UNIVERSIDADE ESTADUAL PAULISTA - UNESP

CÂMPUS DE JABOTICABAL

CHARACTERISTIC GENOMIC TRAITS OF BACTERIAL GENERA ASSOCIATED WITH SUGARCANE

Michelli Inácio Gonçalves Funnicelli

Tecnóloga em Biocombustíveis

2023

UNIVERSIDADE ESTADUAL PAULISTA - UNESP

CÂMPUS DE JABOTICABAL

CHARACTERISTIC GENOMIC TRAITS OF BACTERIAL GENERA ASSOCIATED WITH SUGARCANE

Discente: Michelli Inácio Gonçalves Funnicelli

Orientador: Prof. Dr. Daniel Guariz Pinheiro

Coorientadora: Prof. Dra. Eliana Gertrudes de Macedo Lemos

Tese apresentada à Faculdade de Ciências Agrárias e Veterinárias – UNESP, Câmpus de Jaboticabal, como parte das exigências para a obtenção do título de Doutor em Microbiologia Agropecuária

2023

F982c

Funnicelli, Michelli Inácio Gonçalves
Characteristic genomic traits of bacterial genera associated with sugarcane / Michelli Inácio Gonçalves Funnicelli. -- , 2023
104 p.

Tese (doutorado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal,
Orientadora: Daniel Guariz Pinheiro
Coorientadora: Eliana Gertrudes de Macedo Lemos

1. Metagenômica. 2. Comunidades bacterianas. 3. Bactérias promotoras de crescimento vegetal. 4. Diversidade funcional. 5. Cana-de-açúcar. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Impacto potencial desta pesquisa: Essa pesquisa tem o potencial de gerar impactos substanciais em diversas áreas, desde a agricultura até a biotecnologia. As bactérias associadas a plantas desempenham um papel crucial na promoção da saúde e do crescimento vegetal, auxiliando na aquisição de nutrientes, na resistência a patógenos e no aumento da tolerância a condições adversas. A caracterização genômica dessas bactérias pode revelar informações valiosas sobre funções associadas à adaptação com planta, além de apontar novos insights relacionados a genes e categorias funcionais pouco caracterizadas para serem exploradas em futuros estudos ou aplicações. Os resultados obtidos fornecem informações detalhadas sobre os genes e mecanismos que gêneros específicos de bactérias usam para interagir com a planta. Isso pode levar ao desenvolvimento de estratégias de manejo mais eficazes para melhorar a produtividade de cana-de-açúcar ou para outras culturas de forma sustentável, reduzindo a dependência de produtos químicos agrícolas. O estudo também pode ter implicações importantes na biotecnologia, uma vez que a caracterização genômica de bactérias associadas à cana-de-açúcar pode abrir caminho para a criação de microrganismos geneticamente modificados a fim de melhorar ainda mais a resiliência das plantas.

The potential impact of this research: This research has the potential to generate substantial impacts across various fields, from agriculture to biotechnology. Bacteria associated with plants play a crucial role in promoting plant health and growth by aiding in nutrient acquisition, enhancing resistance to pathogens, and increasing tolerance to adverse conditions. Genomic characterization of these bacteria may unveil valuable insights into functions associated with plant adaptation, offering new perspectives on poorly characterized genes and functional categories for exploration in future studies or applications. The obtained results provide detailed information about the genes and mechanisms that specific genera of bacteria employ to interact with plants. This can lead to the development of more effective management strategies to enhance sugarcane productivity or other crops sustainably, thereby reducing dependence on agricultural chemicals. The study may also have significant implications in biotechnology, as the genomic characterization of bacteria associated with sugarcane could pave the way for creating genetically modified microorganisms to further enhance plant resilience.

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Jaboticabal

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: CHARACTERISTIC GENOMIC TRAITS OF BACTERIAL GENERA ASSOCIATED WITH SUGARCANE

AUTORA: MICHELLI INACIO GONÇALVES FUNNICELLI
ORIENTADOR: DANIEL GUARIZ PINHEIRO
COORIENTADORA: ELIANA GERTRUDES DE MACEDO LEMOS

Aprovada como parte das exigências para obtenção do Título de Doutora em Microbiologia Agropecuária, pela Comissão Examinadora:

Prof. Dr. DANIEL GUARIZ PINHEIRO (Participaçao Virtual)
Departamento de Biotecnologia Agropecuaria e Ambiental / FCAV UNESP Jaboticabal

Prof. Dr. ALESSANDRO DE MELLO VARANI (Participaçao Virtual)
Departamento de Biotecnologia Agropecuária e Ambiental / FCAV UNESP Jaboticabal

Profa. Dra. MARIA CAROLINA QUECINE VERDI (Participaçao Virtual)
Departamento de Genética / Escola Superior de Agricultura Luiz de Queiroz - USP

Prof. Dr. EVERLON CID RIGOBELO (Participaçao Virtual)
Departamento de Ciencias da Producao Agricola / FCAV UNESP Jaboticabal

Prof. Dr. GISELLE GOMES MONTEIRO FRACETTO (Participaçao Virtual)
Depto de Agronômia / Universidade Federal Rural de Pernambuco

Jaboticabal, 02 de março de 2023

gov.br Documento assinado digitalmente
DANIEL GUARIZ PINHEIRO
Data: 04/01/2024 19:09:06-0300
Verifique em https://validar.iti.gov.br

gov.br Documento assinado digitalmente
ALESSANDRO DE MELLO VARANI
Data: 11/01/2024 19:11:28-0300
Verifique em https://validar.iti.gov.br

gov.br Documento assinado digitalmente
MARIA CAROLINA QUECINE VERDI
Data: 18/01/2024 15:55:50-0300
Verifique em https://validar.iti.gov.br

gov.br Documento assinado digitalmente
EVERLON CID RIGOBELO
Data: 09/01/2024 17:23:23-0300
Verifique em https://validar.iti.gov.br

gov.br Documento assinado digitalmente
GISELLE GOMES MONTEIRO FRACETTO
Data: 15/01/2024 10:29:59-0300
Verifique em https://validar.iti.gov.br

DADOS CURRICULARES DO AUTOR

Michelli Inácio Gonçalves Funnicelli – Nascida em 16 de outubro de 1991, em São José do Rio Preto, São Paulo, Brasil. Graduou-se em Tecnologia em Biocombustíveis, em dezembro de 2015 pela Faculdade de Tecnologia, Fatec de Jaboticabal. Em março de 2016 ingressou no curso de Mestrado em Microbiologia Agropecuária, na Faculdade de Ciências Agrárias e Veterinárias - Universidade Estadual Paulista “Júlio de Mesquita Filho” – FCAV/ UNESP, na cidade de Jaboticabal, sob a orientação da Profa. Dra. Eliana G. de Macedo Lemos e obteve o título de mestre em 2018. No mesmo ano foi aprovada no curso de Doutorado pelo mesmo programa sob a orientação do Prof. Dr Daniel Guariz Pinheiro .

AGRADECIMENTOS

Ao Prof. Dr. Daniel Guariz Pinheiro, por todo conhecimento compartilhado, orientação e paciência em nossas discussões, e sobretudo, por acreditar em minhas competências e confiar em meu trabalho.

Ao Profa. Dra. Eliana Gertrudes de Macedo Lemos pela coorientação, por sempre me apoiar e por todo o suporte de infraestrutura computacional.

Aos meus amigos do Laboratório de bioinformática, Lucas Amoroso, Sabrina Dibelli e Luis Teheran pela convivência, momentos de descontração e suporte durante esses anos.

Às minhas amigas Natália Sarmanho, Pâmela Maldaner, Bárbara Bonfá, Tatiane Fernanda pela amizade dentro e fora da faculdade, apoio e as diversas risadas compartilhadas.

A todos os meus familiares, especialmente à minha mãe, Irene Inácio Gonçalves, pelo amor incondicional e por acreditar e me apoiar sempre.

Ao Gustavo Alan Siniscalchi, pelo companheirismo em momentos em que eu mais precisei.

À Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista "Júlio de Mesquita Filho" (FCAV/Unesp) pelo suporte acadêmico e infraestrutura utilizada durante o período do mestrado.

À CAPES pela bolsa concedida. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), pelo financiamento da pesquisa (no. 2017/09008-5).

E todos que acompanharam e me apoiaram durante essa caminhada e que de alguma forma contribuíram para a realização deste trabalho.

Meus sinceros agradecimentos a todos!

CONTENTS

RESUMO ..... **iii**
ABSTRACT ..... **v**
CHAPTER 1 - General considerations **1**
1. INTRODUCTION **1**
2. LITERATURE REVIEW **2**
2.1. Soil and plant-associated microorganisms 2
2.2. Sugarcane: Importance 4
2.3. Microbiome associated with sugarcane 4
2.4. Plant Growth-Promoting Bacteria 7
2.5. Mechanisms of action of plant growth-promoting bacteria 10
2.5.1. Direct mechanisms 10
2.5.2. Indirect mechanisms 15
2.6. Metagenomics applied in studies of plant-associated microbiome 17
3. REFERENCES **19**
APPENDIX **29**
CHAPTER 2 – Unveiling genomic features linked to traits of plant-growth-promoting bacterial communities from sugarcane **43**
1. INTRODUCTION **44**
2. MATERIAL AND METHODS **46**
2.1. Selected bacterial communities 46
2.2. DNA extraction 46
2.3. Construction of a metagenomic library and sequencing on the PacBio Sequel II platform 47
2.4. Quality filtering of long reads 47
2.5. Metagenome assembly and genome retrieval 48
2.6. Taxonomic assignment and phylogenomic analysis 49
2.7. Gene prediction and functional annotation 49
2.8. Genes related to plant growth-promoting traits 50
2.9. Data availability 51
3. RESULTS AND DISCUSSION **51**
3.1. Metagenome-assembled genomes and taxonomy assignment 51
3.2. Functional Profile 56
3.3. Functional PGPT annotation 62
4. CONCLUSIONS **67**
5. REFERENCES **68**
APPENDIX **78**
CHAPTER 3 – Classification of genomic features of plant-associated bacteria using machine learning **80**
1. INTRODUCTION **81**
2. MATERIAL AND METHODS **82**
2.1. Genomic data compilation and metadata 82
2.2. Construction of the phylogenetic tree 83
2.3. Inference of orthologous gene groups 83
2.4. Functional annotation of the representative orthogroups 84

2.5. Enrichment analysis of orthogroups according to lifestyle..... 84
2.6. Classification of plant-associated (PA) genomes and features (genes) selection..... 85
2.7. Enrichment of functional categories of Clusters of Orthologous Groups (COG)..... 85
3. RESULTS..... 86
3.1. Bacterial genome groups and phylogenomic analysis..... 86
3.2. Orthologous gene enrichment..... 87
3.3. Performance evaluation of the machine learning approaches..... 89
3.4. Functional categories regarding plant-associated bacteria gene content..... 91
4. DISCUSSION..... 93
5. CONCLUSION..... 95
6. REFERENCES..... 97
APPENDIX..... 102

TRAÇOS GENÔMICOS CARACTERÍSTICOS DE GÊNEROS BACTERIANOS ASSOCIADOS À CANA-DE-AÇÚCAR

RESUMO - Os microrganismos associados a plantas são cruciais para a saúde e produtividade vegetal. Estudos de *metabarcoding* e dados genômicos relacionados ao microbioma vegetal de diversas espécies têm revelado informações relevantes sobre a composição e estrutura da microbiota em resposta a diferentes condições e estresses ambientais. Isso tem ampliado nossa compreensão sobre as interações microrganismo-planta, destacando o potencial das tecnologias derivadas desses estudos para aplicações na agricultura. No Brasil, a cana-de-açúcar, de alta relevância econômica, serve como um campo promissor para explorar microrganismos benéficos. Deste modo, o presente trabalho teve como objetivo investigar genomas bacterianos isolados de comunidades previamente reconhecidas como promotoras de crescimento vegetal, provenientes da cana-de-açúcar. Para isso, empregamos a abordagem de montagens de genomas a partir de metagenomas (MAGs) na investigação de traços benéficos dos membros presentes dessas comunidades bacterianas, a fim de identificar os principais mecanismos relacionados às características de promoção de crescimento vegetal, tais como disponibilização de nutrientes, síntese de fitohormônios, entre outros, e de destacar candidatos para avaliação em campo devido às suas características complementares e benéficas para as plantas. Após determinar os gêneros bacterianos presentes nas comunidades, foi realizada uma análise de genômica comparativa de espécies pertencentes a esses gêneros, provenientes ou não de plantas, para determinar características gênicas associadas à colonização de plantas. A análise genômica dos 17 genomas identificados permitiu determinar as principais características associadas aos mecanismos subjacentes às interações bactéria-planta. Cinco dos genomas estudados que foram identificados como *Enterobacter asburiae* (MAG06 e MAG10), *Citrobacter werkmanii* (MAG17), *Rhizobium punense* (MAG03), e *Achromobacter animicus* (MAG14) destacaram-se pela alta densidade de genes associados a características benéficas. Em relação à análise genômica voltada para compreender os mecanismos envolvidos na associação com as plantas, dos 13 gêneros estudados, 9 apresentam enriquecimento de genes, permitindo seguir com modelos de classificação para a predição de genomas associados às plantas. Nesse contexto, a abordagem de *random forest* (RF) superou os modelos de regressão logística (RL). A análise de enriquecimento de genes ortólogos além das assinaturas de categorias COG compartilhadas entre genomas revelaram um conjunto significativo de genes relacionados a Transporte e metabolismo de aminoácidos (E), Transporte e metabolismo de lipídios (I), Biossíntese, transporte e catabolismo de metabólitos secundários (Q), Transcrição (K) e Mecanismos de transdução de sinal (T). Os resultados sugerem que as bactérias investigadas têm um potencial significativo para promover o crescimento de plantas, posicionando-os como candidatos promissores para práticas agrícolas sustentáveis. Estudos como este têm sido cruciais para o avanço da agricultura sustentável, uma vez que os microrganismos

desempenham papéis fundamentais na promoção da saúde e do crescimento das plantas. Uma compreensão mais profunda das interações microrganismos-plantas permitirá o desenvolvimento de estratégias mais eficazes para promover a saúde do solo, reduzindo a dependência de fertilizantes químicos e aumentando a resistência das plantas a doenças e fatores de estresse ambiental.

Palavras-chave: Bactérias Promotoras de Crescimento Vegetal (BPCV); Genômica comparativa; Metagenômica; Sequenciamento de Nova Geração (SNG)

CHARACTERISTIC GENOMIC TRAITS OF BACTERIAL GENERA ASSOCIATED WITH SUGARCANE

ABSTRACT - Plant-associated microorganisms are crucial to plant health and productivity. Metabarcoding studies and genomic data related to the plant microbiome of several species have revealed valuable information about the composition and structure of the microbiota in response to different conditions and environmental stresses. This has been expanding our understanding of plant-microorganism interactions, highlighting the potential of technologies derived from these studies for applications in agriculture. In Brazil, sugarcane cultivation is of great economic importance and has been an interesting source for prospecting plant growth-promoting microorganisms. Therefore, the present work aimed to investigate bacterial genomes isolated from communities previously recognized as promoters of plant growth, derived from sugarcane. To this end, we employed the genome assembly approach from metagenomes (MAGs) to investigate beneficial characteristics of members present in these bacterial communities, in order to identify the main mechanisms related to plant growth-promoting characteristics, such as nutrient availability, synthesis of phytohormones, among others, and highlight candidates for field evaluation due to their complementary and beneficial characteristics for plants. After determining the bacterial genera in the communities, a comparative genomic analysis of species belonging to these genera, whether or not originating from plants, was carried out to determine gene characteristics associated with plant colonization. The genomic analysis of the 17 identified genomes allowed us to determine the main characteristics associated with the mechanisms underlying bacteria-plant interactions. Five of the studied genomes, identified as *Enterobacter asburiae* (MAG06 and MAG10), *Citrobacter werkmanii* (MAG17), *Rhizobium punense* (MAG03), and *Achromobacter animicus* (MAG14), exhibited a high density of genes associated with *Plant Growth-Promoting Traits* (PGPTs). Regarding genomic analysis aimed at understanding other mechanisms involved in association with plants, out of the 13 genera studied, only 9 exhibited gene enrichment, enabling the use of classification models to predict genomes associated with plants. The random forest approach outperformed logistic regression models by employing machine learning methods to predict genomes associated with plants. The enrichment analysis of orthologous genes along with the signature of COG categories shared among genomes revealed a significant set of genes related to Amino acid transport and metabolism (E), Lipid transport and metabolism (I), Biosynthesis, transport and catabolism of secondary metabolites (Q), Transcription (K) and Signal transduction mechanisms (T). The results found here demonstrate that the identified bacteria have the potential to promote plant growth, positioning them as promising candidates for sustainable agricultural practices. The study of microorganisms has been crucial for the advancement of sustainable agriculture, as they play fundamental roles in promoting the health and growth of plants. A deeper understanding of the interactions of microorganisms-plants will allow the development of more effective strategies to promote soil health, reducing dependence on chemical fertilizers and increasing plant resistance to diseases and environmental stressors.

Keywords: Plant Growth-Promoting Bacteria (PGPB); Comparative genomics; Metagenomic; Next-Generation Sequencing (NGS)

CHAPTER 1 - General considerations

1. INTRODUCTION

Microorganisms occur in a variety of environments, including plant tissues that harbor a diverse microbial life, such as bacteria and fungi. Interactions between microorganisms and plants can be simplified as beneficial, neutral, or detrimental (Hassani et al. 2018), and a relationship can be disrupted when environmental conditions change (Cheng et al. 2019).

The abundance and diversity of associated taxa of the rhizosphere and other compartments are shaped by the plant and the soil is the main reservoir of plant-colonizing microorganisms (Compant et al. 2010, Banerjee and van der Heijden 2023). However, other factors can also influence the composition of the associated microbiota, such as environmental variability (soil pH, C:N ratio, soil carbon, water content), cropping practices (monoculture or crop rotation, type of fertilization) (Dastogeer et al. 2020).

Scientists have been isolating and studying microorganisms associated with various plant species for a long time. They are interested in the beneficial effects these microorganisms can promote for the health and resilience of plants. Several studies have reported the plant health benefits mediated by the associated microbiota, such as disease suppression (Mendes et al. 2011), increased nutrient acquisition (Muthukumarasamy et al. 2017), growth promotion (Armanhi et al. 2018), and induction of systemic resistance (Yu et al. 2022), among others.

The sugarcane crop (*Saccharum* spp.) is considered one of the most important activities for the world economy and is a key crop in the Brazilian economy due to the production of sugar and ethanol (Heinrichs et al. 2017). It is recognized as one of the most important energy crops, a recognition intensified by the advent of ethanol production (Tew and Cobill 2008).

The extensive sugarcane cultivation imposes a high demand for fertilizers and pesticides, which has a significant environmental impact. In this sense, the manipulation of the sugarcane microbiome has been explored as a promising

alternative for sustainable cultivation (Armanhi et al. 2018, de Souza et al. 2019, de Carvalho et al. 2021, Teheran-Sierra et al. 2021).

Advances in next-generation sequencing (NGS) technologies have been crucial to understanding microbial diversity in environmental samples (Scholz et al. 2012), and long-read sequencing technology has contributed substantially to the analysis of microbial communities (Cuscó et al. 2021, Haro-Moreno et al. 2021).

Therefore, the present study investigated bacterial genomes from bacterial communities already recognized as plant growth promoters isolated from sugarcane. For this purpose, we used the approach of genome assemblies from metagenomes to investigate beneficial traits of the members present in the bacterial communities to determine the main mechanisms related to the characteristics of plant growth promotion. After determining the bacterial genera in the communities associated with sugarcane, we performed a comparative genomic analysis of species belonging to the identified genera, whether or not they were derived from plants, to determine the genetic characteristics associated with plant adaptation.

2. LITERATURE REVIEW

2.1. Soil and plant-associated microorganisms

Microorganisms are present in a wide range of environments, including those considered more hostile. The soil is a complex environment and one of the largest reservoirs of microbial genetic diversity, and dynamics, which largely remains under-explored (Torsvik and Øvreås 2002). Additionally, it is considered a cornerstone of One Health concept (Banerjee and van der Heijden 2023).

Soil microorganisms play several crucial roles in nutrient cycling, soil carbon sequestration, maintenance of fertility, and soil microbiota can directly or indirectly affect the health of plants and animals in terrestrial ecosystems (Fierer 2017). Additionally, they are a rich source of biologically active metabolites and enzymes (Daniel 2004), which have numerous biotechnological applications in different fields.

The plants and their associated microorganisms coevolve in response to reciprocal adaptation, and this relationship has been recognized to be one of the

main determinants of plant health and productivity (Berg et al. 2017, 2020). While plants have their own mechanisms to mitigate biotic and abiotic stresses in nature, the microorganisms within them can play a key role in fostering numerous benefits. These include disease suppression, activation of the plant's immune system, induction of systemic resistance, enhanced nutrient acquisition, improved tolerance to abiotic stresses, adaptation to environmental fluctuations, and promotion of mycorrhizal associations (Hassani et al. 2018).

The plant system shows a multitude of niches that enable the growth and proliferation of diverse microorganisms (Trivedi et al. 2020). Plants actively recruit microorganisms from surrounding microbial reservoirs (Compant et al. 2019), and the rhizosphere is considered a hotspot of plant–microbe interactions, due to the intensity of microbial activity. It represents one of the most complex ecosystems because root exudates and mucilage attract and select a high diversity of soil microorganisms (Mitter et al. 2016).

Plant microbiome components may form separate communities, such as those found in the plant rhizosphere, endosphere, and phyllosphere. This is due to the environment-dependent selection process and response to microbes in each of these niches (Andreote et al. 2014). Additionally, microbial diversity presents a gradient when comparing aboveground and belowground compartments, meaning that phyllosphere microbial communities have relatively low species diversity with a high rate of change (Lebeis 2015). In the study described by Trivedi et al. (2020), the profile of bacterial groups in the soil exhibited similarity with the rhizosphere, with a slight increment in the Proteobacteria phylum. Similarly, with the fungal community, the profile differs in terms of belowground versus aboveground, colonized by a diversity of fungi mainly belonging to the phyla Ascomycota and Basidiomycota. However, a subset of microorganisms called the core microbiome, which constitutes groups of microorganisms that are particularly widespread within a host population, represent a particularly critical component for host biological functions (Risely 2020), and a few members are keystone taxa that influence and drive the community structure (Banerjee et al. 2018). This has contributed to understanding plant–microbiome structure and provided a starting point for assembling SynComs to manipulate plant–microbiome interactions for increased growth and productivity (Trivedi et al. 2020).

2.2. Sugarcane: Importance

Sugarcane is a large perennial tropical grass belonging to the genus *Saccharum*, a member of the Andropogoneae tribe of the grass family (Poaceae) and has a very close genetic relationship to sorghum and other grass family members (Godshall and Legendre 2003, Hoang et al. 2015).

The cultivation of sugarcane (*Saccharum* spp.) is considered one of the most important activities for the global economy. This crop has a significant economic impact on the Brazilian agribusiness sector. In addition to sugar and ethanol production, sugarcane produces numerous valuable by-products such as bagasse, press mud, molasses, and spent wash (vinasse), demonstrating the versatility of this crop.

During the 2021-2022 crop year, around 585 million tons of sugarcane were produced in the country, in a harvest area of 8.317,3 million hectares. This volume represents a retraction of 10.6% compared to the previous harvest (2020/2021) (CONAB 2022a).

