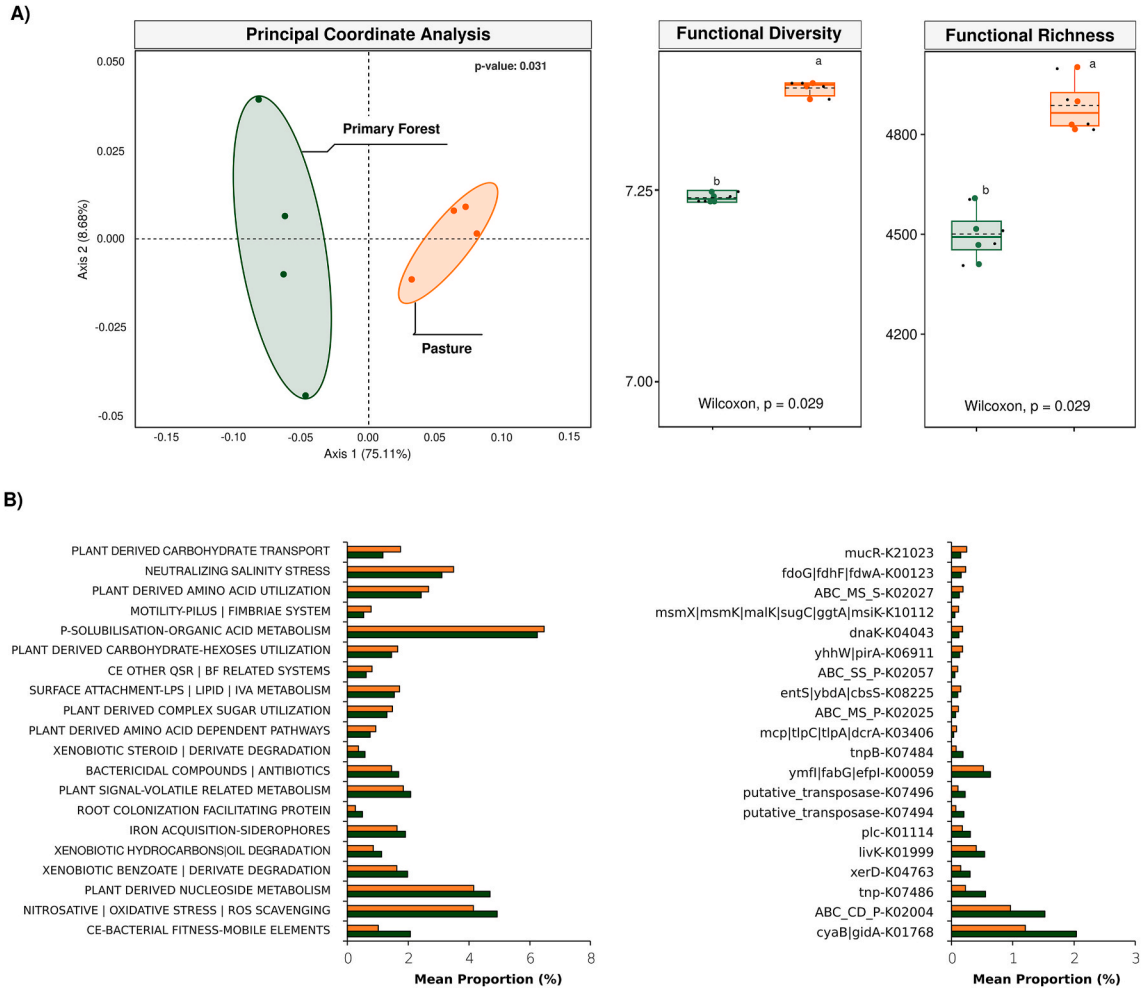


Fig. 3. Overview of Plant-Growth Promotion Gene (PGPG) Abundance in Amazon Environments A) Heatmap Display: Exhibits a general view of significant PGPGs in different samples from two environments. B) Bar Chart Highlighting the variance in gene abundance between the two environments. C) Diversity and Richness Analysis: Showcases a Principal Coordinate Analysis for environmental comparison, along with a boxplot depicting Functional Richness and Diversity indices.