The harvest survey for 2022/23 estimated a production of 572.9 million tons, indicating a 1% decrease from the previous harvest. It also shows a 2.6% decrease in the area devoted to sugar and alcohol production, primarily due to the increased demand for areas dedicated to soybeans and corn cultivation, driven by attractive grain prices (CONAB 2022b).

2.3. Microbiome associated with sugarcane

The plant microbiome has gained prominence due to unexplored diversity (microbial dark matter) (Mapelli et al. 2022), highlighting the importance of microorganisms' roles in host associations with implications for the plant's health, development, and productivity (Levy et al. 2018).

In recent years, there has been a significant increase in studies aimed at understanding plant-associated microbiomes, with a notable swell, especially from 2019 to 2021 (Figure 1).

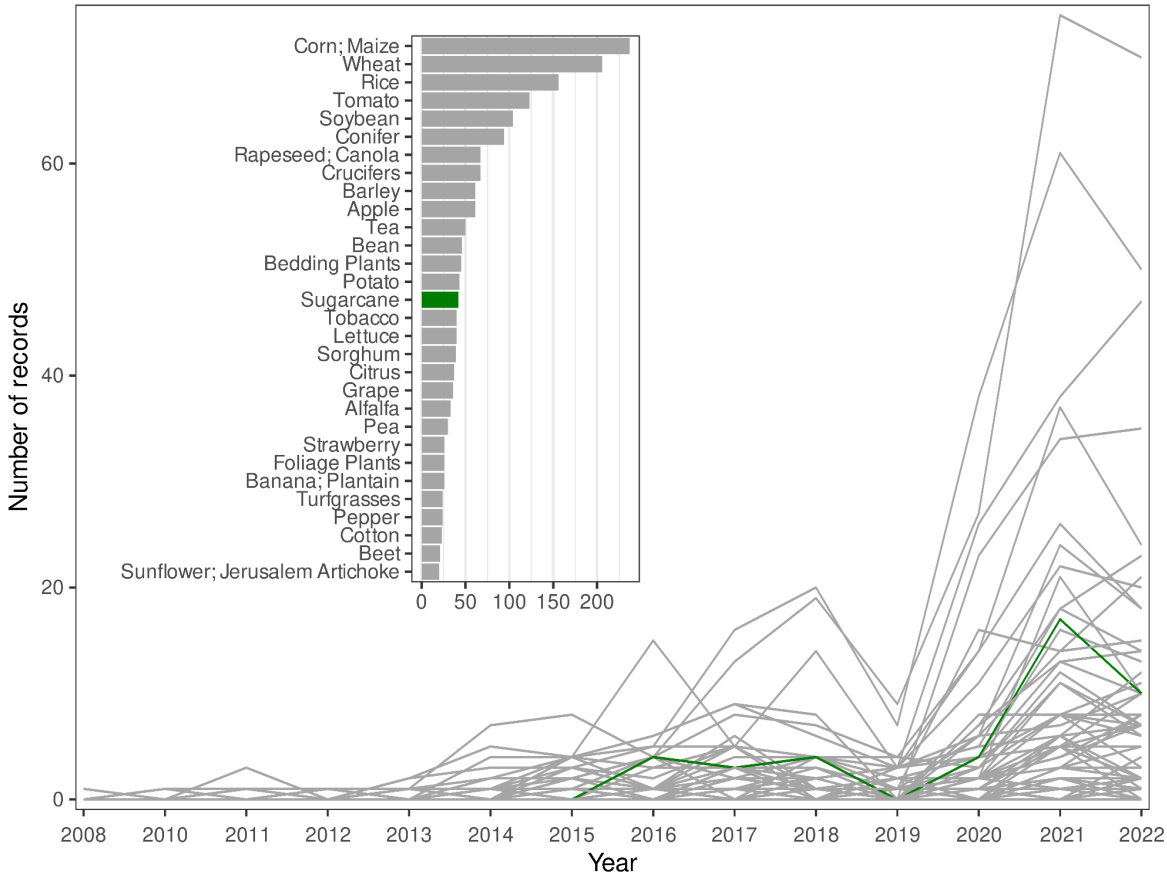

Figure 1. The trend in the number of publications (NP) related to the plant-associated microbiome from 2008 to 2021 in the Web of Science Core Collection[1] (Methodology of automatic searches based on terms are available in Appendix A). The bar plot displays the total number of articles per plant (only plants with over 20 records) and the line plot displays the cumulative number per year. Studies associated with sugarcane are highlighted in green.

At the top of the list of plants with the most studied microbiomes (Figure 1) are those related to staple foods, such as cereals like maize (corn), wheat, and rice, which are acknowledged as the most relevant crops worldwide and contribute to approximately 50% of the total dietary calories (Erenstein et al. 2022). This highlights the importance of microbiota for the plant's health, development, and productivity.

[1] https://www.webofscience.com/wos/woscc/basic-search

The sugarcane is widely acknowledged as a valuable model system for microbiome studies, owing to its significant economic importance in Brazil and other countries, particularly for sugar and biofuel production. However, there is a lack of comprehensive studies focused on this crop. Fortunately, some studies have shed light on the structure and composition of the microbial assemblage associated with sugarcane.

The study conducted by Hamonts *et al*. (2018) investigated the microbiome of distinct varieties of sugarcane at various stages to identify the factors influencing the assembly of the community under field conditions and the behavior of the microbiome associated with the occurrence of Yellow Canopy Syndrome (YCS). The study's findings revealed that the plant compartment was the primary determinant of sugarcane-associated microbial assemblages, followed by growing region, crop age, variety, and the incidence of Yellow Canopy Syndrome (YCS). Furthermore, the occurrence of the syndrome significantly influenced a core microbiome.

In another study on the sugarcane microbiome, Souza *et al.* (2016) provide a detailed inventory of the structure and composition of bacterial and fungal communities. The study's findings reveal that the core microbiome comprises a small subset of species, which represent a large proportion (more than 90% of the total relative abundance) in any organ. However, differences between organs reflect variations in relative taxa abundances. Furthermore, these groups play critical biological roles in plant growth and fermentation. Interestingly, some of the most abundant taxa in central microbiomes have not been isolated or studied before. In contrast to the bacterial component of the microbiome, only a small portion of fungal diversity has been investigated, suggesting that a substantial portion remains unknown, especially regarding the role of fungal colonizers in plant growth, development, and response to biotic and abiotic stress.

The farming system can also influence the composition and structure of the sugarcane microbiome. In De Carvalho *et al.* (2021), the richness and diversity of microorganisms were slightly greater in conventional systems. However, the farming system alone did not represent a weighty factor in driving the fungal and bacterial communities' assembly, but in combination with the plant environmental niche. Among the differentially abundant taxa, several genera with the potential to promote

plant growth and protection were identified, evidencing a potential source for isolation and prospection of beneficial microorganisms for agriculture.

Microorganisms associated with sugarcane have exhibited a large potential for promoting plant growth, as demonstrated in a study conducted by Armanhi *et al.* (2018). In this study, a synthetic bacterial community assembled from the sugarcane microbiome was found to be compatible with corn, dominating over 50% of the microbial abundance in the rhizosphere. This dominance led to a remarkable increase in biomass 3.4 times greater in the inoculated plants group than in the control. Moreover, the genomic analysis of this synthetic community revealed that commonly studied PGP traits were not the sole determinants of robust colonization. Therefore, the observed PGP traits in the experiments could be related to other unknown features, as suggested by de Souza *et al.* (2019).

2.4. Plant Growth-Promoting Bacteria

The global food demand is continuously increasing, and with this increase, environmental impacts caused by agriculture also are expanding (Tilman et al. 2011). To deal with this, sustainable alternatives need to be intensified. Beneficial microorganisms have been highlighted for sustainable agriculture, playing a significant role in increasing the productivity of various crops, helping to maintain soil fertility, and reducing pollutants derived from the intensive use of chemical inputs.

Several plant-associated microorganisms can provide various benefits to hosts, and they have been highlighted as promising for improving plant health and growth. These are commonly referred to as plant growth-promoting microorganisms (PGPM) which are divided into plant growth-promoting [rhizo]bacteria (PGPB or PGPR) and plant growth-promoting fungi (PGPF) (Compant et al. 2010). All of them have mechanisms that promote beneficial effects for plants, helping to maintain the health of the host.

The genera *Azospirillum*, *Bacillus*, *Burkholderia*, *Enterobacter*, *Gluconacetobacter*, *Herbaspirillum*, *Klebsiella*, *Paenibacillus*, *Pantoea*,

Pseudomonas, *Rhizobium*, *Stenotrophomonas* represent genera commonly associate with sugarcane beneficial traits (Table 1) and likely constitute part of the sugarcane microbiome (Yeoh et al. 2016, Armanhi et al. 2018, de Carvalho et al. 2021, Teheran-Sierra et al. 2021).

Table 1. Identified sugarcane-associated bacterial genera and their traits related to plant growth promotion. The methodology of automatic searches based on terms and references of records in the table is described in Appendix A and Appendix B, respectively. IAA = indoleacetic acid production, NIT = nitrogen fixation, PHOS = phosphate solubilization, ACC = ACC deaminase, SID = siderophore production, STR = abiotic/biotic stress tolerance, CK = cytokinin production, HCN = HCN production, AM = ammonia production, GA = gibberellin production, BIOC = biocontrol and PG = plant growth promotion.

Genus	Traits	References
Achromobacter	IAA; NIT; PHOS; SID; BIOC; PG	1–3
Acidomonas	IAA; BIOC	4
Acinetobacter	IAA; NIT; PHOS; BIOC; PG	2,5,6
Agrobacterium	IAA; NIT; PHOS; ACC; SID; BIOC; PG	2,7,8
Arthrobacter	ACC; BIOC; PG	9
Asaia	NIT; PHOS; SID; AM; BIOC; PG	4
Azospirillum	IAA; NIT; PHOS; SID; BIOC; PG	10-21
Azotobacter	NIT	12
Bacillus	IAA; NIT; PHOS; ACC; SID; STR; CK; HCN; AM; BIOC; PG	1,3,4,9,16,17,19,20,22–43
Beijerinckia	NIT; PG	44
Bradyrhizobium	NIT; BIOC; PG	20,45–49
Brevibacillus	IAA; NIT; PHOS; BIOC; PG	1,39,42
Burkholderia	IAA; NIT; PHOS; ACC; SID; AM; BIOC; PG	1,4,7,13,20–22,27,39,42,44,49–69
Chryseobacterium	IAA; PHOS; ACC; PG	70
Clostridium	NIT	71
Cohnella	IAA; PHOS; BIOC; PG	42
Curtobacterium	NIT; BIOC	42,60
Delftia	IAA; PHOS; SID; HCN; BIOC; PG	72
Dyella	IAA; NIT; PHOS; PG	27
Enterobacter	IAA; NIT; PHOS; ACC; SID; HCN; BIOC; PG	2,3,10,26,27,40,44,62,63,73–83
Erwinia	IAA; NIT; PHOS; PG	27
Gluconacetobacter	IAA; NIT; PHOS; SID; BIOC; PG	7,11,13,15,16,18,19,49,53,61,69,77,84,85,85–103
Herbaspirillum	IAA; NIT; PHOS; SID; HCN; GA; BIOC; PG	1,11,13,15,16,21,36,39,49,50,5

		3,56,69,72,74,90,93,101,104–107
Ideonella	NIT; BIOC	49
Klebsiella	IAA; NIT; PHOS; ACC; SID; HCN; BIOC; PG	44,45,50,75,77,82,108–111
Kosakonia	IAA; NIT; PHOS; GA; BIOC; PG	22,56,78,107,112–114
Leifsonia	BIOC	60
Mesorhizobium	NIT; BIOC; PG	20,60
Methylobacter	NIT; PHOS; PG	39
Methylobacterium	IAA; NIT; PHOS; CK; PG	39,40,115–118
Microbacterium	IAA; NIT; ACC; BIOC; PG	9,60,119,120
Microbispora	IAA; NIT; PHOS; ACC; SID; BIOC; PG	40,121
Nguyenibacter	NIT; PHOS; SID; AM; BIOC; PG	4
Nitrospirillum	NIT; SID; BIOC; PG	61,94,122,123
Ochrobactrum	BIOC	60
Paenibacillus	IAA; NIT; PHOS; SID; BIOC; PG	23,24,40,42
Pannonibacter	IAA; NIT; PHOS; ACC; SID; HCN; PG	75
Pantoea	IAA; NIT; PHOS; ACC; SID; HCN; BIOC; PG	2,27,63,72,73,119,124–128
Paraburkholderia	NIT; GA; BIOC; PG	56,61,94,107,129
Pseudomonas	IAA; NIT; PHOS; ACC; SID; HCN; AM; GA; BIOC; PG	2,9,17,19,22,23,27,31,36,38,40,43,44,50,56,75,105,107,108,119,126,130–136
Rahnella	IAA; NIT; PHOS; ACC; SID; HCN; PG	75
Raoultella	NIT; PG	137
Rhizobium	IAA; NIT; PHOS; ACC; SID; HCN; BIOC; PG	8,20,49,62,75,126
Roseateles	IAA; NIT; PHOS; PG	65
Serratia	NIT	45
Shinella	NIT; BIOC; PG	2,80
Sphingobium	IAA; PHOS; PG	62
Sphingomonas	NIT; BIOC; PG	60,138
Staphylococcus	IAA; PHOS; PG	3
Stenotrophomonas	IAA; NIT; PHOS; ACC; SID; HCN; BIOC; PG	2,7,8,8,27,43,63,75,112,139
Streptomyces	IAA; NIT; PHOS; ACC; SID; BIOC; PG	40,140
Tanticharoenia	NIT; PHOS; SID; AM; BIOC; PG	4
Xanthomonas	IAA; NIT; PHOS; ACC; SID; HCN; BIOC; PG	2,60,75

2.5. Mechanisms of action of plant growth-promoting bacteria

There are direct and indirect mechanisms related to plant growth promotion, here we will focus on the most common traits present in bacteria that have these mechanisms.

2.5.1. Direct mechanisms

Biological Nitrogen Fixation

Nitrogen is an essential element of life and is one of the most important nutrients that limit crop yield. Plants cannot directly assimilate atmospheric nitrogen or di-nitrogen (N_2). However, certain archaea or bacteria known as diazotrophs have the ability to convert atmospheric nitrogen into ammonia (NH_3), which is usable by plants (Soumare et al. 2020). This process, known as biological nitrogen fixation (BNF), is catalyzed by nitrogenases, which are complex metalloenzymes with conserved structural and mechanistic features (Dixon and Kahn 2004). The most prevalent nitrogenase complex consists of two proteins: the catalytic molybdenum-iron protein (MoFeP) encoded by *nifDK* genes and its specific reductase, the iron protein (FeP) encoded by *nifH* gene. Additionally, several regulatory proteins involved in nitrogen fixation are encoded by the *nif* genes (Dixon and Kahn 2004, Soumare et al. 2020). Some species also possess homologous alternative nitrogenases, such as vanadium-containing nitrogenase (FeV) and/or iron-only nitrogenase (FeFe).

The diazotrophs are found in a range of habitats, including free-living in soils and within plants as endophytes or endosymbionts (Mus et al. 2016). Several groups participate in nitrogen fixation, and the most well-known association (plant–microbe mutualism) is the symbiosis between rhizobia and legumes, as well as the actinorhizal symbioses between filamentous nitrogen-fixing soil bacteria of the genus *Frankia* and diverse group of dicotyledonous plants (Actinorhizal plants) (Mus et al. 2016).

Synthetic nitrogen fertilizers are commonly used to alleviate the nitrogen limitation in crops, consequently increasing yields. However, intensive fertilization can lead to several environmental impacts (Sainju et al. 2019).

Nitrogen-fixing bacteria, known as diazotrophs, have emerged as an alternative to replace and reduce the use of synthetic nitrogen fertilizers. Efforts and interest are increasing to facilitate higher levels of Biological Nitrogen Fixation (BNF) in non-legume crops. The study conducted by Wen et al. (2021) highlighted the successful development and commercialization of the first microbial product composed of *Klebsiella variicola* 137-1036 (Kv137-1036), modified through synthetic biology to enable BNF for corn (Zea mays) in fertilized fields. This represents a significant step towards the development of sustainable agricultural practices that reduce our reliance on synthetic nitrogen fertilizers.

Phosphate and Potassium Solubilization

Phosphorus is an essential macronutrient required for key metabolic processes in plants, including cell division, energy generation, macromolecule biosynthesis, membrane integrity, signal transduction, and photosynthesis (Rawat et al. 2021). Although it is a common element in the soil and abundant, a huge portion is predominantly in insoluble forms that are not readily absorbed by plants. This necessitates the use of phosphate fertilizers to address phosphorus deficiency.

Phosphorus occurs in soil in two forms organic, which is derived from biological metabolic processes, and inorganic, comprising phosphates bound to minerals like calcium phosphate, aluminum phosphate, and iron phosphate (Rawat et al. 2021). Due to their highly reactive nature, these forms undergo various transformations, rendering P inaccessible to plants (De Zutter et al. 2022).

Phosphate-solubilizing microorganisms (PSMs) possess metabolic capacity to enhance the bioaccessibility of various recalcitrant P forms in soils. Different mechanisms involved in inorganic phosphate solubilization include the production of organic acids, inorganic acids, H_2S, exopolysaccharides (EPS), and siderophores. Meanwhile, organic phosphate solubilization is mediated by the secretion of

extracellular enzymes, including non-specific acid phosphatases (NSAPs), phytases, phosphatases, as well as C–P lyases.

Among the diversity of PSM the genera *Paenibacillus*, *Bacillus*, *Pseudomonas*, *Lactococcus*, *Enterobacter,* and *Alcaligenes* are considered the main bacterial genera capable of solubilizing organic and inorganic phosphates, while *Aspergillus* and *Penicillium* are the main representative fungal genera (Li et al. 2021).

The bioavailability of potassium to plants highly depends on the type of mineral. Two phyllosilicate minerals within the mica family, biotite and muscovite, are of particular interest, due to their richness in essential nutrients, such as Mg, K, Mn, and Zn. These minerals are more readily soluble than others (Sattar et al. 2019). Different mechanisms are associated with K-solubilization by bacteria/rhizobacteria, the major mechanisms being the production of organic and inorganic acids and the production of protons (acidolysis mechanism), similar to what occurs in phosphate solubilization (Etesami et al. 2017).

Siderophores production

Most organisms require iron as a trace element. However, it is often unavailable for direct assimilation by plants and microorganisms. In an aerobic environment, iron exists predominantly as Fe^{3+}, typically in the form of insoluble hydroxides or oxyhydroxides (Rajkumar et al. 2010). Some bacteria acquire iron for their growth and development by synthesizing and secreting siderophores, which are small ferric iron-binding molecules, effectively acting as solubilizing agents for iron from minerals or organic insoluble compounds (Kong and Glick 2017). Additionally, these molecules can also chelate numerous other metals (Johnstone and Nolan 2015).

Siderophore-producing bacteria can either directly stimulate plant growth by improving iron availability or indirectly by inhibiting the activity of plant pathogens in the rhizosphere by limiting iron availability in the environment (Kong and Glick 2017).

Some species are known for their ability to produce siderophores including *Streptomyces* spp. (Terra et al. 2021), *Burkholderia* spp., *Bacillus* spp. (Khan et al. 2016), and *Pseudomonas* spp. (Cornelis and Matthijs 2007). These species have been studied for their role in promoting plant growth, which includes the production of siderophores. In a related study, Ghazy and El-Nahrawy (2021) conducted research focusing on strains of *Bacillus subtilis* and *Pseudomonas koreensis,* selected for their superior performance in siderophore production and antagonistic activity against *Cephalosporium maydis* in an *in vitro* test. In the corn field experiment on the incidence of maize late wilt, the inoculation with the mixture of isolates led to significant increases in catalase (CAT), peroxidase (POX), and polyphenol oxidase (PPO) activities, as well as total chlorophyll and carotenoids, compared to control treatments. Additionally, there was a greater effect in reducing infection and increasing the thickness of the sclerenchymatous sheath layer surrounding the vascular bundles in the corn stalk. These results reflected an increase in productivity and productivity parameters.

Phytohormones production

Phytohormones are crucial plant growth regulators that generally facilitate physiological processes under both normal and stressful conditions. Several genera of bacteria in the rhizosphere and plant-associated produce or modulate essential phytohormones such as gibberellins (GAs), cytokinins, abscisic acid (ABA), auxins, and ethylene.

Auxins are plant hormones that constitute a group of low-molecular-weight molecules, playing a central role in controlling plant growth and development across various environmental conditions (Gomes and Scortecci 2021). The most common naturally occurring auxin is indole-3-acetic acid (IAA) (Çakmakçı et al. 2020, Leontovyčová et al. 2020), which is produced by plants, bacteria, and fungi through at least three different tryptophan-dependent IAA production pathways. This hormone plays a significant role in establishing and maintaining beneficial interaction between plants and associated microorganisms. However, some pathogens can synthesize and secrete IAA as a virulence mechanism during disease development (Fu and Wang 2011).

Cytokinins (CKs) are derivatives of adenine molecules that play a crucial role in influencing various traits of plant growth, development, and physiology (Akhtar et al. 2020). Several studies have reported the secretion of this plant hormone in different bacteria, including *Azospirillum brasilense* (Zaheer et al. 2022), *Pseudomonas fluorescens* (Großkinsky et al. 2016), and *Bacillus subtilis* (Liu et al. 2013). However, the mechanisms involved in CK synthesis are still not well understood (Frébortová and Frébort 2021). Similarly, little has been uncovered about the biosynthesis of Gibberellins (GAs) in bacteria. Gibberellins belong to another class of regulators of growth and developmental processes in vascular plants. It is known that some members from the rhizobia group contain a putative GA biosynthetic operon/gene cluster (Nett et al. 2017, 2020).

Ethylene (ET) governs plant adaptation to stress, but it comes at the expense of growth and development (Ilangumaran and Smith 2017). Under stress conditions, plants increase ET synthesis, which then triggers defense mechanisms. However, due to the multiple effects of this hormone on plant phenotype, elevated levels of ET lead to a series of pleiotropic effects (Ravanbakhsh et al. 2018). To counteract the detrimental effects of ET, many PGPBs encode in their genome the enzyme 1-aminocyclopropane-1-carboxylate (ACC) deaminase. This enzyme has the ability to regulate the production of ethylene (Ali et al. 2014, Chandra et al. 2019, Konkolewska et al. 2020), reducing the impact of various deleterious effects on plants (Glick 2014).

2.5.2. Indirect mechanisms

Induction of Systemic Resistance (ISR)

The induced systemic resistance is a plant's defensive response against a broad spectrum of pathogens and herbivorous insects. This response is also stimulated by beneficial microorganisms (Pieterse et al. 2014). Plants have an innate ability to detect and recognize potential invading microorganisms by identifying pathogen- or microbe-associated molecular patterns (MAMPs or PAMPs) through transmembrane pattern recognition receptors (PRRs). These receptors activate immune defense, known as pattern-triggered immunity (PAMP-triggered immunity - PTI) (Nishad et al. 2020, Yu et al. 2022). Non-pathogenic bacteria can regulate induced systemic resistance (ISR) through signaling pathways that are salicylic acid (SA)-independent mechanisms, such as the jasmonic acid (JA)/ET-dependent pathway (Pieterse et al. 2014, Elías et al. 2018, Yanti et al. 2019, Nguyen et al. 2020). Moreover, certain PGPRs have been reported to trigger an SA-dependent type of ISR that resembles pathogen-induced systemic acquired resistance (SAR) (Ryu et al. 2004a, Niu et al. 2016, Sun et al. 2021).