Amazon soils have to adapt to degrade complex polymers such as lignin and cellulose (Lammel et al., 2015). The prevalence of transposase and integrase/recombinase genes reflects the dynamic nature of microbial genomes in response to the complex forest environment, facilitating horizontal gene transfer and genome plasticity, which are crucial for microbial adaptation and evolution in heterogeneous soil environments.

In pasture, genes related to carbohydrate and amino acid transport, abiotic stress neutralization, motility, chemotaxis, and phosphorus mineralization were more prevalent compared to native forests (Fig. 3A; Fig. 3B). The microbial community in pasture soils demonstrated adaptation to explore the niches created by the increase in topsoil root volume by pasture grasses, their root exudates, animal residues, and the inherent soil matrix. Forest-to-pasture conversions increase pH, available phosphorus, and exchangeable calcium and magnesium levels (Mendes et al., 2015a, b; Pedrinho et al., 2019, 2023). These changes are attributed to pasture management practices, such as liming and fertilization, to enhance grass growth, which significantly alter soil chemistry and increase soil nutrient availability compared to forest soils. Genes associated with the exopolysaccharide biofilm development (*muc*R) and motility regulation and chemotaxis (*mcp*) were abundant in pasture soils compared to forest soils. These mechanisms enable microbes to colonize and adapt to the rhizosphere efficiently (Little et al., 2019; Whiteley and Lee, 2015).

Moreover, genes associated with formate hydrogenase major subunit (*fdwa*), multiple sugar transport system (ABC_MS), and simple sugar transport system (ABC_SS) showed an increase in abundance, likely due to the rise in carbon lability through root exudation as most grass species of pastures have been reported to exudate large quantities of simple organic molecules (Kroeger et al., 2018). Another gene abundant in pasture was the molecular chaperone *DnaK,* which ensures protein stability under stress conditions and is vital for microbial resilience in fluctuating stressed environments. In such environments, severe drought stress can trigger microorganisms to produce enzymes and compounds that contribute to stress resistance, contributing to microbial community resilience and rhizosphere interactions (Tartaglia et al., 2023).

3.4. Future perspectives

Looking ahead, the potential of *PGPg_finder* to revolutionize our understanding and utilization of microbial communities in agriculture is immense. Mining organisms in unexplored environments with the *PGPG_finder* tool can yield significant gains in isolating microorganisms capable of performing various agricultural, environmental, and ecological functions of interest. This approach harnesses the untapped potential of diverse habitats to discover organisms with novel capabilities. Future developments could focus on expanding the database to include a more comprehensive array of plant-growth-promoting genes, enhancing the pipeline's accuracy and efficiency. Integrating machine learning algorithms could predict the impact of specific microbial communities on plant health and yield, tailoring microbial inoculants to crops and environmental conditions. Ultimately, this could lead to a new era of precision agriculture, where microbial management strategies are integral to water management, nutrient use, and pest control, thereby ensuring sustainable food production in harmony with the environment.

3.5. Conclusion

In conclusion, the development and application of *PGPg_finder* represents a significant advancement in exploring microbial contributions to plant growth and soil health, particularly in the context of varying land uses such as pastures and native forests. By enabling a detailed and efficient analysis of Plant-Growth Promotion Genes (PGPG) within both genomic and metagenomic datasets, this pipeline not only enhances our understanding of microbial diversity and function but also underscores the vital role of microorganisms in nutrient cycling, stress resistance, and ecosystem sustainability. The insights from contrasting PGPG prevalence in different environments highlight the complex interactions between microbial communities and their habitats, offering valuable perspectives for sustainable agricultural practices and conservation efforts. As we continue to face global challenges like climate change and food security, tools like *PGPg_finder* are crucial for harnessing the potential of beneficial microorganisms to support resilient and productive ecosystems.

CRediT authorship contribution statement

Thierry Alexandre Pellegrinetti: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Gabriel Gustavo Tavares Nunes Monteiro:** Writing – review & editing, Data curation, Conceptualization. **Leandro Nascimento Lemos:** Writing – review & editing, Validation, Methodology. **Renato Augusto Corrêa dos Santos:** Writing – review & editing, Validation, Methodology. **Artur Gomes Barros:** Writing – review & editing, Validation, Methodology. **Lucas William Mendes:** Writing – review & editing, Writing – original draft, Supervision, Resources, Investigation, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The genomes dataset is hosted on the NCBI database under the bioproject PRJNA988090, and metagenomes are available in MG-RAST under the code mgp83361.

Acknowledgments

This study was supported by a grant from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP 2019/16043-7, 2020/12890-4, 2023/07080-1, 2021/11057-0, 2023/11133-3), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq 408191/2018-0, 402050/2023-1), and PRPI-USP (22.1.08498.01.0). LWM also thank CNPq for the Productivity Research Grant (CNPq 307670/2021-0). We thank Prof. Tsai Siu Mui for the lab support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rhisph.2024.100905.

References

Bais, H.P., Weir, T.L., Perry, L.G., Gilroy, S., Vivanco, J.M., 2006. The role of root exudates in rhizosphere interactions with plants and other organisms. Annu. Rev. Plant Biol. 57, 233–266.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

Brettin, T., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Olsen, G.J., Olson, R., Overbeek, R., Parrello, B., Pusch, G.D., 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci. Rep. 5, 8365.

Buchfink, B., Reuter, K., Drost, H.-G., 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat. Methods 18, 366–368. https://doi.org/10.1038/s41592-021-01101-x.

Bushnell, B., 2014. BBMap: A Fast, Accurate, Splice-Aware Aligner (No. LBNL-7065E). Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States).

Carrión, V.J., Perez-Jaramillo, J., Cordovez, V., Tracanna, V., De Hollander, M., Ruiz-Buck, D., Raaijmakers, J.M., 2019. Pathogen-induced activation of disease-suppressive functions in the endophytic root microbiome. Science 366 (6465), 606–612.