A variety of bacterial factors are involved in inducing systemic resistance by PGPB, with outer membrane lipopolysaccharides (LPS), siderophore production, and salicylic acid production recognized as the most important (Ramamoorthy et al. 2001).

Many studies have demonstrated the effectiveness of PGPB in inducing systemic resistance (ISR) in plants. The research conducted by Arencibia et al., (2006) showed that *Gluconacetobacter diazotrophicus* possesses and/or produces elicitor molecules that activate sugarcane's defense mechanisms, resulting in resistance to *Xanthomonas albilineans*. Specifically, it was found that the bacterium affected the sugarcane defense system through the production or influence on several key components: genes of the ethylene signaling pathway, proteins regulated by auxins, β-1,3 Glucanase proteins, and ubiquitin genes. Also, Jetiyanon and Plianbangchang (2013) reported that LPS of *Enterobacter asburiae* strain RS83 plays a role in the induction of early defensive-related enzymes in lettuce against soft rot disease caused by *Pectobacterum carotovorum* subsp. carotovorum (Pcc).

Antimicrobials and lytic enzymes

Some bacteria have the ability to suppress phytopathogens' growth and proliferation by producing a series of antimicrobial compounds and enzymes with antagonistic activity. These compounds include a wide range of substances such as ribosomal peptides, polyketides (PKs), non-ribosomal peptides (NRPs), volatile organic compounds (VOCs), and terpenes.

The ribosomally synthesized and posttranslationally modified peptides (RiPPs), have attracted interest as alternatives to conventional antibiotics. The biosynthesis of these compounds begins with the synthesis of linear peptides in ribosomes, which are subsequently modified by enzymes (Li and Rebuffat 2020). While NRPs and PKs are mechanisms independent of the ribosome, both are synthesized by large enzyme complexes called nonribosomal peptide synthetases (NRPSs), which condense amino acids (and sometimes other organic acids) to form nonribosomal peptides. Polyketide synthases (PKSs), on the other hand, condense small carboxylic acids, primarily acetate, and propionate, to form polyketides (Little and Hertweck 2022).

The terpenoids are the most abundant natural products and comprise a diverse class with diverse biological functions (Wang et al. 2018, Helfrich et al. 2019). Terpenoid skeletons consist of C5 isoprene units, which are derived from the precursors isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). In bacteria, these units commonly occur via the 2-C-methyl-D-erythritol 4-phosphate pathway (MEP) (Wang et al. 2018).

Volatile organic compounds (VOCs) consist of carbon-containing compounds with low molecular weight that readily evaporate at normal temperatures and pressures (Wang et al. 2013), They belong to various compound classes, such as alkenes, alcohols, ketones, benzenoids, pyrazines, sulfides, and terpenes (Schulz-Bohm et al. 2017). This compound can have multiple functions in addition to antagonistic activity against another organism, can have an effect to promote plant growth (Park et al. 2015, Almeida et al. 2023), and induce systemic resistance (Ryu et al. 2004b). Responsible for inter- and intra-organism communication, these

compounds play a role in a variety of interactions between plants and their symbiotes, including antagonistic and mutualistic relationships, both below and above ground (Kanchiswamy et al. 2015).

Additionally, there are lytic enzymes produced by PGPBs, including chitinases, cellulases, β-1,3 glucanases, proteases, and lipases that can hydrolyze the polysaccharides that make up the cell wall of various phytopathogenic fungi (Philip et al. 2020, Sharma et al. 2020) and bacteria (Zhang et al. 2015, Dong et al. 2022).

2.6. Metagenomics applied in studies of plant-associated microbiomes

For a long time, microorganisms were discovered using classical microbiological culture-dependent approaches, which allowed the identification of specific groups of microorganisms and bioactive compounds. However, it is important to note that culture-dependent approaches have a limitation for a comprehensive microbial community discovery. This occurs because only those microorganisms capable of growing in the culture media will be recovered, leading to an incomplete representation of the local microbial diversity (Tegtmeier et al. 2021). It is also worth mentioning that there is not always a correlation between the relative abundance of a microorganism in its natural environment and its ability to grow in synthetic culture media (Poyet et al. 2019). This means that a microorganism may be abundant in its natural habitat but may not grow well or at all in laboratory conditions. In this way, alternative approaches, such as culture-independent methods, are necessary to obtain a more comprehensive understanding of microbial communities.

The culture-independent methods have become possible due to technological advances in molecular biology and high-throughput omics. These methods have greatly promoted efforts to elucidate the structure and dynamics inherent in microbiomes, such as plant-associated microbiomes (Levy et al. 2018). In recent years, plant microbiome research has increased drastically, providing significant insights concerning the roles of plant-associated microbiota, particularly in terms of their agricultural relevance on plant growth promotion and pathogenesis.

Metagenomic approaches serve as the foundation for various other

techniques and the increasing abundance of omics data information from a wide range of environments offers numerous opportunities to advance our understanding of the field of plant-associated microbiomes.

Many studies describing microbial communities rely on short-read amplicon sequencing, which has limitations in understanding functional capabilities and providing low taxonomic resolution at the species level (Soto-Giron et al. 2021). In contrast, long-read metagenomic sequencing has emerged as a powerful tool, enabling the retrieval of metagenome-assembled genomes (MAGs) with high completeness. This approach has greatly improved our understanding of microbial populations and their interactions with hosts or their environment, as well as facilitated the discovery of novel species, thus contributing to the reduction of what is often referred to as the ‘microbial dark matter’ (Setubal 2021). In microbiome analyses using marker genes (metabarcoding), high-throughput sequencing of long reads positively impacts the increase in the phylogenetic signal, which is attributed to greater coverage of the target gene, which in turn leads to a significant rise in the resolution of the taxonomic profile of the community. Furthermore, long-read sequencing has been instrumental in enhancing taxonomic resolution at both the species and genotype levels (Armanhi et al. 2016, Walder et al. 2017, Martijn et al. 2019, Tedersoo et al. 2020).

3. REFERENCES

Akhtar SS, Mekureyaw MF, Pandey C, Roitsch T (2020) Role of Cytokinins for Interactions of Plants With Microbial Pathogens and Pest Insects. **Frontiers in Plant Science** 10: 1777. https://doi.org/10.3389/fpls.2019.01777

Ali S, Charles TC, Glick BR (2014) Amelioration of high salinity stress damage by plant growth-promoting bacterial endophytes that contain ACC deaminase. **Plant Physiology and Biochemistry** 80: 160–167. https://doi.org/10.1016/j.plaphy.2014.04.003

Almeida OAC, De Araujo NO, Mulato ATN, Persinoti GF, Sforça ML, Calderan-Rodrigues MJ, Oliveira JVDC (2023) Bacterial volatile organic compounds (VOCs) promote growth and induce metabolic changes in rice. **Frontiers in Plant Science** 13: 1056082. https://doi.org/10.3389/fpls.2022.1056082

Andreote FD, Gumiere T, Durrer A (2014) Exploring interactions of plant microbiomes. **Scientia Agricola** 71: 528–539. https://doi.org/10.1590/0103-9016-2014-0195

Arencibia AD, Vinagre F, Estevez Y, Bernal A, Perez J, Cavalcanti J, Santana I, Hemerly AS (2006) *Gluconoacetobacter diazotrophicus* Elicitate a Sugarcane Defense Response Against a Pathogenic Bacteria *Xanthomonas albilineans*. **Plant Signaling & Behavior** 1: 265–273. https://doi.org/10.4161/psb.1.5.3390

Armanhi JSL, de Souza RSC, Damasceno N de B, de Araújo LM, Imperial J, Arruda P (2018) A Community-Based Culture Collection for Targeting Novel Plant Growth-Promoting Bacteria from the Sugarcane Microbiome. **Frontiers in Plant Science** 8: 2191. https://doi.org/10.3389/fpls.2017.02191

Armanhi JSL, de Souza RSC, de Araújo LM, Okura VK, Mieczkowski P, Imperial J, Arruda P (2016) Multiplex amplicon sequencing for microbe identification in community-based culture collections. **Scientific Reports** 6: 29543. https://doi.org/10.1038/srep29543

Banerjee S, van der Heijden MGA (2023) Soil microbiomes and one health. **Nature Reviews Microbiology** 21: 6–20. https://doi.org/10.1038/s41579-022-00779-w

Banerjee S, Schlaeppi K, van der Heijden MGA (2018) Keystone taxa as drivers of microbiome structure and functioning. **Nature Reviews Microbiology** 16: 567–576. https://doi.org/10.1038/s41579-018-0024-1

Berg G, Köberl M, Rybakova D, Müller H, Grosch R, Smalla K (2017) Plant microbial diversity is suggested as the key to future biocontrol and health trends. **FEMS Microbiology Ecology** 93: fix050. https://doi.org/10.1093/femsec/fix050

Berg G, Rybakova D, Fischer D, Cernava T, Vergès M-CC, Charles T, Chen X, Cocolin L, Eversole K, Corral GH, Kazou M, Kinkel L, Lange L, Lima N, Loy A, Macklin JA, Maguin E, Mauchline T, McClure R, Mitter B, Ryan M, Sarand I, Smidt H, Schelkle B, Roume H, Kiran GS, Selvin J, Souza RSC de, van Overbeek L, Singh BK, Wagner M, Walsh A, Sessitsch A, Schloter M (2020) Microbiome definition re-visited: old concepts and new challenges. **Microbiome** 8: 103.

https://doi.org/10.1186/s40168-020-00875-0

Çakmakçı R, Mosber G, Milton AH, Alatürk F, Ali B (2020) The Effect of Auxin and Auxin-Producing Bacteria on the Growth, Essential Oil Yield, and Composition in Medicinal and Aromatic Plants. **Current Microbiology** 77: 564–577. https://doi.org/10.1007/s00284-020-01917-4

de Carvalho LAL, Teheran-Sierra LG, Funnicelli MIG, da Silva RC, Campanari MFZ, de Souza RSC, Arruda P, Soares MA, Pinheiro DG (2021) Farming systems influence the compositional, structural, and functional characteristics of the sugarcane-associated microbiome. **Microbiological Research** 252: 126866. https://doi.org/10.1016/j.micres.2021.126866

Chandra D, Srivastava R, Gupta VVSR, Franco CMM, Sharma AK (2019) Evaluation of ACC-deaminase-producing rhizobacteria to alleviate water-stress impacts in wheat (*Triticum aestivum* L.) plants. **Canadian Journal of Microbiology** 65: 387–403. https://doi.org/10.1139/cjm-2018-0636

Cheng YT, Zhang L, He SY (2019) Plant-Microbe Interactions Facing Environmental Challenge. **Cell Host & Microbe** 26: 183–192. https://doi.org/10.1016/j.chom.2019.07.009

Compant S, Clément C, Sessitsch A (2010) Plant growth-promoting bacteria in the rhizo- and endosphere of plants: Their role, colonization, mechanisms involved and prospects for utilization. **Soil Biology and Biochemistry** 42: 669–678. https://doi.org/10.1016/j.soilbio.2009.11.024

Compant S, Samad A, Faist H, Sessitsch A (2019) A review on the plant microbiome: Ecology, functions, and emerging trends in microbial application. **Journal of Advanced Research** 19: 29–37. https://doi.org/10.1016/j.jare.2019.03.004

CONAB (2022a) 8 Acomp. safra bras. cana, v. 8 - Safra 2021/22, n. 4 - Quarto levantamento. Conab Brasília. Available from: https://www.conab.gov.br/info-agro/safras/cana/boletim-da-safra-de-cana-de-acucar/item/download/45236_22907ba0da29e3364b66c913e9557ec9.

CONAB (2022b) 9 Acomp. safra bras. cana, v. 9 - Safra 2022/23, n. 2 - Segundo levantamento. Conab Brasília. Available from: https://www.conab.gov.br/info-agro/safras/cana/boletim-da-safra-de-cana-de-acucar/item/download/43933_4c496920bffcfbfb45deb447d553a497.

Cornelis P, Matthijs S (2007) Pseudomonas siderophores and their biological significance. In: Microbial siderophores. **Springer**, 193–203.

Cuscó A, Pérez D, Viñes J, Fàbregas N, Francino O (2021) Long-read metagenomics retrieves complete single-contig bacterial genomes from canine feces. **BMC Genomics** 22: 330. https://doi.org/10.1186/s12864-021-07607-0

Daniel R (2004) The soil metagenome – a rich resource for the discovery of novel natural products. **Current Opinion in Biotechnology** 15: 199–204. https://doi.org/10.1016/j.copbio.2004.04.005

Dastogeer KMG, Tumpa FH, Sultana A, Akter MA, Chakraborty A (2020) Plant microbiome–an account of the factors that shape community composition and diversity. **Current Plant Biology** 23: 100161. https://doi.org/10.1016/j.cpb.2020.100161

De Zutter N, Ameye M, Bekaert B, Verwaeren J, De Gelder L, Audenaert K (2022) Uncovering New Insights and Misconceptions on the Effectiveness of Phosphate Solubilizing Rhizobacteria in Plants: A Meta-Analysis. **Frontiers in Plant Science** 13: 858804. https://doi.org/10.3389/fpls.2022.858804

Dixon R, Kahn D (2004) Genetic regulation of biological nitrogen fixation. **Nature Reviews Microbiology** 2: 621–631. https://doi.org/10.1038/nrmicro954

Dong H, Xu X, Gao R, Li Y, Li A, Yao Q, Zhu H (2022) Myxococcus xanthus R31 Suppresses Tomato Bacterial Wilt by Inhibiting the Pathogen *Ralstonia solanacearum* With Secreted Proteins. **Frontiers in Microbiology** 12: 801091. https://doi.org/10.3389/fmicb.2021.801091

Elías JM, Guerrero-Molina MF, Martínez-Zamora MG, Díaz-Ricci JC, Pedraza RO (2018) Role of ethylene and related gene expression in the interaction between strawberry plants and the plant growth-promoting bacterium *Azospirillum brasilense*. Papen H (Ed.). **Plant Biology** 20: 490–496. https://doi.org/10.1111/plb.12697

Erenstein O, Jaleta M, Sonder K, Mottaleb K, Prasanna BM (2022) Global maize production, consumption and trade: trends and R&D implications. **Food Security** 14: 1295–1319. https://doi.org/10.1007/s12571-022-01288-7

Etesami H, Emami S, Alikhani HA (2017) Potassium solubilizing bacteria (KSB):: Mechanisms, promotion of plant growth, and future prospects: A review. **Journal of soil science and plant nutrition** 17: 897–911. https://doi.org/10.4067/S0718-95162017000400005

Fierer N (2017) Embracing the unknown: disentangling the complexities of the soil microbiome. **Nature Reviews Microbiology** 15: 579–590. https://doi.org/10.1038/nrmicro.2017.87

Frébortová J, Frébort I (2021) Biochemical and Structural Aspects of Cytokinin Biosynthesis and Degradation in Bacteria. **Microorganisms** 9: 1314. https://doi.org/10.3390/microorganisms9061314

Fu J, Wang S (2011) Insights into Auxin Signaling in Plant-Pathogen Interactions. **Frontiers in Plant Science** 2. https://doi.org/10.3389/fpls.2011.00074

Ghazy N, El-Nahrawy S (2021) Siderophore production by *Bacillus subtilis* MF497446 and *Pseudomonas koreensis* MG209738 and their efficacy in controlling *Cephalosporium maydis* in maize plant. **Archives of Microbiology** 203: 1195–1209. https://doi.org/10.1007/s00203-020-02113-5

Glick BR (2014) Bacteria with ACC deaminase can promote plant growth and help to feed the world. **Microbiological Research** 169: 30–39. https://doi.org/10.1016/j.micres.2013.09.009

Godshall MA, Legendre BL (2003) SUGAR | Sugarcane. In: Encyclopedia of Food

Sciences and Nutrition. **Elsevier**, 5645–5651. https://doi.org/10.1016/B0-12-227055-X/01160-3

Gomes GLB, Scortecci KC (2021) Auxin and its role in plant development: structure, signalling, regulation and response mechanisms. Wicke S (Ed.). **Plant Biology** 23: 894–904. https://doi.org/10.1111/plb.13303

Großkinsky DK, Tafner R, Moreno MV, Stenglein SA, García de Salamone IE, Nelson LM, Novák O, Strnad M, van der Graaff E, Roitsch T (2016) Cytokinin production by *Pseudomonas fluorescens* G20-18 determines biocontrol activity against Pseudomonas syringae in Arabidopsis. **Scientific Reports** 6: 23310. https://doi.org/10.1038/srep23310

Hamonts K, Trivedi P, Garg A, Janitz C, Grinyer J, Holford P, Botha FC, Anderson IC, Singh BK (2018) Field study reveals core plant microbiota and relative importance of their drivers. **Environmental Microbiology** 20: 124–140. https://doi.org/10.1111/1462-2920.14031

Haro-Moreno JM, López-Pérez M, Rodriguez-Valera F (2021) Enhanced Recovery of Microbial Genes and Genomes From a Marine Water Column Using Long-Read Metagenomics. **Frontiers in Microbiology** 12: 708782. https://doi.org/10.3389/fmicb.2021.708782

Hassani MA, Durán P, Hacquard S (2018) Microbial interactions within the plant holobiont. **Microbiome** 6: 58. https://doi.org/10.1186/s40168-018-0445-0

Heinrichs R, Otto R, Magalhães A, Meirelles GC (2017) Importance of sugarcane in Brazilian and world bioeconomy. In: Knowledge-Driven Developments in the Bioeconomy. **Springer**, 205–217.

Helfrich EJN, Lin G-M, Voigt CA, Clardy J (2019) Bacterial terpene biosynthesis: challenges and opportunities for pathway engineering. **Beilstein Journal of Organic Chemistry** 15: 2889–2906. https://doi.org/10.3762/bjoc.15.283

Hoang NV, Furtado A, Botha FC, Simmons BA, Henry RJ (2015) Potential for Genetic Improvement of Sugarcane as a Source of Biomass for Biofuels. **Frontiers in Bioengineering and Biotechnology** 3. https://doi.org/10.3389/fbioe.2015.00182

Ilangumaran G, Smith DL (2017) Plant Growth Promoting Rhizobacteria in Amelioration of Salinity Stress: A Systems Biology Perspective. **Frontiers in Plant Science** 8: 1768. https://doi.org/10.3389/fpls.2017.01768

Jetiyanon K, Plianbangchang P (2013) Lipopolysaccharide of Enterobacter asburiae strain RS83: A bacterial determinant for induction of early defensive enzymes in Lactuca sativa against soft rot disease. **Biological Control** 67: 301–307. https://doi.org/10.1016/j.biocontrol.2013.09.014

Johnstone TC, Nolan EM (2015) Beyond iron: non-classical biological functions of bacterial siderophores. **Dalton Transactions** 44: 6320–6339. https://doi.org/10.1039/C4DT03559C

Kanchiswamy CN, Malnoy M, Maffei ME (2015) Chemical diversity of microbial volatiles and their potential for plant growth and productivity. **Frontiers in Plant**

Science 6. https://doi.org/10.3389/fpls.2015.00151

Khan A, Doshi HV, Thakur MC (2016) *Bacillus* spp.: a prolific siderophore producer. **Bacilli and agrobiotechnology**: 309–323.

Kong Z, Glick BR (2017) The Role of Plant Growth-Promoting Bacteria in Metal Phytoremediation. In: Advances in Microbial Physiology. **Elsevier**, 97–132. https://doi.org/10.1016/bs.ampbs.2017.04.001

Konkolewska A, Piechalak A, Ciszewska L, Antos-Krzemińska N, Skrzypczak T, Hanć A, Sitko K, Małkowski E, Barałkiewicz D, Małecka A (2020) Combined use of companion planting and PGPR for the assisted phytoextraction of trace metals (Zn, Pb, Cd). **Environmental Science and Pollution Research** 27: 13809–13825. https://doi.org/10.1007/s11356-020-07885-3

Lebeis SL (2015) Greater than the sum of their parts: characterizing plant microbiomes at the community-level. **Current Opinion in Plant Biology** 24: 82–86. https://doi.org/10.1016/j.pbi.2015.02.004

Leontovyčová H, Trdá L, Dobrev PI, Šašek V, Gay E, Balesdent M-H, Burketová L (2020) Auxin biosynthesis in the phytopathogenic fungus Leptosphaeria maculans is associated with enhanced transcription of indole-3-pyruvate decarboxylase LmIPDC2 and tryptophan aminotransferase LmTAM1. **Research in Microbiology** 171: 174–184. https://doi.org/10.1016/j.resmic.2020.05.001

Levy A, Salas Gonzalez I, Mittelviefhaus M, Clingenpeel S, Herrera Paredes S, Miao J, Wang K, Devescovi G, Stillman K, Monteiro F, Rangel Alvarez B, Lundberg DS, Lu T-Y, Lebeis S, Jin Z, McDonald M, Klein AP, Feltcher ME, Rio TG, Grant SR, Doty SL, Ley RE, Zhao B, Venturi V, Pelletier DA, Vorholt JA, Tringe SG, Woyke T, Dangl JL (2018) Genomic features of bacterial adaptation toplants. **Nature Genetics** 50: 138–150. https://doi.org/10.1038/s41588-017-0012-9

Li J, Lu J, Wang H, Fang Z, Wang X, Feng S, Wang Z, Yuan T, Zhang S, Ou S, Yang X, Wu Z, Du X, Tang L, Liao B, Shu W, Jia P, Liang J (2021) A comprehensive synthesis unveils the mysteries of phosphate‐solubilizing microbes. **Biological Reviews** 96: 2771–2793. https://doi.org/10.1111/brv.12779

Li Y, Rebuffat S (2020) The manifold roles of microbial ribosomal peptide–based natural products in physiology and ecology. **Journal of Biological Chemistry** 295: 34–54. https://doi.org/10.1074/jbc.REV119.006545

Little RF, Hertweck C (2022) Chain release mechanisms in polyketide and non-ribosomal peptide biosynthesis. **Natural Product Reports** 39: 163–205. https://doi.org/10.1039/D1NP00035G

Liu F, Xing S, Ma H, Du Z, Ma B (2013) Cytokinin-producing, plant growth-promoting rhizobacteria that confer resistance to drought stress in *Platycladus orientalis* container seedlings. **Applied Microbiology and Biotechnology** 97: 9155–9164. https://doi.org/10.1007/s00253-013-5193-2

Mapelli F, Mengoni A, Riva V, Borin S (2022) Bacterial culturing is crucial to boost sustainable agriculture. **Trends in Microbiology**: S0966842X22002888. https://doi.org/10.1016/j.tim.2022.10.005

Martijn J, Lind AE, Schön ME, Spiertz I, Juzokaite L, Bunikis I, Pettersson OV, Ettema TJG (2019) Confident phylogenetic identification of uncultured prokaryotes through long read amplicon sequencing of the 16S-ITS-23S rRNA operon. **Environmental Microbiology** 21: 2485–2498. https://doi.org/10.1111/1462-2920.14636

Mendes R, Kruijt M, de Bruijn I, Dekkers E, van der Voort M, Schneider JHM, Piceno YM, DeSantis TZ, Andersen GL, Bakker PAHM, Raaijmakers JM (2011) Deciphering the Rhizosphere Microbiome for Disease-Suppressive Bacteria. **Science** 332: 1097–1100. https://doi.org/10.1126/science.1203980

Mitter B, Pfaffenbichler N, Sessitsch A (2016) Plant–microbe partnerships in 2020. **Microbial Biotechnology** 9: 635–640. https://doi.org/10.1111/1751-7915.12382

Mus F, Crook MB, Garcia K, Garcia Costas A, Geddes BA, Kouri ED, Paramasivan P, Ryu M-H, Oldroyd GED, Poole PS, Udvardi MK, Voigt CA, Ané J-M, Peters JW (2016) Symbiotic Nitrogen Fixation and the Challenges to Its Extension to Nonlegumes. Kelly RM (Ed.). **Applied and Environmental Microbiology** 82: 3698–3710. https://doi.org/10.1128/AEM.01055-16

Muthukumarasamy R, Revathi G, Vadivelu M, Arun K (2017) Isolation of bacterial strains possessing nitrogen-fixation, phosphate and potassium-solubilization and their inoculation effects on sugarcane. **Indian Journal of Experimental Biology** 55: 161–170.