Correa, L.O., Bezerra, A.F.M., Honorato, L.R.S., Cortêz, A.C.A., Souza, J.V.B., Souza, E.S., 2021. Amazonian soil fungi are efficient degraders of glyphosate herbicide; novel isolates of Penicillium, Aspergillus, and Trichoderma. Braz. J. Biol. 83, e242830.
da Cunha, I.D.C.M., da Silva, A.V.R., Boleta, E.H.M., Pellegrinetti, T.A., Zagatto, L.F.G., Zagatto, S.D.S.S., et al., 2024. The interplay between the inoculation of plant growth-promoting rhizobacteria and the rhizosphere microbiome and their impact on plant phenotype. Microbiol. Res. 283, 127706.
Delaux, P.M., Schornack, S., 2021. Plant evolution driven by interactions with symbiotic and pathogenic microbes. Science 371 (6531), eaba6605.
De Vries, J., Archibald, J.M., 2018. Plant evolution: landmarks on the path to terrestrial life. New Phytol. 217, 1428–1434.
Er, X.Y., Seow, T.W., Lim, C.K., Ibrahim, Z., Sarip, S.M., 2018. Biological treatment of closed landfill leachate treatment by using Brevibacillus panacihumi strain ZB1. In: IOP Conference Series: Earth and Environmental Science. IOP Publishing, 012012.
Fadiji, A.E., Santoyo, G., Yadav, A.N., Babalola, O.O., 2022. Efforts towards overcoming drought stress in crops: revisiting the mechanisms employed by plant growth-promoting bacteria. Front. Microbiol. 13.
Gherardi, L., Sala, O., 2020. Global patterns and climatic controls of belowground net carbon fixation. Proc. Natl. Acad. Sci. USA 117 (33), 20038–20043. https://doi.org/10.1073/pnas.2006715117.
Govarthanan, M., Mythili, R., Selvankumar, T., Kamala-Kannan, S., Rajasekar, A., Chang, Y.-C., 2016. Bioremediation of heavy metals using an endophytic bacterium Paenibacillus sp. RM isolated from the roots of Tridax procumbens. 3 Biotech 6, 1–7.
Heckman, D.S., Geiser, D.M., Eidell, B.R., Stauffer, R.L., Kardos, N.L., Hedges, S.B., 2001. Molecular evidence for the early colonization of land by fungi and plants. Science 293, 1129–1133.
Hu, X., Gu, H., Liu, J., Wei, D., Zhu, P., Zhou, B., Chen, X., Jin, J., Liu, X., Wang, G., 2022. Metagenomics reveals divergent functional profiles of soil carbon and nitrogen cycling under long-term addition of chemical and organic fertilizers in the black soil region. Geoderma 418, 115846.
Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinf. 11, 119. https://doi.org/10.1186/1471-2105-11-119.
Jones, M.R., Good, J.M., 2016. Targeted capture in evolutionary and ecological genomics. Mol. Ecol. 25, 185–202.
Khan, M.A., Asaf, S., Khan, A.L., Jan, R., Kang, S.-M., Kim, K.-M., Lee, I.-J., 2020. Thermotolerance effect of plant growth-promoting Bacillus cereus SA1 on soybean during heat stress. BMC Microbiol. 20, 1–14.
Kroeger, M.E., Delmont, T.O., Eren, A.M., Meyer, K.M., Guo, J., Khan, K., Rodrigues, J.L., Bohannan, B.J., Tringe, S.G., Borges, C.D., 2018. New biological insights into how deforestation in Amazonia affects soil microbial communities using metagenomics and metagenome-assembled genomes. Front. Microbiol. 9, 1635.
Kwak, M.-J., Kong, H.G., Choi, K., Kwon, S.-K., Song, J.Y., Lee, J., Lee, P.A., Choi, S.Y., Seo, M., Lee, H.J., 2018. Rhizosphere microbiome structure alters to enable wilt resistance in tomato. Nat. Biotechnol. 36, 1100–1109.
Lammel, D.R., Nüsslein, K., Tsai, S.M., Cerri, C.C., 2015. Land use, soil and litter chemistry drive bacterial community structures in samples of the rainforest and Cerrado (Brazilian Savannah) biomes in Southern Amazonia. Eur. J. Soil Biol. 66, 32–39.
Langille, M.G., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A., Clemente, J.C., Burkepile, D.E., Vega Thurber, R.L., Knight, R., 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat. Biotechnol. 31, 814–821.
Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923.
Leontidou, K., Genitsaris, S., Papadopoulou, A., Kamou, N., Bosmali, I., Matsi, T., Madesis, P., Vokou, D., Karamanoli, K., Mellidou, I., 2020. Plant growth promoting rhizobacteria isolated from halophytes and drought-tolerant plants: genomic characterisation and exploration of phyto-beneficial traits. Sci. Rep. 10, 14857.
Li, D., Luo, R., Liu, C.-M., Leung, C.-M., Ting, H.-F., Sadakane, K., Yamashita, H., Lam, T.-W., 2016. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods, Pan-omics analysis of biological data 102, 3–11. https://doi.org/10.1016/j.ymeth.2016.02.020.
Little, R.H., Woodcock, S.D., Campilongo, R., Fung, R.K.Y., Heal, R., Humphries, L., Pacheco-Moreno, A., Paulusch, S., Stigliano, E., Vikeli, E., Ward, D., Malone, J.G., 2019. Differential regulation of genes for cyclic-di-GMP metabolism orchestrates adaptive changes during rhizosphere colonization by Pseudomonas fluorescens. Front. Microbiol. 10.
Lugtenberg, B., Kamilova, F., 2009. Plant-growth-promoting rhizobacteria. Annu. Rev. Microbiol. 63, 541–556.
Majeed, A., Abbasi, M.K., Hameed, S., Imran, A., Rahim, N., 2015. Isolation and characterization of plant growth-promoting rhizobacteria from wheat rhizosphere and their effect on plant growth promotion. Front. Microbiol. 6, 198.
Mendes, Lucas William, de Lima Brossi, M.J., Kuramae, E.E., Tsai, S.M., 2015a. Land-use system shapes soil bacterial communities in Southeastern Amazon region. Appl. Soil Ecol. 95, 151–160.
Mendes, L.W., Mendes, R., Raaijmakers, J.M., Tsai, S.M., 2018. Breeding for soil-borne pathogen resistance impacts active rhizosphere microbiome of common bean. ISME J. 12, 3038–3042.
Mendes, Lucas W., Tsai, S.M., Navarrete, A.A., de Hollander, M., van Veen, J.A., Kuramae, E.E., 2015b. Soil-Borne microbiome: linking diversity to function. Microb. Ecol. 70, 255–265. https://doi.org/10.1007/s00248-014-0559-2.
Mendes, R., Garbeva, P., Raaijmakers, J.M., 2013. The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. FEMS Microbiol. Rev. 37, 634–663.
Mendes, L.W., De Chaves, M.G., de Fonseca, M.C., Mendes, R., Raai- jmakers, J.M., Tsai, S.M., 2019. Resistance breeding of common bean shapes the physiology of the rhizosphere microbiome. Front. Microbiol. 10, 2252.
Nascimento, N., West, T.A., Börner, J., Ometto, J., 2019. What drives intensification of land use at agricultural frontiers in the Brazilian Amazon? Evidence from a decision game. Forests 10, 464.
Navarrete, A.A., Venturini, A.M., Meyer, K.M., Klein, A.M., Tiedje, J.M., Bohannan, B.J. M., Nüsslein, K., Tsai, S.M., Rodrigues, J.L.M., 2015. Differential response of acidobacteria subgroups to forest-to-pasture conversion and their biogeographic patterns in the western Brazilian Amazon. Front. Microbiol. 6.
Parks, D.H., Tyson, G.W., Hugenholtz, P., Beiko, R.G., 2014. STAMP: statistical analysis of taxonomic and functional profiles. Bioinformatics 30, 3123–3124. https://doi.org/10.1093/bioinformatics/btu494.
Patz, S., Gautam, A., Becker, M., Ruppel, S., Rodríguez-Palenzuela, P., Huson, D.H., 2021. PLaBAse: a comprehensive web resource for analyzing the plant growth-promoting potential of plant-associated bacteria. bioRxiv 2021 (12), 13, 472471.
Patz, S., Rauh, M., Gautam, A., Huson, D.H., 2024. mgPGPT: Metagenomic Analysis of Plant Growth-Promoting Traits. https://doi.org/10.1101/2024.02.17.580828.
Pedrinho, A., Mendes, L.W., do Rêgo Barros, F.M., Merloti, L.F., e Martins, M.M., Cotta, S.R., Andreote, F.D., Tsai, S.M., 2023. Impacts of deforestation and forest regeneration on soil bacterial communities associated with phosphorus transformation processes in the Brazilian Amazon region. Ecol. Indicat. 146, 109779.
Pedrinho, A., Mendes, L.W., Merloti, L.F., Andreote, F.D., Tsai, S.M., 2020. The natural recovery of soil microbial community and nitrogen functions after pasture abandonment in the Amazon region. FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Ecol. 96, fiaa149 https://doi.org/10.1093/femsec/fiaa149.