Nett RS, Nguyen H, Nagel R, Marcassa A, Charles TC, Friedberg I, Peters RJ (2020) Unraveling a Tangled Skein: Evolutionary Analysis of the Bacterial Gibberellin Biosynthetic Operon. Kent AD (Ed.). **mSphere** 5: e00292-20. https://doi.org/10.1128/mSphere.00292-20

Nett RS, Montanares M, Marcassa A, Lu X, Nagel R, Charles TC, Hedden P, Rojas MC, Peters RJ (2017) Elucidation of gibberellin biosynthesis in bacteria reveals convergent evolution. **Nature Chemical Biology** 13: 69–74. https://doi.org/10.1038/nchembio.2232

Nguyen NH, Trotel-Aziz P, Villaume S, Rabenoelina F, Schwarzenberg A, Nguema-Ona E, Clément C, Baillieul F, Aziz A (2020) *Bacillus subtilis* and *Pseudomonas fluorescens* Trigger Common and Distinct Systemic Immune Responses in Arabidopsis thaliana Depending on the Pathogen Lifestyle. **Vaccines** 8: 503. https://doi.org/10.3390/vaccines8030503

Nishad R, Ahmed T, Rahman VJ, Kareem A (2020) Modulation of Plant Defense System in Response to Microbial Interactions. **Frontiers in Microbiology** 11: 1298. https://doi.org/10.3389/fmicb.2020.01298

Niu D, Wang X, Wang Y, Song X, Wang J, Guo J, Zhao H (2016) *Bacillus cereus* AR156 activates PAMP-triggered immunity and induces a systemic acquired resistance through a NPR1 -and SA-dependent signaling pathway. **Biochemical and Biophysical Research Communications** 469: 120–125. https://doi.org/10.1016/j.bbrc.2015.11.081

Park Y-S, Dutta S, Ann M, Raaijmakers JM, Park K (2015) Promotion of plant growth by *Pseudomonas fluorescens* strain SS101 via novel volatile organic compounds.

Biochemical and Biophysical Research Communications 461: 361–365. https://doi.org/10.1016/j.bbrc.2015.04.039

Philip NV, Koteshwara A, Kiran GA, Raja S, Subrahmanyam VM, Chandrashekar HR (2020) Statistical Optimization for Coproduction of Chitinase and Beta 1, 4-Endoglucanase by Chitinolytic *Paenibacillus elgii* PB1 Having Antifungal Activity. **Applied Biochemistry and Biotechnology** 191: 135–150. https://doi.org/10.1007/s12010-020-03235-8

Pieterse CMJ, Zamioudis C, Berendsen RL, Weller DM, Van Wees SCM, Bakker PAHM (2014) Induced Systemic Resistance by Beneficial Microbes. **Annual Review of Phytopathology** 52: 347–375. https://doi.org/10.1146/annurev-phyto-082712-102340

Poyet M, Groussin M, Gibbons SM, Avila-Pacheco J, Jiang X, Kearney SM, Perrotta AR, Berdy B, Zhao S, Lieberman TD, Swanson PK, Smith M, Roesemann S, Alexander JE, Rich SA, Livny J, Vlamakis H, Clish C, Bullock K, Deik A, Scott J, Pierce KA, Xavier RJ, Alm EJ (2019) A library of human gut bacterial isolates paired with longitudinal multiomics data enables mechanistic microbiome research. **Nature Medicine** 25: 1442–1452. https://doi.org/10.1038/s41591-019-0559-3

Rajkumar M, Ae N, Prasad MNV, Freitas H (2010) Potential of siderophore-producing bacteria for improving heavy metal phytoextraction. **Trends in Biotechnology** 28: 142–149. https://doi.org/10.1016/j.tibtech.2009.12.002

Ramamoorthy V, Viswanathan R, Raguchander T, Prakasam V, Samiyappan R (2001) Induction of systemic resistance by plant growth promoting rhizobacteria in crop plants against pests and diseases. **Crop Protection** 20: 1–11. https://doi.org/10.1016/S0261-2194(00)00056-9

Ravanbakhsh M, Sasidharan R, Voesenek LACJ, Kowalchuk GA, Jousset A (2018) Microbial modulation of plant ethylene signaling: ecological and evolutionary consequences. **Microbiome** 6: 52. https://doi.org/10.1186/s40168-018-0436-1

Rawat P, Das S, Shankhdhar D, Shankhdhar SC (2021) Phosphate-Solubilizing Microorganisms: Mechanism and Their Role in Phosphate Solubilization and Uptake. **Journal of Soil Science and Plant Nutrition** 21: 49–68. https://doi.org/10.1007/s42729-020-00342-7

Risely A (2020) Applying the core microbiome to understand host–microbe systems. Tate A (Ed.). **Journal of Animal Ecology** 89: 1549–1558. https://doi.org/10.1111/1365-2656.13229

Ryu C-M, Murphy JF, Mysore KS, Kloepper JW (2004a) Plant growth-promoting rhizobacteria systemically protect *Arabidopsis thaliana* against *Cucumber mosaic virus* by a salicylic acid and NPR1-independent and jasmonic acid-dependent signaling pathway. **The Plant Journal** 39: 381–392. https://doi.org/10.1111/j.1365-313X.2004.02142.x

Ryu C-M, Farag MA, Hu C-H, Reddy MS, Kloepper JW, Paré PW (2004b) Bacterial Volatiles Induce Systemic Resistance in Arabidopsis. 134: **Plant Physiology** 1017–1026. https://doi.org/10.1104/pp.103.026583

Sainju UM, Ghimire R, Pradhan GP (2019) Nitrogen fertilization I: Impact on crop, soil, and environment. **Nitrogen Fixation**.

Sattar A, Naveed M, Ali M, Zahir ZA, Nadeem SM, Yaseen M, Meena VS, Farooq M, Singh R, Rahman M, Meena HN (2019) Perspectives of potassium solubilizing microbes in sustainable food production system: A review. **Applied Soil Ecology** 133: 146–159. https://doi.org/10.1016/j.apsoil.2018.09.012

Scholz MB, Lo C-C, Chain PS (2012) Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. **Current Opinion in Biotechnology** 23: 9–15. https://doi.org/10.1016/j.copbio.2011.11.013

Schulz-Bohm K, Martín-Sánchez L, Garbeva P (2017) Microbial Volatiles: Small Molecules with an Important Role in Intra- and Inter-Kingdom Interactions. Frontiers in Microbiology 8: 2484. https://doi.org/10.3389/fmicb.2017.02484

Setubal JC (2021) Metagenome-assembled genomes: concepts, analogies, and challenges. **Biophysical Reviews** 13: 905–909. https://doi.org/10.1007/s12551-021-00865-y

Sharma S, Kumar S, Khajuria A, Ohri P, Kaur R, Kaur R (2020) Biocontrol potential of chitinases produced by newly isolated *Chitinophaga* sp. S167. **World Journal of Microbiology and Biotechnology** 36: 90. https://doi.org/10.1007/s11274-020-02864-9

Shen W, Ren H (2021) TaxonKit: A practical and efficient NCBI taxonomy toolkit. **Journal of Genetics and Genomics** 48: 844–850. https://doi.org/10.1016/j.jgg.2021.03.006

Soto-Giron MJ, Kim J-N, Schott E, Tahmin C, Ishoey T, Mincer TJ, DeWalt J, Toledo G (2021) The Edible Plant Microbiome represents a diverse genetic reservoir with functional potential in the human host. **Scientific Reports** 11: 24017. https://doi.org/10.1038/s41598-021-03334-4

Soumare A, Diedhiou AG, Thuita M, Hafidi M, Ouhdouch Y, Gopalakrishnan S, Kouisni L (2020) Exploiting Biological Nitrogen Fixation: A Route Towards a Sustainable Agriculture. **Plants** 9: 1011. https://doi.org/10.3390/plants9081011

de Souza RSC, Armanhi JSL, Damasceno N de B, Imperial J, Arruda P (2019) Genome Sequences of a Plant Beneficial Synthetic Bacterial Community Reveal Genetic Features for Successful Plant Colonization. **Frontiers in Microbiology** 10: 1779. https://doi.org/10.3389/fmicb.2019.01779

de Souza RSC, Okura VK, Armanhi JSL, Jorrín B, Lozano N, da Silva MJ, González-Guerrero M, de Araújo LM, Verza NC, Bagheri HC, Imperial J, Arruda P (2016) Unlocking the bacterial and fungal communities assemblages of sugarcane microbiome. **Scientific Reports** 6: 28774. https://doi.org/10.1038/srep28774

Sun K, Xie X-G, Lu F, Zhang F-M, Zhang W, He W, Dai C-C (2021) Peanut preinoculation with a root endophyte induces plant resistance to soil-borne pathogen *Fusarium oxysporum* via activation of salicylic acid-dependent signaling. **Plant and Soil** 460: 297–312. https://doi.org/10.1007/s11104-020-04807-7

Tedersoo L, Anslan S, Bahram M, Kõljalg U, Abarenkov K (2020) Identifying the 'unidentified' fungi: a global-scale long-read third-generation sequencing approach. **Fungal Diversity** 103: 273–293. https://doi.org/10.1007/s13225-020-00456-4

Tegtmeier D, Hurka S, Mihajlovic S, Bodenschatz M, Schlimbach S, Vilcinskas A (2021) Culture-Independent and Culture-Dependent Characterization of the Black Soldier Fly Gut Microbiome Reveals a Large Proportion of Culturable Bacteria with **Potential for Industrial Applications. Microorganisms** 9: 1642. https://doi.org/10.3390/microorganisms9081642

Teheran-Sierra LG, Funnicelli MIG, de Carvalho LAL, Ferro MIT, Soares MA, Pinheiro DG (2021) Bacterial communities associated with sugarcane under different agricultural management exhibit a diversity of plant growth-promoting traits and evidence of synergistic effect. **Microbiological Research** 247: 126729. https://doi.org/10.1016/j.micres.2021.126729

Terra L, Ratcliffe N, Castro HC, Vicente ACP, Dyson P (2021) Biotechnological Potential of *Streptomyces* Siderophores as New Antibiotics. **Current Medicinal Chemistry** 28: 1407–1421. https://doi.org/10.2174/0929867327666200510235512

Tew TL, Cobill RM (2008) Genetic improvement of sugarcane (Saccharum spp.) as an energy crop. In: Genetic improvement of bioenergy crops. **Springer**, 273–294.

Tilman D, Balzer C, Hill J, Befort BL (2011) Global food demand and the sustainable intensification of agriculture. **Proceedings of the National Academy of Sciences** 108: 20260–20264. https://doi.org/10.1073/pnas.1116437108

Torsvik V, Øvreås L (2002) Microbial diversity and function in soil: from genes to ecosystems. **Current Opinion in Microbiology** 5: 240–245. https://doi.org/10.1016/S1369-5274(02)00324-7

Trivedi P, Leach JE, Tringe SG, Sa T, Singh BK (2020) Plant–microbiome interactions: from community assembly to plant health. **Nature Reviews Microbiology** 18: 607–621. https://doi.org/10.1038/s41579-020-0412-1

Walder F, Schlaeppi K, Wittwer R, Held AY, Vogelgsang S, van der Heijden MGA (2017) Community Profiling of Fusarium in Combination with Other Plant-Associated Fungi in Different Crop Species Using SMRT Sequencing. **Frontiers in Plant Science** 8: 2019. https://doi.org/10.3389/fpls.2017.02019

Wang C, Liwei M, Park J-B, Jeong S-H, Wei G, Wang Y, Kim S-W (2018) Microbial Platform for Terpenoid Production: *Escherichia coli* and Yeast. **Frontiers in Microbiology** 9: 2460. https://doi.org/10.3389/fmicb.2018.02460

Wang C, Wang Z, Qiao X, Li Z, Li F, Chen M, Wang Y, Huang Y, Cui H (2013) Antifungal activity of volatile organic compounds from *Streptomyces alboflavus* TD-1. **FEMS Microbiology Letters** 341: 45–51. https://doi.org/10.1111/1574-6968.12088

Wen A, Havens KL, Bloch SE, Shah N, Higgins DA, Davis-Richardson AG, Sharon J, Rezaei F, Mohiti-Asli M, Johnson A, Abud G, Ane J-M, Maeda J, Infante V, Gottlieb SS, Lorigan JG, Williams L, Horton A, McKellar M, Soriano D, Caron Z, Elzinga H, Graham A, Clark R, Mak S-M, Stupin L, Robinson A, Hubbard N, Broglie R, Tamsir A, Temme K (2021) Enabling Biological Nitrogen Fixation for Cereal Crops in

Fertilized Fields. **ACS Synthetic Biology 10: 3264–3277. https://doi.org/10.1021/acssynbio.1c00049**

Yanti Y, Warnita, Reflin (2019) Involvement of Jasmonic Acid in the Induced Systemic Resistance of Tomato against *Ralstonia syzigiisub* sp. indonesiensis by Indigenous Endophyte Bacteria. **IOP Conference Series: Earth and Environmental Science** 347: 012024. https://doi.org/10.1088/1755-1315/347/1/012024

Yeoh YK, Paungfoo-Lonhienne C, Dennis PG, Robinson N, Ragan MA, Schmidt S, Hugenholtz P (2016) The core root microbiome of sugarcanes cultivated under varying nitrogen fertilizer application: N fertilizer and sugarcane root microbiota. **Environmental Microbiology** 18: 1338–1351. https://doi.org/10.1111/1462-2920.12925

Yu Y, Gui Y, Li Z, Jiang C, Guo J, Niu D (2022) Induced Systemic Resistance for Improving Plant Immunity by Beneficial Microbes. **Plants** (Basel, Switzerland) 11: 386. https://doi.org/10.3390/plants11030386

Zaheer MS, Ali HH, Iqbal MA, Erinle KO, Javed T, Iqbal J, Hashmi MIU, Mumtaz MZ, Salama EAA, Kalaji HM, Wróbel J, Dessoky ES (2022) Cytokinin Production by Azospirillum brasilense Contributes to Increase in Growth, Yield, Antioxidant, and Physiological Systems of Wheat (Triticum aestivum L.). **Frontiers in Microbiology** 13: 886041. https://doi.org/10.3389/fmicb.2022.886041

Zhang S, Chakrabarty PK, Fleites LA, Rayside PA, Hopkins DL, Gabriel DW (2015) Three New Pierce's Disease Pathogenicity Effectors Identified Using *Xylella fastidiosa* Biocontrol Strain EB92-1. Arnold DL (Ed.). **PLOS ONE** 10: e0133796. https://doi.org/10.1371/journal.pone.0133796

APPENDIX

APPENDIX A. Supplementary methodology

Plant microbiome studies

To obtain a summary of the work related to phytomicrobiomes, we collected records from the Web of Science [2] online tool, using the following search term on title, abstract and keywords: ‘plant microbiome’ OR (('core microbiota') AND (plant) OR (('core plant microbiome') OR ('core plant microbiota'))). The plant or group was identified based on a match with a list of common or scientific plant names obtained from the Apsnet [3] website.

Bacterial genera associated with sugarcane

The data of bacterial genera described as plant-growth promoters isolated from sugarcane were collected from the Web of Science online tool, using the following search terms on title, abstract, and keywords: (((PGPR) OR (PGPB) OR ("plant growth-promoting") OR ("plant growth promotion") OR ("plant growth*") OR ("plant growth-promoting bacteria") OR ("plant growth-promoting rhizobacteria")) AND ((sugar*cane) OR (Saccharum) OR (isolate from sugar*cane) OR (isolated from sugar*cane) OR (sugarcane endophytic) OR (sugarcane epiphytic))).

A list of bacterial genera available in the NCBI Taxonomy was retrieved using the TaxonKit tool (Shen and Ren 2021). This list was then used to identify matching genera in the records. Additionally, we conducted searches using terms related to traits associated with plant growth promotion from the recovered queries.

The matching results for species or genera recognized as sugarcane pathogens (species: *Acidovorax avenae*, *Leifsonia xyli*, *Herbaspirillum rubrisubalbicans*; genus *Phytoplasma*, *Mycoplasma*) were not accounted for.

[2] http://www.webofknowledge.com
[3] https://www.apsnet.org/edcenter/resources/commonnames/Pages/default.aspx

APPENDIX B. The related references in Table 1.

1. Antunes JEL, Lyra MCCP, Ollero FJ, Freitas ADS, Oliveira LMS, Araújo ASF, et al. Diversity of plant growth-promoting bacteria associated with sugarcane. Genet Mol Res [Internet]. 2017 [cited 2022 Dec 22];16(2). Available from: http://www.funpecrp.com.br/gmr/year2017/vol16-2/pdf/gmr-16-02-gmr.16029662.pdf

2. Taulé C, Mareque C, Barlocco C, Hackembruch F, Reis VM, Sicardi M, et al. The contribution of nitrogen fixation to sugarcane (Saccharum officinarum L.), and the identification and characterization of part of the associated diazotrophic bacterial community. Plant Soil. 2012 Jul;356(1–2):35–49.

3. Santos RM dos, Rigobelo EC. Growth-Promoting Potential of Rhizobacteria Isolated From Sugarcane. Front Sustain Food Syst. 2021 Apr 16;5:596269.

4. Pitiwittayakul N, Wongsorn D, Tanasupawat S. Characterisation of Plant Growth-Promoting Endophytic Bacteria from Sugarcane and Their Antagonistic Activity against Fusarium moniliforme. Trop Life Sci Res. 2021 Sep 30;32(3):97–118.

5. Silva JC da, Santos LDS, Faria PSA, Silva FG, Rubio Neto A, Martins PF, et al. Multifunctional characteristics of Acinetobacter lwoffii Bac109 for growth promotion and colonization in micropropagated sugarcane. Pesqui Agropecuária Trop. 2021;51:e69373.

6. CG Bhakta Institute of Biotechnology, Uka Tarsadia University, Bardoli, 394350 - Gujarat, India, Patel P, Shah R, CG Bhakta Institute of Biotechnology, Uka Tarsadia University, Bardoli, 394350 - Gujarat, India, Modi K, ASPEE SHAKILAM Agricultural Biotechnology Institute, Navsari Agricultural University, Surat, Gujarat, India. Isolation and characterization of plant growth promoting potential of Acinetobacter sp. RSC7 isolated from Saccharum officinarum cultivar Co 671. J Exp Biol Agric Sci. 2017 Sep 10;5(4):483–91.

7. Beneduzi A, Moreira F, Costa PB, Vargas LK, Lisboa BB, Favreto R, et al. Diversity and plant growth promoting evaluation abilities of bacteria isolated from sugarcane cultivated in the South of Brazil. Appl Soil Ecol. 2013 Jan;63:94–104.

8. Xing YX, Wei CY, Mo Y, Yang LT, Huang SL, Li YR. Nitrogen-Fixing and Plant Growth-Promoting Ability of Two Endophytic Bacterial Strains Isolated from Sugarcane Stalks. Sugar Tech. 2016 Aug;18(4):373–9.

9. Pereira LB, Andrade GS, Meneghin SP, Vicentini R, Ottoboni LMM. Prospecting Plant Growth-Promoting Bacteria Isolated from the Rhizosphere of Sugarcane Under Drought Stress. Curr Microbiol. 2019 Nov;76(11):1345–54.

10. Reinhardt Érica L, Ramos PL, Manfio GP, Barbosa HR, Pavan C, Moreira-Filho CA. Molecular characterization of nitrogen-fixing bacteria isolated from brazilian agricultural plants at São Paulo state. Braz J Microbiol. 2008 Sep;39(3):414–22.

11. de Bellone SC, Bellone CH. Presence of endophytic diazotrophs in sugarcane juice. World J Microbiol Biotechnol. 2006 Oct;22(10):1065–8.

12. Tejera N, Lluch C, Martìnez-Toledo MV, Gonzàlez-López J. Isolation and characterization of Azotobacter and Azospirillum strains from the sugarcane rhizosphere. Plant Soil. 2005 Mar;270(1):223–32.

13. Kennedy I. Non-symbiotic bacterial diazotrophs in crop-farming systems: can their potential for plant growth promotion be better exploited? Soil Biol Biochem. 2004 Aug;36(8):1229–44.

14. PEDRAZA RO, DIAZ RICCI JC. Genetic stability of Azospirillum brasilense after passing through the root interior of sugarcane. Symbiosis. 2003;

15. Boddey RM, de Oliveira OC, Urquiaga S, Reis VM, de Olivares FL, Baldani VLD, et al. Biological nitrogen fixation associated with sugar cane and rice: Contributions and prospects for improvement. Plant Soil. 1995 Jul;174(1–2):195–209.

16. Nascimento FC, Kandasamy S, Lazarovits G, Rigobelo EC. Effect of Chemical Fertilization on the Impacts of Plant Growth-Promoting Rhizobacteria in Maize Crops. Curr Microbiol. 2020 Dec;77(12):3878–87.

17. Rosa PAL, Galindo FS, Oliveira CE da S, Jalal A, Mortinho ES, Fernandes GC, et al. Inoculation with Plant Growth-Promoting Bacteria to Reduce Phosphate Fertilization Requirement and Enhance Technological Quality and Yield of Sugarcane. Microorganisms. 2022 Jan 17;10(1):192.

18. Singh D, Sharma A, Saini GK. Biochemical and molecular characterisation of the bacterial endophytes from native sugarcane varieties of Himalayan region. 3 Biotech. 2013 Jun;3(3):205–12.

19. Chauhan H, Bagyaraj DJ, Sharma A. PLANT GROWTH-PROMOTING BACTERIAL ENDOPHYTES FROM SUGARCANE AND THEIR POTENTIAL IN PROMOTING GROWTH OF THE HOST UNDER FIELD CONDITIONS. Exp Agric. 2013 Jan;49(1):43–52.

20. Pisa G, Magnani GS, Weber H, Souza EM, Faoro H, Monteiro RA, et al. Diversity of 16S rRNA genes from bacteria of sugarcane rhizosphere soil. Braz J Med Biol Res. 2011 Dec;44(12):1215–21.

21. Oliveira ALM, Stoffels M, Schmid M, Reis VM, Baldani JI, Hartmann A. Colonization of sugarcane plantlets by mixed inoculations with diazotrophic bacteria. Eur J Soil Biol. 2009 Jan;45(1):106–13.

22. Rossetto L, Pierangeli GMF, Kuramae EE, Xavier MA, Cipriano MAP, Silveira APD da. Sugarcane pre-sprouted seedlings produced with beneficial bacteria and arbuscular mycorrhizal fungi. Bragantia. 2021;80:e2721.

23. Michavila G, Alibrandi P, Cinà P, Welin B, Castagnaro AP, Chalfoun NR, et al. Plant growth-promoting bacteria isolated from sugarcane improve the survival of micropropagated plants during acclimatisation. Ital J Agron [Internet]. 2022 Jun 29 [cited 2022 Dec 22];17(2). Available from: https://www.agronomy.it/index.php/agro/article/view/2006

24. de Melo LHV, Rocha FYO, Vidal MS, Gitahy P de M, Arruda GM, Barreto CP,

et al. Diversity and biotechnological potential of endophytic Bacillus species originating from the stem apoplast fluid of sugarcane plants. Appl Soil Ecol. 2021 Oct;166:103985.

25. Singh RK, Singh P, Li HB, Song QQ, Guo DJ, Solanki MK, et al. Diversity of nitrogen-fixing rhizobacteria associated with sugarcane: a comprehensive study of plant-microbe interactions for growth enhancement in Saccharum spp. BMC Plant Biol. 2020 Dec;20(1):220.

26. Magnani GS, Didonet CM, Cruz LM, Picheth CF, Pedrosa FO, Souza EM. Diversity of endophytic bacteria in Brazilian sugarcane. Genet Mol Res. 2010;9(1):250–8.

27. de Lima DRM, dos Santos IB, Oliveira JTC, da Costa DP, de Queiroz JVJ, Romagnoli EM, et al. Genetic diversity of N-fixing and plant growth-promoting bacterial community in different sugarcane genotypes, association habitat and phenological phase of the crop. Arch Microbiol. 2021 Apr;203(3):1089–105.

28. Zainab N, Amna, Din BU, Javed MT, Afridi MS, Mukhtar T, et al. Deciphering metal toxicity responses of flax (Linum usitatissimum L.) with exopolysaccharide and ACC-deaminase producing bacteria in industrially contaminated soils. Plant Physiol Biochem. 2020 Jul;152:90–9.

29. Di Y ning, Kui L, Singh P, Liu L feng, Xie L yan, He L lian, et al. Identification and Characterization of Bacillus subtilis B9: A Diazotrophic Plant Growth-Promoting Endophytic Bacterium Isolated from Sugarcane Root. J Plant Growth Regul [Internet]. 2022 May 7 [cited 2022 Dec 22]; Available from: https://link.springer.com/10.1007/s00344-022-10653-x

30. Amna, Xia Y, Farooq MA, Javed MT, Kamran MA, Mukhtar T, et al. Multi-stress tolerant PGPR Bacillus xiamenensis PM14 activating sugarcane (Saccharum officinarum L.) red rot disease resistance. Plant Physiol Biochem. 2020 Jun;151:640–9.

31. Shastri B, Kumar R, Lal RJ. Isolation, Characterization and Identification of Indigenous Endophytic Bacteria Exhibiting PGP and Antifungal Traits from the Internal Tissue of Sugarcane Crop. Sugar Tech. 2020 Aug;22(4):563–73.

32. Dash DM, Osborne JW. Biodegradation of monocrotophos by a plant growth promoting Bacillus aryabhattai (VITNNDJ5) strain in artificially contaminated soil. Int J Environ Sci Technol. 2020 Mar;17(3):1475–90.

33. Wang Z, Yu Z -X., Solanki MK, Yang L -T., Xing Y -X., Dong D -F., et al. Diversity of sugarcane root-associated endophytic Bacillus and their activities in enhancing plant growth. J Appl Microbiol. 2020 Mar;128(3):814–27.

34. Hazarika DJ, Goswami G, Gautom T, Parveen A, Das P, Barooah M, et al. Lipopeptide mediated biocontrol activity of endophytic Bacillus subtilis against fungal phytopathogens. BMC Microbiol. 2019 Dec;19(1):71.

35. Zia MA, Yasmin H, Shair F, Jabeen Z, Mumtaz S, Hayat Z, et al. Glucanolytic

Rhizobacteria Produce Antifungal Metabolites and Elicit ROS Scavenging System in Sugarcane. Sugar Tech. 2019 Apr;21(2):244–55.

36. Antunes JEL, Freitas ADSD, Oliveira LMS, Lyra MDCCPD, Fonseca MAC, Santos CERS, et al. Sugarcane inoculated with endophytic diazotrophic bacteria: effects on yield, biological nitrogen fixation and industrial characteristics. An Acad Bras Ciênc. 2019;91(4):e20180990.

37. Fonseca M de C da, Bossolani JW, de Oliveira SL, Moretti LG, Portugal JR, Scudeletti D, et al. Bacillus subtilis Inoculation Improves Nutrient Uptake and Physiological Activity in Sugarcane under Drought Stress. Microorganisms. 2022 Apr 13;10(4):809.

38. Pirhadi M. IMPACT OF SOIL SALINITY ON DIVERSITY AND COMMUNITY OF SUGARCANE ENDOPHYTIC PLANT GROWTH PROMOTING BACTERIA (SACCHARUM OFFICINARUM L. VAR. CP48). Appl Ecol Environ Res. 2018;16(1):725–39.

39. Antunes JEL, Freitas ADS de, Bonifacio A, Oliveira LM de S, Santos CE de R e S, Lyra M do CCP de, et al. Isolation and Characterization of Plant Growth-Promotion Diazotrophic Endophytic Bacteria Associated to Sugarcane (Saccharum officinarum L.) Grown in Paraíba, Brazil. Braz Arch Biol Technol. 2022;65:e22200439.

40. Kruasuwan W, Thamchaipenet A. Diversity of Culturable Plant Growth-Promoting Bacterial Endophytes Associated with Sugarcane Roots and Their Effect of Growth by Co-Inoculation of Diazotrophs and Actinomycetes. J Plant Growth Regul. 2016 Dec;35(4):1074–87.

41. Rocha FYO, Negrisoli Júnior AS, de Matos GF, Gitahy P de M, Rossi CN, Vidal MS, et al. Endophytic Bacillus Bacteria Living in Sugarcane Plant Tissues and Telchin licus licus Larvae (Drury) (Lepidoptera: Castniidae): The Symbiosis That May Open New Paths in the Biological Control. Front Microbiol. 2021 May 12;12:659965.

42. de los Milagros Orberá Ratón T, Yano R, Rodríguez Gámez O, Floh EIS, de Jesús Serrat Díaz M, Barbosa HR. Isolation and characterisation of aerobic endospore forming Bacilli from sugarcane rhizosphere for the selection of strains with agriculture potentialities. World J Microbiol Biotechnol. 2012 Apr;28(4):1593–603.

43. Hassan MN, Afghan S, Hafeez FY. Suppression of red rot caused by Colletotrichum falcatum on sugarcane plants using plant growth-promoting rhizobacteria. BioControl. 2010 Aug;55(4):531–42.

44. de Oliveira ZM, Floh EIS, Ferrara FIS, Barbosa HR. Diazotrophyc rhizobacteria isolated from sugarcane can release amino acids in a synthetic culture medium. Biol Fertil Soils. 2011 Nov;47(8):957–62.

45. Ando S, Goto M, Meunchang S, Thongra-ar P, Fujiwara T, Hayashi H, et al. Detection of nifH Sequences in Sugarcane (Saccharum officinarum L.) and Pineapple (Ananas comosus [L.] Merr.). Soil Sci Plant Nutr. 2005 Apr;51(2):303–8.

46. de Alencar Menezes Júnior I, Feitosa de Matos G, Moura de Freitas K, da

Conceição Jesus E, Rouws LFM. Occurrence of diverse Bradyrhizobium spp. in roots and rhizospheres of two commercial Brazilian sugarcane cultivars. Braz J Microbiol. 2019 Jul;50(3):759–67.

47. de Matos GF, Zilli JE, de Araújo JLS, Parma MM, Melo IS, Radl V, et al. Bradyrhizobium sacchari sp. nov., a legume nodulating bacterium isolated from sugarcane roots. Arch Microbiol. 2017 Nov;199(9):1251–8.

48. Rouws LFM, Leite J, de Matos GF, Zilli JE, Coelho MRR, Xavier GR, et al. Endophytic B radyrhizobium spp. isolates from sugarcane obtained through different culture strategies: Bradyrhizobium spp. isolates from sugarcane. Environ Microbiol Rep. 2014 Aug;6(4):354–63.

49. Fischer D, Pfitzner B, Schmid M, Simões-Araújo JL, Reis VM, Pereira W, et al. Molecular characterisation of the diazotrophic bacterial community in uninoculated and inoculated field-grown sugarcane (Saccharum sp.). Plant Soil. 2012 Jul;356(1–2):83–99.

50. Govindarajan M, Kwon SW, Weon HY. Isolation, molecular characterization and growth-promoting activities of endophytic sugarcane diazotroph Klebsiella sp. GR9. World J Microbiol Biotechnol. 2007 Jul;23(7):997–1006.

51. Govindarajan M, Balandreau J, Muthukumarasamy R, Revathi G, Lakshminarasimhan C. Improved Yield of Micropropagated Sugarcane Following Inoculation by Endophytic Burkholderia vietnamiensis. Plant Soil. 2006 Feb;280(1–2):239–52.

52. Caballero-Mellado J, Martínez-Aguilar L, Paredes-Valdez G, Santos PE de los. Burkholderia unamae sp. nov., an N2-fixing rhizospheric and endophytic species. Int J Syst Evol Microbiol. 2004 Jul 1;54(4):1165–72.

53. Boddey RM, Urquiaga S, Alves BJR, Reis V. Endophytic nitrogen fixation in sugarcane: present knowledge and future applications. Plant Soil. 2003 May;252(1):139–49.

54. Cui G, Yin K, Lin N, Liang M, Huang C, Chang C, et al. Burkholderia gladioli CGB10: A Novel Strain Biocontrolling the Sugarcane Smut Disease. Microorganisms. 2020 Dec 7;8(12):1943.

55. Luvizotto DM, Marcon J, Andreote FD, Dini-Andreote F, Neves AAC, Araújo WL, et al. Genetic diversity and plant-growth related features of Burkholderia spp. from sugarcane roots. World J Microbiol Biotechnol. 2010 Oct;26(10):1829–36.

56. Labanca ERG, Andrade SAL, Kuramae EE, Silveira APD. The modulation of sugarcane growth and nutritional profile under aluminum stress is dependent on beneficial endophytic bacteria and plantlet origin. Appl Soil Ecol. 2020 Dec;156:103715.

57. Malviya MK, Li CN, Solanki MK, Singh RK, Htun R, Singh P, et al. Comparative analysis of sugarcane root transcriptome in response to the plant growth-promoting Burkholderia anthina MYSP113. Singh AK, editor. PLOS ONE. 2020 Apr 8;15(4):e0231206.

58. Malviya MK, Solanki MK, Li CN, Htun R, Singh RK, Singh P, et al. Beneficial Linkages of Endophytic Burkholderia anthina MYSP113 Towards Sugarcane Growth Promotion. Sugar Tech. 2019 Oct;21(5):737–48.

59. Gonçalves PJR de O, Hume CCD, Ferreira AJ, Tsui S, Brocchi M, Wren BW, et al. Environmental interactions are regulated by temperature in Burkholderia seminalis TC3.4.2R3. Sci Rep. 2019 Apr 2;9(1):5486.

60. Rezamahalleh HM, Khodakaramian G, Hassanzadeh N. Diversity of Endophytic and Epiphytic Bacteria From Sugarcane in Khuzestan, Iran. Braz Arch Biol Technol. 2019;62:e19180407.

61. dos Santos SG, Chaves VA, da Silva Ribeiro F, Alves GC, Reis VM. Rooting and growth of pre-germinated sugarcane seedlings inoculated with diazotrophic bacteria. Appl Soil Ecol. 2019 Jan;133:12–23.

62. Rodrigues AA, Araújo MVF, Soares RS, Oliveira BFRD, Ribeiro IDA, Sibov ST, et al. Isolation and prospection of diazotrophic rhizobacteria associated with sugarcane under organic management. An Acad Bras Ciênc. 2018 Dec;90(4):3813–29.

63. Lima DRM de, Santos IB dos, Oliveira JTC, Barbosa JG, Diniz WP da S, Farias ARB de, et al. Tolerance of potentially diazotrophic bacteria to adverse environmental conditions and plant growth-promotion in sugarcane. Arch Agron Soil Sci. 2018 Sep 19;64(11):1534–48.

64. Department of Agricultural Microbiology, Tamil Nadu Agricultural University, Coimbatore-03, Arthee R, Marimuthu P, Department of Agricultural Microbiology, Tamil Nadu Agricultural University, Coimbatore-03. Studies on endophytic Burkholderia sp. from sugarcane and its screening for plant growth promoting potential. J Exp Biol Agric Sci. 2017 May 20;5(2):242–57.

65. Muthukumarasamy R, Revathi G, Vadivelu M, Arun K. Isolation of bacterial strains possessing nitrogen-fixation, phosphate and potassium-solubilization and their inoculation effects on sugarcane. Indian J Exp Biol. 2017 Mar;55(3):161–70.

66. Paungfoo-Lonhienne C, Lonhienne TGA, Yeoh YK, Donose BC, Webb RI, Parsons J, et al. Crosstalk between sugarcane and a plant-growth promoting Burkholderia species. Sci Rep. 2016 Dec;6(1):37389.

67. Pereira A de A, Silva M de B, Oliveira J de S, Ramos A de S, Freire M, Freire FJ, et al. Salinity influence on the growth and production of indole acetic acid by endophytic Burkholderia spp. from sugarcane. Biosci J. 2012;28(1, Suppl. 1):112–21.

68. Castro-González R, Martínez-Aguilar L, Ramírez-Trujillo A, Estrada-de los Santos P, Caballero-Mellado J. High diversity of culturable Burkholderia species associated with sugarcane. Plant Soil. 2011 Aug;345(1–2):155–69.

69. Rothballer M, Schmid M, Hartmann A. Diazotrophic Bacterial Endophytes in Gramineae and Other Plants. In: Pawlowski K, editor. Prokaryotic Symbionts in Plants [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2007 [cited 2022 Dec 22]. p. 273–302. (Microbiology Monographs; vol. 8). Available from: http://link.springer.com/10.1007/7171_2007_103

70. Bhise KK, Bhagwat PK, Dandge PB. Synergistic effect of Chryseobacterium gleum sp. SUK with ACC deaminase activity in alleviation of salt stress and plant growth promotion in Triticum aestivum L. 3 Biotech. 2017 Jun;7(2):105.

71. Herlet J, Schwarz WH, Zverlov VV, Liebl W, Kornberger P. Addition of β-galactosidase boosts the xyloglucan degradation capability of endoglucanase Cel9D from Clostridium thermocellum. Biotechnol Biofuels. 2018 Dec;11(1):238.

72. da Silveira APD, Iório R de PF, Marcos FCC, Fernandes AO, de Souza SACD, Kuramae EE, et al. Exploitation of new endophytic bacteria and their ability to promote sugarcane growth and nitrogen nutrition. Antonie Van Leeuwenhoek. 2019 Feb;112(2):283–95.

73. Singh P, Singh RK, Li HB, Guo DJ, Sharma A, Lakshmanan P, et al. Diazotrophic Bacteria Pantoea dispersa and Enterobacter asburiae Promote Sugarcane Growth by Inducing Nitrogen Uptake and Defense-Related Gene Expression. Front Microbiol. 2021 Jan 12;11:600417.

74. Asis CA, Kubota M, Ohta H, Arima Y, Chebotar VK, Tsuchiya K ichi, et al. Isolation and partial characterization of endophytic diazotrophs associated with Japanese sugarcane cultivar. Soil Sci Plant Nutr. 2000 Sep;46(3):759–65.

75. Mehnaz S. Genetic and Phenotypic Diversity of Plant Growth Promoting Rhizobacteria Isolated from Sugarcane Plants Growing in Pakistan. J Microbiol Biotechnol. 2010 Dec;1614–23.

76. Guo DJ, Singh RK, Singh P, Li DP, Sharma A, Xing YX, et al. Complete Genome Sequence of Enterobacter roggenkampii ED5, a Nitrogen Fixing Plant Growth Promoting Endophytic Bacterium With Biocontrol and Stress Tolerance Properties, Isolated From Sugarcane Root. Front Microbiol. 2020 Sep 22;11:580081.

77. Paredes-Villanueva J, Del Rosario JL, Urcia-Pulido M, Zavaleta-Armas J. Plant growth promoter collection of Gluconacetobacter diazotrophicus from northern coast of Peru. Sci Agropecu. 2020 Mar 31;11(1):15–21.

78. Taulé C, Luizzi H, Beracochea M, Mareque C, Platero R, Battistoni F. The Mo- and Fe-nitrogenases of the endophyte Kosakonia sp. UYSO10 are necessary for growth promotion of sugarcane. Ann Microbiol. 2019 Jul;69(7):741–50.

79. Kruasuwan W, Thamchaipenet A. 1-Aminocyclopropane-1-carboxylate (ACC) Deaminase-Producing Endophytic Diazotrophic Enterobacter sp. EN-21 Modulates Salt–Stress Response in Sugarcane. J Plant Growth Regul. 2018 Sep;37(3):849–58.

80. Taulé C, Castillo A, Villar S, Olivares F, Battistoni F. Endophytic colonization of sugarcane (Saccharum officinarum) by the novel diazotrophs Shinella sp. UYSO24 and Enterobacter sp. UYSO10. Plant Soil. 2016 Jun;403(1–2):403–18.

81. Meesilp N, Jaisil P, Milintawisamai N, Niamsanit S. Effects of a Formulated Bio-product Containing the Nitrogen Fixing Bacterial Strain Enterobacter oryzae 3LSO1 on Sugarcane Growth. Sugar Tech. 2016 Jun;18(3):242–8.

82. Lin L, Li Z, Hu C, Zhang X, Chang S, Yang L, et al. Plant Growth-Promoting Nitrogen-Fixing Enterobacteria Are in Association with Sugarcane Plants Growing in

Guangxi, China. Microbes Environ. 2012;27(4):391–8.

83. Zhu B, Chen M, Lin L, Yang L, Li Y, An Q. Genome Sequence of Enterobacter sp. Strain SP1, an Endophytic Nitrogen-Fixing Bacterium Isolated from Sugarcane. J Bacteriol. 2012 Dec 15;194(24):6963–4.

84. Pedraza RO. Recent advances in nitrogen-fixing acetic acid bacteria. Int J Food Microbiol. 2008 Jun;125(1):25–35.

85. Lery LMS, Coelho A, von Kruger WMA, Gonçalves MSM, Santos MF, Valente RH, et al. Protein expression profile of Gluconacetobacter diazotrophicus PAL5, a sugarcane endophytic plant growth-promoting bacterium. PROTEOMICS. 2008 Nov;8(22):4833–4833.

86. Cocking EC, Stone PJ, Davey MR. Intracellular colonization of roots of Arabidopsis and crop plants by Gluconacetobacter diazotrophicus. Vitro Cell Dev Biol - Plant. 2006 Jan;42(1):74–82.

87. Lee S, Flores-Encarnación M, Contreras-Zentella M, Garcia-Flores L, Escamilla JE, Kennedy C. Indole-3-Acetic Acid Biosynthesis Is Deficient in Gluconacetobacter diazotrophicus Strains with Mutations in Cytochrome c Biogenesis Genes. J Bacteriol. 2004 Aug 15;186(16):5384–91.

88. Perin L, Baldani JI, Reis VM. Diversidade de Gluconacetobacter diazotrophicus isolada de plantas de cana-de-açúcar cultivadas no Brasil. Pesqui Agropecuária Bras. 2004 Aug;39(8):763–70.

89. Cocking EC. Endophytic colonization of plant roots by nitrogen-fixing bacteria. Plant Soil. 2003 May;252(1):169–75.

90. VARGAS C, MUNIZ DE PADUA VL, DE MATOS NOGUEIRA E, VINAGRE F, MASUDA HP, RODRIGUES DA SILVA F, et al. Signaling pathways mediating the association between sugarcane and endophytic diazotrophic bacteria: a genomic approach. Symbiosis. 2003;

91. Dong Z, Zelmer CD, Canny MJ, McCully ME, Luit B, Pan B, et al. Evidence for protection of nitrogenase from O2 by colony structure in the aerobic diazotroph Gluconacetobacter diazotrophicus. Microbiology. 2002 Aug 1;148(8):2293–8.

92. Pan B, Vessey JK. Response of the Endophytic Diazotroph Gluconacetobacter diazotrophicus on Solid Media to Changes in Atmospheric Partial O 2 Pressure. Appl Environ Microbiol. 2001 Oct;67(10):4694–700.

93. Aguiar NO, Olivares FL, Novotny EH, Canellas LP. Changes in metabolic profiling of sugarcane leaves induced by endophytic diazotrophic bacteria and humic acids. PeerJ. 2018 Sep 5;6:e5445.

94. dos Santos SG, da Silva Ribeiro F, da Fonseca CS, Pereira W, Santos LA, Reis VM. Development and nitrate reductase activity of sugarcane inoculated with five diazotrophic strains. Arch Microbiol. 2017 Aug;199(6):863–73.

95. Yoon V, Tian G, Vessey JK, Macfie SM, Dangi OP, Kumer AK, et al. Colonization efficiency of different sorghum genotypes by Gluconacetobacter diazotrophicus. Plant Soil. 2016 Jan;398(1–2):243–56.

96. Alquéres SM, Cardoso AM, Brito-Moreira J, Baldani JI, Martins OB. Transfer RNA-dependent asparagine biosynthesis in Gluconacetobacter diazotrophicus and its influence on biological nitrogen fixation. Plant Soil. 2012 Jul;356(1–2):209–16.

97. Nieto-Peñalver CG, Bertini EV, de Figueroa LIC. Identification of N-acyl homoserine lactones produced by Gluconacetobacter diazotrophicus PAL5 cultured in complex and synthetic media. Arch Microbiol. 2012 Jul;194(7):615–22.

98. Natheer SE, Muthukkaruppan S. Assessing the in vitro zinc solubilization potential and improving sugarcane growth by inoculating Gluconacetobacter diazotrophicus. Ann Microbiol. 2012 Mar;62(1):435–41.

99. De La Cruz CPP, Bird CO, Isulat MD, others. Sprouting, survival and growth of young sugarcane (Saccharum officinarum L.) treated with diazotrophic bacteria (Gluconacetobacter diazotrophicus). Philipp Agric Sci. 2012;95(1):106–11.

100. Lery LMS, Hemerly AS, Nogueira EM, von Krüger WMA, Bisch PM. Quantitative Proteomic Analysis of the Interaction Between the Endophytic Plant-Growth-Promoting Bacterium Gluconacetobacter diazotrophicus and Sugarcane. Mol Plant-Microbe Interactions®. 2011 May;24(5):562–76.

101. Silva-Froufe LG da, Boddey RM, Reis VM. Quantification of natural populations of Gluconacetobacter diazotrophicus and Herbaspirillum spp. In sugar cane (Saccharum spp.) Using differente polyclonal antibodies. Braz J Microbiol [Internet]. 2009 Dec [cited 2022 Dec 22];40(4). Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1517-83822009000400018&lng=en&nrm=iso&tlng=en

102. Bertalan M, Albano R, de Pádua V, Rouws L, Rojas C, Hemerly A, et al. Complete genome sequence of the sugarcane nitrogen-fixing endophyte Gluconacetobacter diazotrophicus Pal5. BMC Genomics. 2009 Dec;10(1):450.

103. Intorne AC, de Oliveira MVV, Lima ML, da Silva JF, Olivares FL, de Souza Filho GA. Identification and characterization of Gluconacetobacter diazotrophicus mutants defective in the solubilization of phosphorus and zinc. Arch Microbiol. 2009 May;191(5):477–83.

104. Marques Júnior RB, Canellas LP, Silva LG da, Olivares FL. Rooting of micro seed pieces by combined use of humic substances and endophytic diazotrophic bacteria in sugar cane. Rev Bras Cienc Solo. 2008;32:1121–8.

105. Downing KJ, Leslie G, Thomson JA. Biocontrol of the Sugarcane Borer Eldana saccharina by Expression of the Bacillus thuringiensis cry1Ac7 and Serratia marcescens chiA Genes in Sugarcane-Associated Bacteria. Appl Environ Microbiol. 2000 Jul;66(7):2804–10.