Pedrinho, A., Mendes, L.W., Merloti, L.F., Da Fonseca, M.D.C., Cannavan, F. de S., Tsai, S. M., 2019. Forest-to-pasture conversion and recovery based on assessment of microbial communities in Eastern Amazon rainforest. FEMS Microbiol. Ecol. 95, fiy236.
Pellegrinetti, T.A., Cunha, I.C.M., Chaves, M.G., Freitas, A.S., Silva, A.V.R., Tsai, S.M., Mendes, L.W., 2023. Draft genome sequences of representative Paenibacillus polymyxa, Bacillus cereus, Fictibacillus sp., and Brevibacillus agri strains isolated from Amazonian dark earth. Microbiology Resource Announcements 12, e00574. https://doi.org/10.1128/MRA.00574-23, 23.
Pellegrinetti, T.A., de Cássia Mesquita da Cunha, I., Chaves, M.G. de, Freitas, A.S. de, Passos, G.S., Silva, A.V.R. da, Cotta, S.R., Tsai, S.M., Mendes, L.W., 2024. Genomic insights of Fictibacillus terranigra sp. nov., a versatile metabolic bacterium from Amazonian Dark Earths. Braz. J. Microbiol. 1–12.
Pieterse, C.M., Zamioudis, C., Berendsen, R.L., Weller, D.M., Van Wees, S.C., Bakker, P. A., 2014. Induced systemic resistance by beneficial microbes. Annu. Rev. Phytopathol. 52, 347–375.
Qian, L., Yu, X., Zhou, J., Gu, H., Ding, J., Peng, Y., He, Q., Tian, Y., Liu, J., Wang, S., 2022. MCycDB: a curated database for comprehensively profiling methane cycling processes of environmental microbiomes. Molecular Ecology Resources 22, 1803–1823.
Saeed, Q., Xiukang, W., Haider, F.U., Kučerik, J., Mumtaz, M.Z., Holatko, J., Naseem, M., Kintl, A., Ejaz, M., Naveed, M., 2021. Rhizosphere bacteria in plant growth promotion, biocontrol, and bioremediation of contaminated sites: a comprehensive review of effects and mechanisms. Int. J. Mol. Sci. 22, 10529.
Santos, M., Nogueira, M., Hungría, M., 2019. Microbial inoculants: reviewing the past, discussing the present and previewing an outstanding future for the use of beneficial bacteria in agriculture. Amb. Express 9 (1). https://doi.org/10.1186/s13568-019-0932-0.
Santos-Júnior, C.D., Sarmento, H., de Miranda, F.P., Henrique-Silva, F., Logares, R., 2020. Uncovering the genomic potential of the Amazon River microbiome to degrade rainforest organic matter. Microbiome 8, 1–18.
Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069.
Shaffer, M., Borton, M.A., McGivern, B.B., Zayed, A.A., La Rosa, S.L., Solden, L.M., Liu, P., Narrowe, A.B., Rodríguez-Ramos, J., Bolduc, B., 2020. DRAM for distilling microbial metabolism to automate the curation of microbiome function. Nucleic Acids Res. 48, 8883–8900.
Souza, H.M. de L., Barreto, L.R., Mota, A.J. da, Oliveira, L.A. de, Barroso, H., dos, S., Zanotto, S.P., 2017. Tolerance to polycyclic aromatic hydrocarbons (PAHs) by filamentous fungi isolated from contaminated sediment in the Amazon region, 39 (4), 481–488.
Sun, J., Lu, F., Luo, Y., Bie, L., Xu, L., Wang, Y., 2023. OrthoVenn3: an integrated platform for exploring and visualizing orthologous data across genomes. Nucleic Acids Res. 51, W397–W403. https://doi.org/10.1093/nar/gkad313.
Sun, R., Wang, F., Hu, C., Liu, B., 2021. Metagenomics reveals taxon-specific responses of the nitrogen-cycling microbial community to long-term nitrogen fertilization. Soil Biol. Biochem. 156, 108214.
Tartaglia, M., Ranauda, M.A., Falzarano, A., Maisto, M., Postiglione, A., Prigioniero, A., Scarano, P., Zuzolo, D., Sciarrillo, R., Guarino, C., 2023. Metatranscriptomics of pastures under drought stress show a rhizospheric meta-organism reshape. Rhizosphere 26, 100687. https://doi.org/10.1016/j.rhisph.2023.100687.
Timmusk, S., Copolovici, D., Copolovici, L., Teder, T., Nevo, E., Behers, L., 2019. Paenibacillus polymyxa biofilm polysaccharides antagonise Fusarium graminearum. Sci. Rep. 9, 662.
Timmusk, S., Grantcharova, N., Wagner, E.G.H., 2005. Paenibacillus polymyxa invades plant roots and forms biofilms. Appl. Environ. Microbiol. 71, 7292–7300.
Tu, Q., Lin, L., Cheng, L., Deng, Y., He, Z., 2019. NCycDB: a curated integrative database for fast and accurate metagenomic profiling of nitrogen cycling genes. Bioinformatics 35, 1040–1048.