106. Olivares FL, Baldani VLD, Reis VM, Baldani JI, Döbereiner J. Occurrence of the endophytic diazotrophs Herbaspirillum spp. in roots, stems, and leaves, predominantly of Gramineae. Biol Fertil Soils. 1996 Feb;21(3):197–200.

107. Leite MFA, Dimitrov MR, Freitas-Iório RP, de Hollander M, Cipriano MAP, Andrade SAL, et al. Rearranging the sugarcane holobiont via plant growth-promoting bacteria and nitrogen input. Sci Total Environ. 2021 Dec;800:149493.

108. Sajjad Mirza M, Ahmad W, Latif F, Haurat J, Bally R, Normand P, et al. Isolation, partial characterization, and the effect of plant growth-promoting bacteria (PGPB) on micro-propagated sugarcane in vitro. Plant Soil. 2001;237(1):47–54.

109. Dantur KI, Chalfoun NR, Claps MP, Tórtora ML, Silva C, Jure Á, et al. The Endophytic Strain Klebsiella michiganensis Kd70 Lacks Pathogenic Island-Like Regions in Its Genome and Is Incapable of Infecting the Urinary Tract in Mice. Front Microbiol. 2018 Jul 16;9:1548.

110. Lin L, Wei C, Chen M, Wang H, Li Y, Li Y, et al. Complete genome sequence of endophytic nitrogen-fixing Klebsiella variicola strain DX120E. Stand Genomic Sci. 2015 Dec;10(1):22.

111. Wei CY, Lin L, Luo LJ, Xing YX, Hu CJ, Yang LT, et al. Endophytic nitrogen-fixing Klebsiella variicola strain DX120E promotes sugarcane growth. Biol Fertil Soils. 2014 May;50(4):657–66.

112. Singh RK, Singh P, Li HB, Guo DJ, Song QQ, Yang LT, et al. Plant-PGPR interaction study of plant growth-promoting diazotrophs Kosakonia radicincitans BA1 and Stenotrophomonas maltophilia COA2 to enhance growth and stress-related gene expression in Saccharum spp. J Plant Interact. 2020 Jan 1;15(1):427–45.

113. Beracochea M, Taulé C, Battistoni F. Draft Genome Sequence of Kosakonia radicincitans UYSO10, an Endophytic Plant Growth-Promoting Bacterium of Sugarcane (Saccharum officinarum). Dennehy JJ, editor. Microbiol Resour Announc. 2019 Oct 24;8(43):e01000-19.

114. Chen M, Zhu B, Lin L, Yang L, Li Y, An Q. Complete genome sequence of Kosakonia sacchari type strain SP1T. Stand Genomic Sci. 2014 Jun 6;9(3):1311–8.

115. Yang T, Shi F, An Y, Sun D, Lu Y. Isolation of a Pink-pigmented Facultative Methylotrophic Bacteria and its Growth-promoting Effect on Pepper Plants. Curr Sci. 2022 Jul 25;123(2):214.

116. Madhaiyan M, Poonguzhali S, Sundaram SP, Sa T. A new insight into foliar applied methanol influencing phylloplane methylotrophic dynamics and growth promotion of cotton (Gossypium hirsutum L.) and sugarcane (Saccharum officinarum L.). Environ Exp Bot. 2006 Aug;57(1–2):168–76.

117. Madhaiyan M, Poonguzhali S, Lee HS, Hari K, Sundaram SP, Sa TM. Pink-pigmented facultative methylotrophic bacteria accelerate germination, growth and yield of sugarcane clone Co86032 (Saccharum officinarum L.). Biol Fertil Soils. 2005 Jul;41(5):350–8.

118. Meena KK, Kumar M, Kalyuzhnaya MG, Yandigeri MS, Singh DP, Saxena AK, et al. Epiphytic pink-pigmented methylotrophic bacteria enhance germination and seedling growth of wheat (Triticum aestivum) by producing phytohormone. Antonie Van Leeuwenhoek. 2012 May;101(4):777–86.

119. Mendes R, Pizzirani-Kleiner AA, Araujo WL, Raaijmakers JM. Diversity of Cultivated Endophytic Bacteria from Sugarcane: Genetic and Biochemical Characterization of Burkholderia cepacia Complex Isolates. Appl Environ Microbiol. 2007 Nov 15;73(22):7259–67.

120. Lin L, Guo W, Xing Y, Zhang X, Li Z, Hu C, et al. The actinobacterium Microbacterium sp. 16SH accepts pBBR1-based pPROBE vectors, forms biofilms, invades roots, and fixes N2 associated with micropropagated sugarcane plants. Appl Microbiol Biotechnol. 2012 Feb;93(3):1185–95.

121. Kruasuwan W, Hoskisson PA, Thamchaipenet A. Draft Genome Sequence of Root-Associated Sugarcane Growth-Promoting Microbispora sp. Strain GKU 823. Genome Announc. 2017 Jul 20;5(29):e00647-17.

122. Terra LA, de Soares CP, Meneses CHSG, Tadra Sfeir MZ, de Souza EM, Silveira V, et al. Transcriptome and proteome profiles of the diazotroph Nitrospirillum amazonense strain CBAmC in response to the sugarcane apoplast fluid. Plant Soil. 2020 Jun;451(1–2):145–68.

123. Schwab S, Terra LA, Baldani JI. Genomic characterization of Nitrospirillum amazonense strain CBAmC, a nitrogen-fixing bacterium isolated from surface-sterilized sugarcane stems. Mol Genet Genomics. 2018 Aug;293(4):997–1016.

124. Loiret FG, Ortega E, Kleiner D, Ortega-Rodes P, Rodes R, Dong Z. A putative new endophytic nitrogen-fixing bacterium Pantoea sp. from sugarcane. J Appl Microbiol. 2004 Sep;97(3):504–11.

125. Zeng Q, Shi G, Nong Z, Ye X, Hu C. Complete Genome Sequence of Pantoea ananatis Strain NN08200, an Endophytic Bacterium Isolated from Sugarcane. Curr Microbiol. 2020 Aug;77(8):1864–70.

126. Kamran S, Shahid I, Baig DN, Rizwan M, Malik KA, Mehnaz S. Contribution of Zinc Solubilizing Bacteria in Growth Promotion and Zinc Content of Wheat. Front Microbiol. 2017 Dec 21;8:2593.

127. da Silva JF, Barbosa RR, de Souza AN, da Motta OV, Teixeira GN, Carvalho VS, et al. Isolation of Pantoea ananatis from sugarcane and characterization of its potential for plant growth promotion. Genet Mol Res. 2015;14(4):15301–11.

128. Loiret FG, Grimm B, Hajirezaei MR, Kleiner D, Ortega E. Inoculation of sugarcane with Pantoea sp. increases amino acid contents in shoot tissues; serine, alanine, glutamine and asparagine permit concomitantly ammonium excretion and nitrogenase activity of the bacterium. J Plant Physiol. 2009 Jul;166(11):1152–61.

129. da Silva PRA, Vidal MS, Soares C de P, Polese V, Tadra-Sfeir MZ, Souza EM de, et al. Sugarcane apoplast fluid modulates the global transcriptional profile of the diazotrophic bacteria Paraburkholderia tropica strain Ppe8. Aroca R, editor. PLOS ONE. 2018 Dec 14;13(12):e0207863.

130. Kumar NR, Arasu VT, Gunasekaran P. Genotyping of antifungal compounds producing plant growth-promoting rhizobacteria, Pseudomonas fluorescens. Curr Sci. 2002;1463–6.

131. Lopes LD, Weisberg AJ, Davis EW, Varize C de S, Pereira e Silva M de C, Chang JH, et al. Genomic and metabolic differences between Pseudomonas putida populations inhabiting sugarcane rhizosphere or bulk soil. Puglisi E, editor. PLOS ONE. 2019 Oct 3;14(10):e0223269.

132. Li HB, Singh RK, Singh P, Song QQ, Xing YX, Yang LT, et al. Genetic Diversity of Nitrogen-Fixing and Plant Growth Promoting Pseudomonas Species Isolated from Sugarcane Rhizosphere. Front Microbiol. 2017 Jul 14;8:1268.

133. Rameshkumar N. Genotypic and Phenotypic Diversity of PGPR Fluorescent Pseudomonads Isolated from the Rhizosphere of Sugarcane (Saccharum officinarum L.). J Microbiol Biotechnol. 2012 Jan 28;22(1):13–24.

134. Singh P, Singh RK, Guo DJ, Sharma A, Singh RN, Li DP, et al. Whole Genome Analysis of Sugarcane Root-Associated Endophyte Pseudomonas aeruginosa B18—A Plant Growth-Promoting Bacterium With Antagonistic Potential Against Sporisorium scitamineum. Front Microbiol. 2021 Feb 5;12:628376.

135. Mehnaz S, Weselowski B, Aftab F, Zahid S, Lazarovits G, Iqbal J. Isolation, characterization, and effect of fluorescent pseudomonads on micropropagated sugarcane. Can J Microbiol. 2009;55(8):1007–11.

136. Villegas LC, Paterno ES, others. Growth enhancement and root colonization of sugarcane by plant growth-promoting bacteria. Philipp J Crop Sci. 2008;33(2):3–24.

137. Luo T, Ou-Yang XQ, Yang LT, Li YR, Song XP, Zhang GM, et al. Raoultella sp. strain L03 fixes N 2 in association with micropropagated sugarcane plants: Raoultella -sugarcane associated N 2 -fixation. J Basic Microbiol. 2016 Aug;56(8):934–40.

138. Dos-Santos CM, Ribeiro NVS, Schwab S, Baldani JI, Vidal MS. The Effect of Inoculation of a Diazotrophic Bacterial Consortium on the Indigenous Bacterial Community Structure of Sugarcane Apoplast Fluid. Curr Microbiol. 2021 Aug;78(8):3079–91.

139. de Lima AMS, de Oliveira ECA, Martins VRS, da Silva LB, de Souza PHN, Freire FJ. Integrated Application of Nitrogen, Molybdenum and Plant Growth-Promoting Rhizobacterium can Enhance the Sugarcane Growth. Sugar Tech. 2022 Dec;24(6):1748–65.

140. Kruasuwan W, Salih TS, Brozio S, Hoskisson PA, Thamchaipenet A. Draft Genome Sequence of Plant Growth-Promoting Endophytic Streptomyces sp. GKU 895 Isolated from the Roots of Sugarcane. Genome Announc. 2017 May 11;5(19):e00358-17.

CHAPTER 2 – Unveiling genomic features linked to traits of plant-growth-promoting bacterial communities from sugarcane

ABSTRACT - The microorganisms are ubiquitous, and those inhabiting plants have been the subject of several studies. Plant-associated bacteria exhibit various biological mechanisms that enable them to colonize host plants and, in some cases, enhance their fitness. In this study, we describe the genomic features predicted to be associated with plant growth-promoting traits in six bacterial communities isolated from sugarcane. The use of highly accurate single-molecule real-time sequencing technology for metagenomic samples from these bacterial communities allowed us to recover 17 genomes. The taxonomic assignments for the binned genomes were performed, revealing taxa distributed across three main phyla: Bacillota, Bacteroidota, and Pseudomonadota, with the latter being the most representative. Subsequently, we functionally annotated the metagenome-assembled genomes (MAGs) to characterize their metabolic pathways related to plant growth-promoting traits. Our study successfully identified the enrichment of important functions related to phosphate and potassium acquisition, modulation of phytohormones, and mechanisms for coping with abiotic stress. These findings could be linked to the robust colonization of these sugarcane endophytes.

Keywords: long-read sequencing; Single-Molecule Real Time (SMRT) sequencing; plant–microbe association; plant-beneficial bacteria;

1. INTRODUCTION

Plants harbor numerous microorganisms that engage in complex interactions with both the plant host and with each other. The plant microbiome has been recognized as one of the main determinants of plant growth, health, and productivity (Berg et al. 2017). The combination of the host plant and its symbiotic microbiota is referred to as a holobiont (Sánchez-Cañizares et al. 2017, Hassani et al. 2018).

Beneficial microorganisms are becoming more prominent in modern agriculture, given their role in increasing the productivity of different crops, helping to maintain soil fertility, and reducing pollutants derived from the intensive use of agrochemical inputs (Kumar et al. 2022). The study of microbiomes associated with soil and plants has been highlighted to expand plant resilience and crop productivity in an agricultural system (Armanhi et al. 2021). Several studies demonstrate considerable evidence of the role of beneficial microorganisms in increasing plant growth (Marks et al. 2015), nutrient acquisition (Emami et al. 2018, Singh et al. 2021), biocontrol (Rojas-Solís et al. 2018), and resistance induction (Sharma et al. 2018). However, our understanding of the relationships inherent to the microbiota associated with plants and soil is still limited., This has led to studies aiming to identify the driving factors that modulate the composition and structure of microbiome communities (Dastogeer et al. 2020, Santos and Olivares 2021).

The phytomicrobiomes of different plants such as sugarcane (de Souza et al. 2016, Hamonts et al. 2018, de Carvalho et al. 2021), maize (Cai et al. 2018), rice (Cai et al. 2018, Wang et al. 2020), and others, have been explored to understand their structures, interactions, and roles in promoting benefits to hosts, consequently improving crop quality and productivity in agriculture (Parray et al. 2022). Despite the knowledge surrounding the structure and dynamics of plant-associated microbiomes, it is still a challenge to understand which species and traits are actively part of the community or are playing key roles (Großkopf and Soyer 2014). To overcome this limitation, researchers have focused on assembling simplified microbial consortia (SMC) or synthetic communities (SynCom) (Großkopf and Soyer 2014, Kang et al. 2020). Furthermore, a new approach known as simple state communities (SSC) has been proposed to simplify the complex communities and enable studies aimed at understanding microbial interactions (Sarkar et al. 2022).

Bacterial consortia have gained interest as potential inoculants since different types of microorganisms can interact synergistically. This can not only increase functional diversity but also enhance metabolic efficiency and environmental adaptation (Tosi et al. 2020). Several studies related to the co-inoculation of different species of microorganisms have observed an improvement in yield (Hungria et al. 2013). Additionally, these studies have presented various mechanisms of action resulting from the compatibility and synergy between the species (Korir et al. 2017, Armanhi et al. 2021, Kaur et al. 2022).

Recently, the study by Teheran-Sierra et al., (2021) reported microbial communities from sugarcane with high potential to promote plant growth. This report was based on several assays commonly related to plant growth-promoting microbes (PGPM) and also observed beneficial traits in plants (*Brachiaria ruziziensis*), besides other characteristics related to indirect traits that could provide benefits to the microorganisms' adaptability in the plant environment. However, there is still a lack of information about the role and function of the bacteria in these communities that possess multiple plant growth-promoting traits associated with sugarcane.

Strategies using metagenome-assembled genome (MAG) have proven helpful in understanding microbial populations and their functional potential within dynamic ecosystems. Moreover, this methodology enables the discovery of novel species (Setubal 2021). Many studies have generated valuable reference genomes for the different microbiomes using MAGs (IMG/M Data Consortium et al. 2020, Almeida et al. 2021, Xie et al. 2021, Su et al. 2022). However, the quality of these genomes can be precarious, and the most recent strategy in metagenomics to overcome this consists of employing long reads, to ensure greater contiguity of MAGs, together with short reads, to polish and improve overall precision (Driscoll et al. 2017, Cuscó et al. 2021, Singleton et al. 2021) or use high accuracy methods, like highly accurate long-read (HiFi sequencing) (Haro-Moreno et al. 2021).

The current study aims to explore the genomes of the community's species, identifying genomic features and linked traits related to such plant-beneficial capabilities. This leads us to question whether understanding genomic characterization in terms of functions can reveal that microbial communities collectively contribute to a greater diversity of functions that are beneficial to plants.

To achieve this, we utilized highly accurate single-molecule real-time sequencing technology (SMRT) on metagenomic samples from the top-rated bacterial communities by Teheran-Sierra et al., (2021). We retrieved MAGs and accessed their functional potential to identify biological mechanisms related to the plant growth-promoting traits of each member.

2. MATERIAL AND METHODS

2.1. Selected bacterial communities

The bacterial communities we are focusing on were obtained from the phytomicrobiome of mature sugarcane (*Saccharum* sp.) variety CTC9001 under distinct agricultural management, as previously described by Teheran-Sierra et al., (2021). The top 6 communities were selected based on overall plant growth promotion performance. The communities C07, C09, and C12 were obtained from the epiphytic surface of the stalk, while C08 and C28 are composed of endophytes. All of them from sugarcane plants cultivated under conventional management systems. The C27 community, on the other hand, was obtained from the endophytic niche of sugarcane leaves under organic management. These communities were selected based on features related to benefits to plants, such as 1-aminocyclopropane-1-carboxylic acid (ACC) deaminase and indole-3-acetic acid (IAA) production, solubilization of phosphorus, and others (Teheran-Sierra et al. 2021). They were also evaluated *in planta* using *Brachiaria ruziziensis,* besides other characteristics related to indirect traits that could provide benefits to the microorganisms' adaptability in the plant environment. In addition, the same study revealed the composition of these communities by a metabarcoding approach based on the sequencing of the V3-V4 hypervariable region of the 16S rRNA gene.

2.2. DNA extraction

The bacterial communities were individually cultured in 20 mL of BRX broth (Teheran-Sierra et al. 2021) from a 10 µL aliquot of the stock. They were incubated for 48 h at 30 °C under constant agitation at 150 rpm. Subsequently, the community cultures were submitted to centrifugation of 3000 x g for 15 min at 20 ºC. The resulting pellets were resuspended in 1X PBS buffer, and the cell concentration was

adjusted to approximately 2 x 10^9 cells. The bacterial suspension was then centrifuged, and the pellets underwent the DNA extraction process. DNA extractions were performed using the MagAttract HMW DNA Kit (QIAGEN, Hilden, Germany) (Cat No./ID: 67563) according to the manufacturer's instructions. Then, the DNA was quantified using NanoDrop 2000® spectrophotometer and Qubit® 3.0 fluorometer (Thermo Fisher Scientific©). The concentrations were adjusted to 50 ng.μL^{-1} to ensure equal sampling conditions for sequencing. The DNA samples were then pooled into two groups, SC1 (composed of C08, C09, and C28 communities) and SC2 (composed of C07, C12, and C27 communities). The DNA pools were quantified in the Qubit® 3.0 fluorometer and subsequently stored at -80 °C.

2.3. Construction of a metagenomic library and sequencing on the PacBio Sequel II platform

The community DNA pools were subjected to the HiFi library construction process using the SMRTbell® Express Template Prep Kit 2.0 (Pacific Biosciences©) according to the Pacific Biosciences standard template preparation protocol for building libraries greater than 10 kb. The recovered fragments were subsequently sequenced on the PacBio Sequel II platform. All library preparation steps were carried out by the Georgia Genomics and Bioinformatics Core (GGBC) laboratory at the University of Georgia (UGA). PacBio SMRT sequencing involves circularized molecules known as SMRTbells within Zero Mode Waveguides (ZMWs), and the reads obtained from each ZMW are referred to as subreads.

2.4. Quality filtering of long reads

The obtained raw reads (subreads) were demultiplexed using *lima* tool[4] (v. 1.9.0) setting the following parameters "*-j 7 --peek-guess --same --min-score* 0 *--dump-removed --split-bam-named*". Subsequently, the sets of subreads, *i.e.* the reads from the same SMRTbell molecule, were submitted to the Circular Consensus Sequencing (CCS) protocol. The sets of subreads for each library (SC1/SC2) were processed using the CCS tool (v. 4.2.0) to generate one representative sequence

[4] https://github.com/PacificBiosciences/barcoding

(consensus read) per ZMW. This process was performed irrespective of the number of passes (*--min-passes* 0) but considering a minimum read quality (*--min-rq* 0.90). Both tools are components of the SMRT® Link software package.

2.5. Metagenome assembly and genome retrieval

The metagenomics data (CCS reads) were assembled using metaFlye (v. 2.8.3) (Kolmogorov et al. 2020) with parameters set to consider PacBio HiFi data ("*--pacbio-hifi"*) and providing an expected sum of all genome lengths as 110 megabases (*"--genome-size* 110m"). Thereafter, the Metagenome-Assembled Genomes (MAGs) binning was performed by metaWRAP (v. 1.2.1) (Uritskiy et al. 2018), utilizing the *binning* module that takes advantage of multiple binning algorithms: MaxBin2 (Wu et al. 2016) (*--maxbin2*), metaBAT2 (Kang et al. 2019) (*--metabat2*), and CONCOCT (Alneberg et al. 2014) (*--concoct*). A refinement step was performed using the *bin_refinement* module, considering a threshold for completeness (*-c* 50; *i.e.* completeness > 50 %) and contamination (*-x* 10; *i.e.* contamination < 10 %).

Additionally, the resulting bins were also refined by removing contigs with divergent genomic properties using RefineM (v0.1.2) (Parks et al. 2017) with default settings. Subsequently, we subjected these bins to evaluation with MAGPurify (v1.026) (Nayfach et al. 2019) to identify contamination in MAGs. Contamination, in this context, refers to the contigs originating from a different species relative to the dominant one present in the MAG.

The CCS reads were mapped back to the refined MAGs using minimap2 (v. 2.15-r905) (Li 2018) allowing for up to 20% sequence divergence (*-x asm20*). These reads mapped within each MAG were retrieved and employed in the complete Circlator pipeline (Hunt et al. 2015) through the Canu assembler (Koren et al. 2017) (*--assembler canu*) in an attempt to achieve circularization of the MAGs.

At last, the integrity and contamination of the MAGs were evaluated by checkM (v1.0.13) (Parks et al. 2015) using the *lineage_wf* workflow to automatically select an appropriate set of lineage-specific marker genes for the assessment. We followed the criteria outlined in the Minimum Information about a

Metagenome-Assembled Genome (MIMAG) of bacteria and archaea (Bowers et al. 2017) to assign an overall quality status for each genome assembled.

2.6. Taxonomic assignment and phylogenomic analysis

The Kraken tool (v.2.0.7-beta) (Wood et al. 2019) was employed for taxonomic assignment. We utilized a custom database comprising the NCBI non-redundant nucleotide sequences, as well as the RefSeq (Pruitt et al. 2005) genome sequences from bacteria, archaea, and viruses (October/2021). Only the whole genomes with assembly levels of “complete genome”, “chromosome”, and “scaffold” were used to build the Kraken database.

We also performed the taxonomic assignment using the ‘*classify_wf*’ function of the GTDB-tk tool (v.1.7) (Chaumeil et al. 2019) with default settings. This function assigns taxonomy based on the GTDB reference database (Release 202) (Parks et al. 2022), using a combination of evidence, including phylogenetic placement, relative evolutionary divergence (RED) (Parks et al. 2018), and average nucleotide identity (ANI) values using the genomes within the GTDB reference tree.

The molecular phylogenetic tree was inferred using the maximum likelihood method (ML). The tree was built from the alignment of 124 single-copy marker genes, which were identified with the BUSCO tool (v5.2.2) (Manni et al. 2021) using the bacterial dataset (bacteria_odb10/2020-03-06). Our phylogenomic analysis was conducted using the BUSCO_Phylogenomics.py[5] script to reconstruct species phylogenies from 639 representative genomes (REPR).

2.7. Gene prediction and functional annotation

The prediction of protein-coding genes, tRNAs, and rRNAs in the MAGs was carried out using Prokka (v1.11), a prokaryotic genome annotation pipeline (Seemann 2014).