Vandenkoornhuyse, P., Quaiser, A., Duhamel, M., Le Van, A., Dufresne, A., 2015. The importance of the microbiome of the plant holobiont. New Phytol. 206, 1196–1206.
Venturini, A.M., Gontijo, J.B., Mandro, J.A., Paula, F.S., Yoshiura, C.A., da França, A.G., Tsai, S.M., 2022. Genome-resolved metagenomics reveals novel archaeal and bacterial genomes from Amazonian forest and pasture soils. Microb. Genom. 8, mgen000853 https://doi.org/10.1099/mgen.0.000853.
Volant, S., Lechat, P., Woringer, P., Motreff, L., Campagne, P., Malabat, C., Kennedy, S., Ghozlane, A., 2020. SHAMAN: a user-friendly website for metataxonomic analysis from raw reads to statistical analysis. BMC Bioinf. 21, 345. https://doi.org/10.1186/s12859-020-03666-4.
Waltz, E., 2017. A new crop of microbe startups raises big bucks, takes on the establishment. Nat. Biotechnol. 35, 1120–1123.
Wang, M., Ge, A.-H., Ma, X., Wang, X., Xie, Q., Wang, L., Song, X., Jiang, M., Yang, W., Murray, J.D., Wang, Y., Liu, H., Cao, X., Wang, E., 2024. Dynamic root microbiome sustains soybean productivity under unbalanced fertilization. Nat. Commun. 15, 1668. https://doi.org/10.1038/s41467-024-45925-5.
Wani, P.A., Rafi, N., Wani, U., Bilikis A, H., Khan, M.S.A., 2023. Simultaneous bioremediation of heavy metals and biodegradation of hydrocarbons by metal resistant Brevibacillus parabrevis OZF5 improves plant growth promotion. Ann. Finance 27, 20–31.
Whiteley, C.G., Lee, D.-J., 2015. Bacterial diguanylate cyclases: structure, function and mechanism in exopolysaccharide biofilm development. Biotechnol. Adv. 33, 124–141. https://doi.org/10.1016/j.biotechadv.2014.11.010.
Wickham, H., 2011. ggplot2. WIREs Computational Statistics 3. https://doi.org/10.1002/wics.147.
Yu, X., Zhou, J., Song, W., Xu, M., He, Q., Peng, Y., Tian, Y., Wang, C., Shu, L., Wang, S., 2021. SCycDB: a curated functional gene database for metagenomic profiling of sulphur cycling pathways. Molecular Ecology Resources 21, 924–940.
Zeng, J., Tu, Q., Yu, X., Qian, L., Wang, C., Shu, L., Liu, F., Liu, S., Huang, Z., He, J., 2022. PCycDB: a comprehensive and accurate database for fast analysis of phosphorus cycling genes. Microbiome 10, 101.
Zhang, J., Kobert, K., Flouri, T., Stamatakis, A., 2013. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30. https://doi.org/10.1093/bioinformatics/btt593.
Zheng, Z., Zheng, J., Peng, D., Sun, M., 2017. Complete genome sequence of Fictibacillus arsenicus G25-54, a strain with toxicity to nematodes. J. Biotechnol. 241, 98–100.
Zhou, H., Ren, Z., Zu, X., Yu, X., Zhu, H., Li, X., Zhong, J., Liu, E., 2021. Efficacy of plant growth-promoting bacteria Bacillus cereus YN917 for biocontrol of rice blast. Front. Microbiol. 12, 684888.