[5] https://github.com/jamiemcg/BUSCO_phylogenomics

The predicted amino acid sequences were subjected to functional annotation with the EnrichM[6] tool (v0.5.0). The '*annotate'* module was used to align proteins using Diamond (v0.9.22) (Buchfink et al. 2015) against the EnrichM v10 database, which incorporates a KO-annotated UniRef100 database. The parameters were configured to set alignment thresholds at 40% for identity (*-id* 0.4), 90% for query coverage (*--aln_query* 0.9), and 90% for reference coverage (*--aln_reference* 0.9). Additionally, the *'classify'* was used to identify the KEGG Orthology (KO) modules recognized as complete.

Gene clusters associated with the biosynthesis of secondary metabolites were identified using antiSMASH (v5.0) (Blin et al. 2019) considering the bacterial domain (*--taxon bacteria*). We also set additional parameters to compare identified clusters against databases of known predictions (*--cb-general --cb- knownclusters --cb-subclusters*), to activate site finder analysis *--asf*), to run Pfam to Gene Ontology mapping module (*--pfam2go*), and to generate phylogenetic trees of identified cluster orthologous groups (*--smcog-trees*).

2.8. Genes related to plant growth-promoting traits

The annotation of proteins related to Plant Growth-Promoting Traits (PGPTs) was performed using the PGPT-Pred web tool for plant-associated bacteria PLaBAse[7] (v.1.02) (Patz et al. 2021).

For annotations of PGPTs, we exclusively considered those identified with KOs previously annotated by the enrichM tool. The PGPT density was estimated by the ratio between the number of PGPTs to the total number of Coding Sequences (CDS).

The enrichment analysis for each genome was performed using PGPT presence/absence data up to level 4 in the PGPT annotation hierarchy. The PGPT frequencies in this level were assessed by comparing the traits annotated in the MAGs (observed PGPTs) to the distribution of all PGPTs (expected PGPTs). The enrichment value was achieved by Pearson's Chi-square test in R (v. 3.6.2). The

[6] https://github.com/geronimp/enrichM
[7] http://plabase.informatik.uni-tuebingen.de/pb/plabase.php

post-hoc analysis based on the residuals of the Chi-squared test was computed using the R package "*chisq.posthoc.test*". The *p-values* were corrected using Benjamini–Hochberg False Discovery Rate (FDR), and a significance threshold of 0.1 (*q-value* ≤ 0.1) was applied to identify enriched features.

The richness of PGPTs was assessed using the Chao1 index, implemented through the fossil package (v.0.4.0) (Vavrek 2011) in the R environment.

2.9. Data availability

The metagenomics data generated for this study were deposited in the NCBI - Sequence Read Archive (SRA), under BioProject ID PRJNA933557. The MAG sequences recovered in this study are available in the GenBank database at the National Center for Biotechnology Information (NCBI) under the accession numbers: JAYGIX000000000 (MAG01), JAYGIY000000000 (MAG02), JAYGIZ000000000 (MAG03), JAYGJA000000000 (MAG04), JAYGJB000000000 (MAG05), JAYGJC000000000 (MAG06), JAYGJD000000000 (MAG07), JAYGJE000000000 (MAG08), JAYGJF000000000 (MAG09), JAYGJG000000000 (MAG10), JAYGJH000000000 (MAG11), JAYGJI000000000 (MAG12), JAYGJJ000000000 (MAG13), JAYGJK000000000 (MAG14), JAYGJL000000000 (MAG15), JAYGJM000000000 (MAG16) and JAYGJN000000000 (MAG17).

3. RESULTS AND DISCUSSION

3.1. Metagenome-assembled genomes and taxonomy assignment

The high-throughput sequencing of the two community pools by PacBio Sequel II platform using HiFi protocol resulted in a total of 25,822,498 subreads, of which 12,297,135 were obtained for the SC1, and 13,525,363 for the SC2 library, both exhibited an average read size of approximately 11 kb (Appendix A). From these sequences, a total of 702,188 CCS (75%) were generated with an average precision of approximately 98.30% and an average length of 8,559 bp. The

sequencing data processing efforts yielded a total of 17 MAGs. According to the completeness and contamination values assessed by CheckM (Table 1), and the quality standards established by the Genomic Standards Consortium (Bowers et al. 2017), five MAGs were circularized and were designated as “finished”, five MAGs as “high-quality”, six as “medium quality”, and the remaining one as “low quality”.

Table 1. Basic assembly statistics of MAGs and sanity report obtained by CheckM tool.

	MAG	Species	Nº contigs	Total size (Mb)	GC	CPT	CTM	HTN	Marker lineage	MIMAG
SC1	MAG01	*Sphingobium yanoikuyae*	15	4.94	64.6%	98%	0.7%	75%	o__Sphingomonadales	HQ Draft
	MAG02	*Stenotrophomonas bentonitica*	1	4.20	67.0%	100%	0.2%	0%	f__Xanthomonadaceae	Finished
	MAG03	*Rhizobium pusense*	2	5.22	59.2%	100%	0.6%	0%	f__Rhizobiaceae	HQ Draft
	MAG04	*Pseudoxanthomonas* sp.	39	4.38	69.6%	90%	3.3%	81%	f__Xanthomonadaceae	HQ Draft
	MAG05	*Stenotrophomonas pavanii*	69	2.44	67.4%	54%	0.6%	100%	f__Xanthomonadaceae	MQ Draft
	MAG06	*Enterobacter asburiae*	1	4.87	56.0%	100%	0.6%	0%	f__Enterobacteriaceae	HQ Draft
SC2	MAG07	*Stenotrophomonas pavanii*	69	4.12	67.4%	65%	5.2%	100%	k__Bacteria	MQ Draft
	MAG08	*Sphingobacterium multivorum*	1	6.32	40.2%	100%	1.5%	0%	p__Bacteroidetes	Finished
	MAG09	*Pantoea dispersa*	58	2.21	58.2%	51%	1.6%	89%	f__Enterobacteriaceae	MQ Draft
	MAG10	*Enterobacter asburiae*	1	5.04	56.0%	99%	1.0%	0%	f__Enterobacteriaceae	Finished
	MAG11	*Chryseobacterium* sp.	55	4.45	36.9%	84%	0.0%	0%	o__Flavobacteriales	MQ Draft
	MAG12	*Paenibacillus* sp.	61	2.62	47.5%	43%	0.0%	0%	k__Bacteria	LQ Draft
	MAG13	*Lactococcus lactis*	1	2.67	35.0%	100%	0.4%	0%	o__Lactobacillales	Finished
	MAG14	*Achromobacter animicus*	1	6.17	65.5%	100%	0.5%	0%	o__Burkholderiales	Finished
	MAG15	*Stenotrophomonas maltophilia*	2	4.52	66.6%	99%	0.0%	0%	f__Xanthomonadaceae	HQ Draft
	MAG16	*Burkholderia gladioli*	107	5.38	68.3%	58%	2.1%	0%	o__Burkholderiales	MQ Draft
	MAG17	*Citrobacter werkmanii*	55	4.02	52.2%	69%	1.7%	0%	k__Bacteria	MQ Draft

GC = Guanine-Cytosine content, CPT = Completeness, CTM = Contamination, and HTN = Heterogeneity

The taxonomic assignment results contain MAGs distributed among three bacterial phyla: Pseudomonadota (n = 13), Bacillota (n = 2), and Bacteroidota (n = 2). These phyla are most predominantly related to phyllosphere communities of various plants (Trivedi et al. 2020). The phylum Pseudomonadota stands out as the most representative, encompassing nine genera: *Achromobacter*, *Burkholderia*, *Citrobacter*, *Enterobacter*, *Pantoea Pseudoxanthomonas*, *Rhizobium*, *Sphingobium* and *Stenotrophomonas*. The genera *Lactococcus* and *Paenibacillus* are representatives of the phylum Bacillota, and the genera *Chryseobacterium* and *Sphingobacterium* belong to the phylum Bacteroidota (Appendix B). Moreover, a phylogenetic tree was reconstructed from 124 bacterial BUSCO single-copy genes to assess the phylogenetic relatedness of the MAGs and representative species (Figure 1).

Most MAGs were classified at the species level with the GTDB-tk tool (Chaumeil et al. 2019) and exhibited an ANI above 95%, a threshold traditionally employed for species discrimination (Goris et al. 2007) (Appendix C). However, the MAG belonging to the genus *Chryseobacterium* (MAG11) showed the highest similarity with the genome of a species not yet characterized (Appendix C). This genome had been previously sequenced and shared an ANI of 97.23% with the genome of *Chryseobacterium* sp. ON_d1 (GenBank ID: GCF_006829085.1). According to the phylogenetic tree, the closest relative is the species *Chryseobacterium gleum* (Figure 1). For MAG12, which belongs to the genus *Paenibacillus*, no taxonomic assignment at the species level was possible. It grouped closely with the species of *Paenibacillus hunanensis* (GenBank ID: GCF 014645615.1), suggesting the potential discovery of a new species (Figure 1).

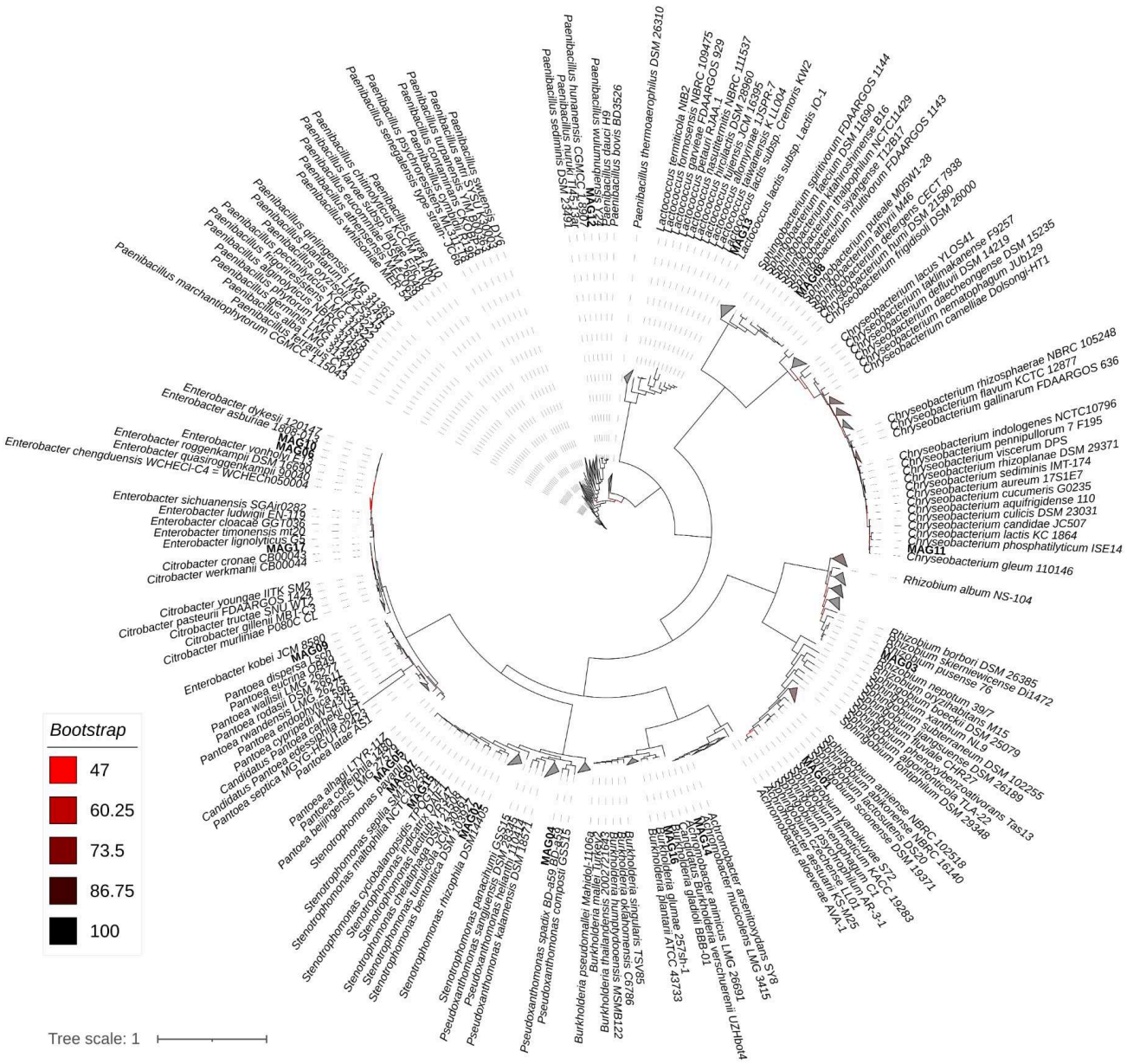

Figure 1. The maximum-likelihood phylogenetic tree constructed using the concatenated alignment of 124 single-copy marker genes found in up to 70% of the species of the 639 representative genomes from the GenBank database, as well as the MAGs belonging to the phyla Pseudomonadota, Bacteroidota, and Bacillota. Clades containing 5 or more representatives are collapsed. The color scale of the branches indicates bootstrap values (based on 1000 replicates), ranging from the lowest support (red) to the highest support (black).

3.2. Functional Profile

The functional diversity of the bacterial genomes was assessed based on the KEGG Ortholog (KO) groups (Table 2 and Appendix D). The metabolic reconstruction reveals a total of 281 complete and 121 incomplete KEGG modules. A detailed presentation of the individual description of pathways/modules by genomes can be found in Supporting Information Appendix E.

Table 2. Descriptive information about structural and functional annotation of the MAGs.

MAG	Species	CDS	tRNAs	16S	23S	5S	tmRNAs	HPs	KOs	PGPTs	PGPTs (Density)
MAG01	*Sphingobium yanoikuyae*	4,576	75	4	4	4	1	1,847	2,147	1,437	0.31
MAG02	*Stenotrophomonas bentonitica*	3,653	70	3	3	4	2	1,308	1,985	1,323	0.36
MAG03	*Rhizobium pusense*	4,881	56	4	4	4	1	1,508	2,861	2,076	0.43
MAG04	*Pseudoxanthomonas* sp.	4,025	56	3	2	2	1	1,534	1,696	1,147	0.29
MAG05	*Stenotrophomonas pavanii*	2,324	51	1	1	2	0	972	1,004	653	0.28
MAG06	*Enterobacter asburiae*	4,392	84	8	8	9	1	707	3,182	2,276	0.52
MAG07	*Stenotrophomonas pavanii*	3,718	58	1	2	3	1	1,392	1,882	1,289	0.35
MAG08	*Sphingobacterium multivorum*	5,328	83	7	7	7	1	2,604	1,938	1,259	0.24
MAG09	*Pantoea dispersa*	2,255	43	4	3	3	1	526	1,136	802	0.36
MAG10	*Enterobacter asburiae*	4,641	84	8	8	9	1	903	3,182	2,281	0.49
MAG11	*Chryseobacterium* sp.	4,341	69	-	-	-	1	2,434	1,271	810	0.19
MAG12	*Paenibacillus* sp.	2,313	11	-	-	-	1	926	910	621	0.27
MAG13	*Lactococcus lactis*	2,640	71	6	6	7	1	991	1,297	828	0.32
MAG14	*Achromobacter animicus*	5,558	63	4	4	5	1	1,738	3,087	2,236	0.40
MAG15	*Stenotrophomonas maltophilia*	4,060	83	4	4	4	1	1,440	2,205	1,488	0.37
MAG16	*Burkholderia gladioli*	4,968	40	-	-	-	1	1,946	1,998	1,460	0.29
MAG17	*Citrobacter werkmanii*	4,007	67	5	5	5	0	743	2,395	1,682	0.42

HPs = Hypothetical proteins not annotated by the Prokka pipeline and eggNOG mapper.
KOs = Represents the count of coding sequences (CDS) annotated with KEGG orthologs.
PGPTs = Represents the count of coding sequences (CDS) identified with Plant Growth-Promoting Traits (PGPTs).
PGPT Density = Ratio between the number of PGPTs to the total number of Coding Sequences (CDS).

Among the KEGG pathways, we identified some important modules that are related to beneficial traits from plant growth-promoting bacteria (PGPBs), such as metabolism of nitrogen (N), solubilization and uptake of phosphate (P), and sulfur (S) metabolism, as well as mechanisms involved in alleviation of abiotic stresses, antioxidants, and biosynthesis of secondary metabolites (terpenoids, pyrrolnitrin, polyamines) (Figure 2A). In addition, we identified the gene encoding ACC deaminase (acdS), commonly recognized as important for plant-bacteria interaction as it is associated with the modulation of ethylene levels in plants. We also identified several other genes encoding enzymes related to the biosynthesis of phytohormones, such as auxin (IAA), cytokinin (CK), and salicylic acid (SA) (Figure 2B).

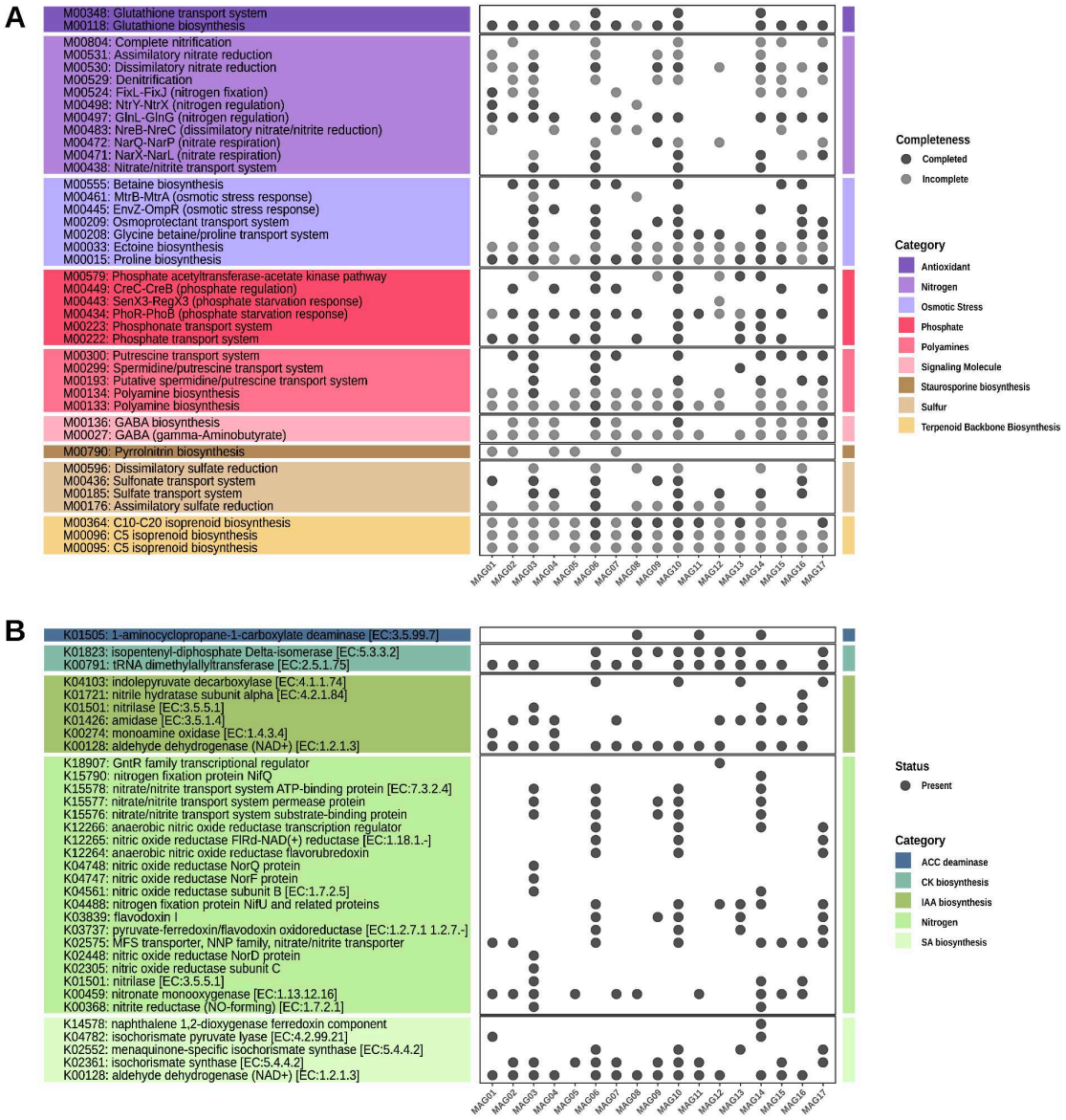

Figure 2. Key genomic features observed in MAGs, which are commonly associated with plant growth-promoting bacteria. These features were represented here as (A) KEGG modules or (B) KEGG orthology (KO) identifiers, which are not included in modules.

Regarding the potential of biological nitrogen fixation, we have not identified nitrogenases, which are essential for this process. However, the genomes of *Rhizobium punsense* (MAG03), *E. asburiae* (MAG06 and MAG10), *P. dispersa* (MAG09), *A. animicus* (MAG14), and *C. werkmanii* (MAG17) can be considered nitrate-reducing bacteria according to the ability of these genomes to transform nitrate (NO_3-) into ammonium (NH_4^+) by dissimilatory nitrate reduction (DNRA) (Figure 2A).

The 1-aminocyclopropane-1-carboxylate (ACC) deaminase (*acdS*; K01505), is an enzyme responsive to modulating the level of ethylene production, decreasing ethylene levels in plants, and mitigating adverse effects of high accumulation of this hormone on plant growth (Ali and Kim 2018). The *acdS* was identified for the MAGs *Sphingobacterium multivorum* (MAG08), *Chryseobacterium* sp. (MAG11), and *Achromobacter animicus* (MAG14) (Figure 2A and Appendix D). The presence of this enzyme in those genera has been mainly related to saline and hydric stress alleviation and plant growth promotion (Tittabutr et al. 2013, Bhise et al. 2017, Araújo et al. 2020, Nascimento et al. 2021a).

The genes *gcd* (K00117), *gdh* (K00034), and *gnl* (K01053) encode enzymes involved in inorganic P solubilization through gluconic acid (GA), the primary organic compound associated with this process. The genomes of *S. yanoikuyae* (MAG01), *R. pusense* (MAG03), *Chryseobacterium* sp. (MAG11), and *A. animicus* (MAG14) carry genes encoding the enzymes responsible for GA production. MAG14 encodes the enzyme 2-ketogluconate reductase (EC 1.1.99.3; K06151 and K06152), which can convert GA into 2-Keto-D-gluconic acid, another organic acid associated with the solubilization of inorganic P (Appendix D). Moreover, these MAGs harbor genes related to P uptake (*phnC* [K02041], *phnD* [K02044], and *phnE* [K02042]) and phosphate transport system (*pstA* [K02038], *pstB* [K02036], *pstC* [K02037], and *pstS* [K02040]), alongside regulatory genes (*phorR* [K07636] and *phorB* [K07657]) (Figure 2A and Table Appendix E). These findings suggest that the MAGs can increase phosphate uptake, enhance soil P turnover, and potentially increase its availability to plants (Liang et al. 2020).

Indole-3-acetic acid (IAA) production was identified in some MAGs (Figure 2A). Bacterial IAA synthesis generally occurs by tryptophan (trp) dependent pathways, which include three major ones: Indole-3-pyruvate (IPyA), tryptamine (TAM), indole-3-acetamide (IAM) (Ludueña et al. 2019, Zhang et al. 2019). However, the tryptophan side chain oxidase (TSO) pathway is most commonly observed in strains of *Pseudomonas fluorescens* (Li et al. 2018).

The MAGs exhibit pathways related to the biosynthesis of tryptophan (Trp) from chorismate (M00023; Appendix E). Except *P. dispersa* (MAG09) and *B. gladioli* (MAG16) lack genes related to Trp synthesis.

The IPyA pathway is considered the most common pathway in bacteria (Zhang et al. 2019). In this study, this pathway was identified in *E. asburiae* related genomes (MAG06, and MAG10), since they encode two main enzymes responsible for the conversion of IAA intermediates. Indole-3-pyruvate is converted to indole-3-acetaldehyde (IAAld) by the key enzyme indolepyruvate decarboxylase (K04103; EC 4.1.1.74) (Schütz et al. 2003, Harris et al. 2018). The resulting compound from decarboxylation is oxidized by the aldehyde dehydrogenase (K00128; EC 1.2.1.3) into Indole-3-acetic acid (IAA) (Figure 2B), a compound commonly found in *Enterobacter* species (Luziatelli et al. 2020).

The presence of amidase (K01426; EC. 3.5.1.4) in the genomes of *S. bentonitica* (MAG02), *R. pusens*e (MAG03), *Pseudoxanthomonas* sp. (MAG04), *S. pavanii* (MAG07), *Paenibacillus sp.* (MAG12), *L. lactis* (MAG013), *A. animicus* (MAG14), *S. maltophilia* (MAG15), and *B. gladioli* (MAG16), suggests their capability to synthesize IAA by the IAM pathway, utilizing the intermediate indole-3-acetamide (Figure 2B). The *Pseudoxanthomonas* sp. (MAG04) probably exhibits IAA synthesis through the TAM pathway, which intermediates the conversion of tryptamine (TAM) to IAAld by monoamine oxidase (K00274; EC 1.4.3.4). The final step is catalyzed by aldehyde dehydrogenase (K00128; EC 1.2.1.3).

The enzymatic and non-enzymatic antioxidant processes consist of mechanisms that attenuate abiotic stress factors in plants, mediated by plant growth-promoting bacteria (Zandi and Schnug 2022). The biosynthesis pathway of polyamines, as well as the transport systems for spermidine and putrescine, were identified in *E. asburiae* (MAG06 and MAG10) and *R. pusense* (MAG03). These compounds have been reported to be involved in cell growth inducing root growth (Wu et al. 2012, Zhou et al. 2016), and frequently are related to plants' responses to various abiotic stresses (Bhise et al. 2017, Selim et al. 2021). Other compounds related to abiotic stress-alleviating effects in plants, acting as osmoprotectants, include proline (Kartik et al. 2021), glycine betaine

(Ahmed et al. 2021), glutathione (Santos et al. 2018), ectoine (Tao et al. 2016), etc. The ectoine biosynthetic cluster was identified in the *S. yanoikuyae* (MAG01) and *A. animicus* (MAG14) (Appendix F). Several clusters of secondary metabolites known to enhance plant health, such as siderophores (Aznar and Dellagi 2015), and non-ribosomal peptides (NRPs) (Ongena and Jacques 2008) were identified in the genomes (Appendix F). The MAGs of *S. yanoikuyae* (MAG01), *Chryseobacterium* sp. (MAG11), and *B. gladioli* (MAG16), stand out for having a high number of biosynthetic gene clusters (BGCs) (n > 5) (Appendix F).

Chitinolytic enzymes, including chitinase [EC:3.2.1.14] and endoglucanase [EC:3.2.1.4], were identified in almost all MAGs, except for *Chryseobacterium* sp. (MAG11) and *Paenibacillus* sp. (MAG12). These enzymes are often associated with antifungal activity (Philip et al. 2020, Sharma et al. 2020).

3.3. Functional PGPT annotation

The genomic features of each MAG related to the beneficial functions in plant growth are based on PGPT mapping of the annotated KOs (Appendix G). The heatmap displays the abundance of main PGPTs at hierarchical level 4 (n = 39) (Appendix H). The abundance profile of traits resulted in MAGs being clustered into two groups. This result also highlights categories which have significant enrichment (*q-value* ≤ 0.1) in at least one genome. The MAGs assigned as *E. asburie* (MAG06 and MAG10), *C. werkmanii* (MAG17), *A. animicus* (MAG14), and *R. punsense* (MAG03) are members of cluster I and were grouped based on the highest abundance and distribution of PGPTs in level 4 hierarchy (Figure 3). This is a consequence of the higher density of PGPTs in these MAGs when compared to the members of cluster II (Table 2), whose PGPT abundances at the same level were relatively lower (Figure 3).

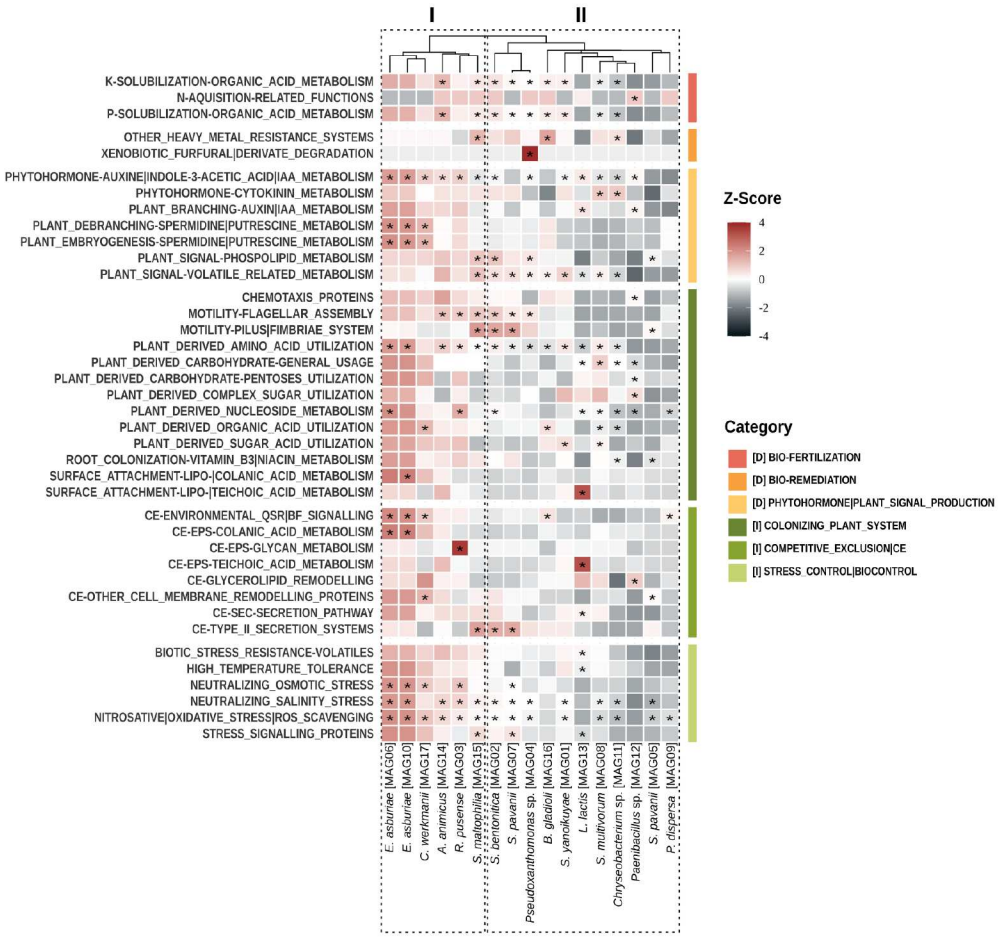

Figure 3. Abundance and enrichment of traits related to plant growth promotion (PGP). The heatmap illustrates the distribution of PGPTs at hierarchical level 4. The PGPTs annotations are based on KEGG Orthology (KO), and the values were normalized by the z-score. Enriched traits within each MAG (*q-value* ≤ 0.1) are marked with asterisks (*). Categories preceded by [D] and [I] indicate direct and indirect effects on plant growth promotion, respectively.

With respect to the direct effect on plant growth promotion, the category of bio-fertilization exhibits enrichment of essential functions aimed at enhancing plant nutrient acquisition, particularly phosphorus (P) and potassium (K) solubilization. These processes are primarily mediated by the production of several organic compounds (Appendix G). Most MAGs exhibited enrichment for these two classes (Figure 3).

Except for the *S. pavanni* MAGs (MAG05 and MAG07), all the others have genomic features related to IAA production, demonstrated either by the

enrichment of the IAA metabolism trait (Figure 3) or by the presence of orthologs associated with the biosynthesis of this molecule (Figure 2B). While *P. dispersa* (MAG09) and *B. gladioli* (MAG16) do not show enrichment for this trait due to the absence of genes related to tryptophan biosynthesis, the ability to produce IAA from exogenous tryptophan or intermediate compounds may still be possible. This is supported by the presence of EC:1.2.1.3 in both genomes, and EC:3.5.1.4, EC:3.5.5.1, and EC:4.2.1.84 in *B. gladioli* (MAG16), which are related to the acid indole acetic pathway.

While *P. dispersa* (MAG09) and *B. gladioli* (MAG16) do not show enrichment for this trait due to the absence of genes related to tryptophan biosynthesis, the ability to produce IAA from exogenous tryptophan or intermediate compounds may still be possible, supported by the presence of genes encoding enzymes related to the acid indole acetic pathway, the EC 1.2.1.3 in both genomes and EC 3.5.1.4, EC 3.5.5.1, EC 4.2.1.84 in *B. gladioli* (MAG16). Additionally, the MAGs assigned with *S. multivorum* (MAG08) and *Chryseobacterium* sp. (MAG11) exhibit enrichment for cytokinin (CK) metabolism, inferred by the identification of genes related to the biosynthesis of isoprenoids (C10-C20) (Figure 3) and the presence of genes such as *miaA*, *miaB*, and *miaE*, responsible for the production of zeatin (Figure 2A: “M00364” and Appendix E), which is directly associated with the synthesis of CKs (Nascimento et al. 2021b).

Among the MAGs, only those assigned to *Stenotrophomonas* spp. showed enrichment for components of the type II secretion system (T2SS) (Figure 3). This reveals the ability to secrete different types of proteins and transport substrates across their cell membrane, which are commonly found in many gram-negative bacteria (Korotkov et al. 2012). Additionally, these *Stenotrophomonas spp.* genomes have fimbriae and flagellar assembly systems. These systems, which also occur in *A. animicus* (MAG14), *R. pusense* (MAG03), and *Pseudoxanthomonas* sp. (MAG04) (Figure 3), enhances bacterial cell adhesion to the root surface – a meaningful feature for endophyte colonization (Ahlawat et al. 2022).

Another factor to be considered in plant colonization is the usage of plant-derived compounds (Figure 3). Protein families related to sugar and amino acid transport were found in microorganisms identified as robust colonizers within a synthetic community derived from the sugarcane microbiome (de Souza et al. 2019). The family of ATP-independent tripartite periplasmic transporters (TRAP transporters) is specialized in the absorption, exchange, or efflux of C4-dicarboxylates (*dctABD)* (Janausch et al. 2002). We found genes related to the uptake of C4-dicarboxylates, citrate (*citABCDEFGTX* and *tctABCDE*), malate (*maeAN*), and fumarate (*dcuBRS*) (Appendix D). These organic acids are commonly found in exudates from sugarcane roots and stems (Singh and Mukerji 2006, Glassop et al. 2007). The TRAP transporters are absent in the MAGs related to *S. bentonitica* (MAG02), *Pseudoxanthomonas* sp. (MAG04), *S. pavanii* (MAG05 and MAG07), *S. multivorum* (MAG08), *Chryseobacterium* sp. (MAG11), *Paenibacillus* sp. (MAG12), *L. lactis* (MAG13), and *S. maltophilia* (MAG15). The absence of TRAP transporters in some genomes may indicate a potential limitation in their ability to efficiently transport and utilize plant-derived C4-dicarboxylates. On the other hand, several genomes showed enrichment of the “plant amino acid utilization” feature, except for the MAGs related to *S. pavanii* (MAG05), *P. dispersa* (MAG09), *Paenibacillus* sp. (MAG12), and *C. werkmanii* (MAG17).

In general, the MAGs exhibit different mechanisms that may be involved in the colonization process, some of which were previously related to robust colonization (de Souza et al. 2019). Furthermore, the broad enrichment for the stress control category (Figure 3) suggests that mechanisms related to ROS scavenging, osmotic, and saline stress neutralization, may be involved with the robust colonization style of sugarcane. This is supported by findings in several studies (Amna et al. 2020, Guo et al. 2020, Chandra et al. 2021).

In Figure 4, we display the functional richness distribution of the PGPTs for each MAG, based on the Chao1 metric. This includes the total richness, which corresponds to the sum of all functions associated with the PGPTs of all MAGs. This analysis provides an overview of the functional potential and the distribution of functional categories found in the bacterial recovered genomes.

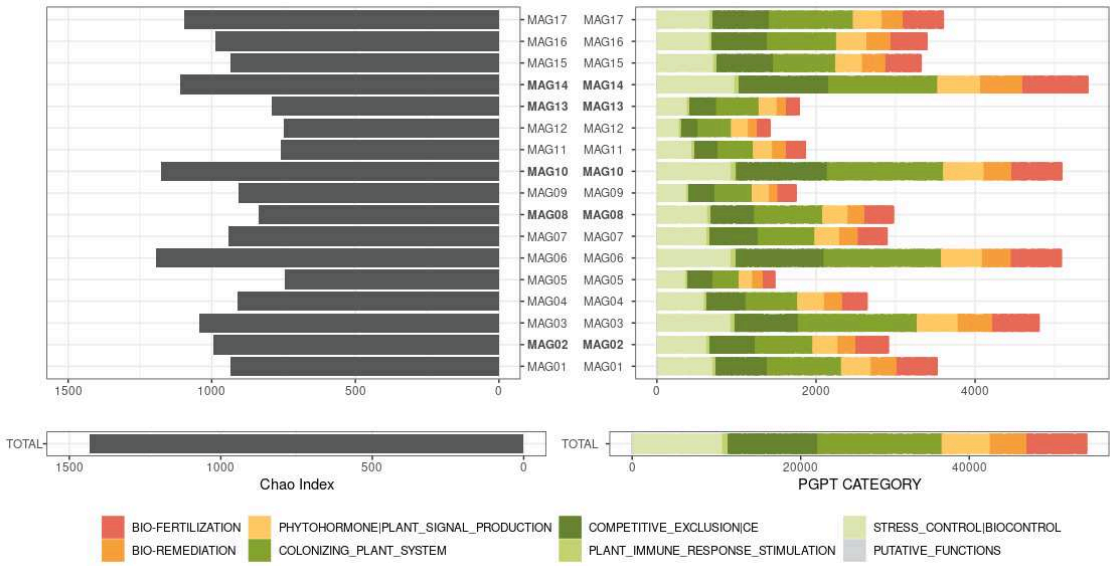

Figure 4. Functional diversity of PGPTs (Plant Growth-Promoting Traits) of each genome and the aggregated sum of all genomes (total). In the bar chart on the left, the Chao1 index corresponds to the measure of functional richness, highlighting the variety of specific functions and their distribution in different genomes. The stacked bar chart on the right represents the distribution of functions of PGPT at hierarchical level 2. The highlighted MAGs on the y-axis indicate that the genomes are completed.

In the context presented, we identified five genomes (MAGs) with functional richness indices greater than 1,000 (Figure 4, bar graph on the left). These include the species: *Rhizobium pusense* (MAG03), *Enterobacter asburie* (MAG06 and MAG10), *Achromobacter animicus* (MAG14), and *Citrobacter werkmanii* (MAG17). It is noteworthy that in previous analyses, this group of MAGs with high functional richness was already observed. However, here we specifically highlight MAG14, which exhibits many PGPT categories, as depicted in Figure 4.

Species of the genus *Achromobacter* are mainly associated with clinical samples but have also been isolated from other environments, including soils (Gomila et al. 2011, Zhang et al. 2023), sewage sludge (Pradeep et al. 2015), effluent water industrial plants (Sreeja Mole S S et al. 2021). Additionally, they can be found in plant association (Kuncharoen et al. 2017, Mohamadpoor et al. 2022), exhibiting endophytic and plant growth-promoting characteristics.

Considering the genus' importance as an emerging opportunistic pathogen, we evaluated the genome of *Achromobacter animicus* (MAG14)

searching for virulence factors using the PathogenFinder tool (Cosentino et al. 2013). This tool indicated a low probability (26%) of MAG14 being a human pathogen. Our results suggest that MAG14 has significant potential as a plant growth promoter, exhibiting multiple functions related to plant interactions. Among them, we emphasize its capacity for ACC deaminase production, and synthesis of indoleacetic acid through the conversion of indole-3-acetamide. Additionally, MAG14 can also produce different organic acids that act in the solubilization of phosphate and potassium. These traits likely contribute to the plant growth properties previously reported for this genus (Jha and Kumar 2009, Romero-Perdomo et al. 2019).

4. CONCLUSIONS

The efforts undertaken in this work allowed the recovery of a series of genomes related to microorganisms with the potential to promote plant growth. Furthermore, an in-depth investigation of their functional profiles was possible, shedding light on the processes of plant-microorganism symbiotic interaction. The findings provide valuable insights concerning the main traits of bacteria associated with plants. Additionally, the results demonstrate the importance of complementarity between individuals that compose microbial communities, highlighting the role of compatibility in the remodeling or construction of simplified or synthetic communities. In future studies, we plan to test these organisms directly on plants to validate the traits and processes estimated by the genetic features.

5. REFERENCES

Ahlawat OP, Yadav D, Kashyap PL, Khippal A, Singh G (2022) Wheat endophytes and their potential role in managing abiotic stress under changing climate. **Journal of Applied Microbiology** 132: 2501–2520. https://doi.org/10.1111/jam.15375

Ahmed S, Heo T-Y, Roy Choudhury A, Walitang DI, Choi J, Sa T (2021) Accumulation of compatible solutes in rice (*Oryza sativa* L.) cultivars by inoculation of endophytic plant growth promoting bacteria to alleviate salt stress. **Applied Biological Chemistry** 64: 68. https://doi.org/10.1186/s13765-021-00638-x

Ali S, Kim W-C (2018) Plant Growth Promotion Under Water: Decrease of Waterlogging-Induced ACC and Ethylene Levels by ACC Deaminase-Producing Bacteria. **Frontiers in Microbiology** 9: 1096. https://doi.org/10.3389/fmicb.2018.01096

Almeida A, Nayfach S, Boland M, Strozzi F, Beracochea M, Shi ZJ, Pollard KS, Sakharova E, Parks DH, Hugenholtz P, Segata N, Kyrpides NC, Finn RD (2021) A unified catalog of 204,938 reference genomes from the human gut microbiome. **Nature Biotechnology** 39: 105–114. https://doi.org/10.1038/s41587-020-0603-3

Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C (2014) Binning metagenomic contigs by coverage and composition. **Nature Methods** 11: 1144–1146. https://doi.org/10.1038/nmeth.3103

Amna, Xia Y, Farooq MA, Javed MT, Kamran MA, Mukhtar T, Ali J, Tabassum T, Rehman S ur, Hussain Munis MF, Sultan T, Chaudhary HJ (2020) Multi-stress tolerant PGPR *Bacillus xiamenensis* PM14 activating sugarcane (Saccharum officinarum L.) red rot disease resistance. **Plant Physiology and Biochemistry** 151: 640–649. https://doi.org/10.1016/j.plaphy.2020.04.016

Araújo VLVPD, Lira Junior MA, Souza Júnior VSD, De Araújo Filho JC, Cury Fracetto FJ, Andreote FD, De Araujo Pereira AP, Mendes Júnior JP, Rêgo Barros FMD, Monteiro Fracetto GG (2020) Bacteria from tropical semiarid temporary ponds promote maize growth under hydric stress. **Microbiological Research** 240: 126564. https://doi.org/10.1016/j.micres.2020.126564

Armanhi JSL, de Souza RSC, Biazotti BB, Yassitepe JE de CT, Arruda P (2021) Modulating Drought Stress Response of Maize by a Synthetic Bacterial Community. **Frontiers in Microbiology** 12: 747541. https://doi.org/10.3389/fmicb.2021.747541

Aznar A, Dellagi A (2015) New insights into the role of siderophores as triggers of plant immunity: what can we learn from animals? **Journal of Experimental Botany** 66: 3001–3010. https://doi.org/10.1093/jxb/erv155

Berg G, Köberl M, Rybakova D, Müller H, Grosch R, Smalla K (2017) Plant microbial diversity is suggested as the key to future biocontrol and health

trends. **FEMS Microbiology Ecology** 93. https://doi.org/10.1093/femsec/fix050

Bhise KK, Bhagwat PK, Dandge PB (2017) Synergistic effect of *Chryseobacterium gleum* sp. SUK with ACC deaminase activity in alleviation of salt stress and plant growth promotion in *Triticum aestivum* L. **3 Biotech** 7: 105. https://doi.org/10.1007/s13205-017-0739-0

Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T (2019) antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. **Nucleic Acids Research** 47: W81–W87. https://doi.org/10.1093/nar/gkz310

Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA et al. (2017) Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. **Nature Biotechnology** 35: 725–731. https://doi.org/10.1038/nbt.3893

Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. Nature Methods 12: 59–60. https://doi.org/10.1038/nmeth.3176

Cai H, Bai Y, Guo C (2018) Comparative genomics of 151 plant-associated bacteria reveal putative mechanisms underlying specific interactions between bacteria and plant hosts. **Genes & Genomics** 40: 857–864. https://doi.org/10.1007/s13258-018-0693-1

de Carvalho LAL, Teheran-Sierra LG, Funnicelli MIG, da Silva RC, Campanari MFZ, de Souza RSC, Arruda P, Soares MA, Pinheiro DG (2021) Farming systems influence the compositional, structural, and functional characteristics of the sugarcane-associated microbiome. **Microbiological Research** 252: 126866. https://doi.org/10.1016/j.micres.2021.126866

Chandra A, Chandra P, Tripathi P (2021) Whole genome sequence insight of two plant growth-promoting bacteria (*B. subtilis* BS87 and *B. megaterium* BM89) isolated and characterized from sugarcane rhizosphere depicting better crop yield potentiality. **Microbiological Research** 247: 126733. https://doi.org/10.1016/j.micres.2021.126733

Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH (2019) GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Hancock J (Ed.). **Bioinformatics**: btz848. https://doi.org/10.1093/bioinformatics/btz848

Cosentino S, Voldby Larsen M, Møller Aarestrup F, Lund O (2013) PathogenFinder - Distinguishing Friend from Foe Using Bacterial Whole Genome Sequence Data. Barbour JD (Ed.). **PLoS ONE** 8: e77302. https://doi.org/10.1371/journal.pone.0077302

Cuscó A, Pérez D, Viñes J, Fàbregas N, Francino O (2021) Long-read metagenomics retrieves complete single-contig bacterial genomes from canine feces. **BMC Genomics** 22: 330. https://doi.org/10.1186/s12864-021-07607-0

Dastogeer KMG, Tumpa FH, Sultana A, Akter MA, Chakraborty A (2020) Plant

microbiome–an account of the factors that shape community composition and diversity. **Current Plant Biology** 23: 100161. https://doi.org/10.1016/j.cpb.2020.100161

Driscoll CB, Otten TG, Brown NM, Dreher TW (2017) Towards long-read metagenomics: complete assembly of three novel genomes from bacteria dependent on a diazotrophic cyanobacterium in a freshwater lake co-culture. Standards in **Genomic Sciences** 12: 9. https://doi.org/10.1186/s40793-017-0224-8

Emami S, Alikhani HA, Pourbabaei AA, Etesami H, Motashare Zadeh B, Sarmadian F (2018) Improved growth and nutrient acquisition of wheat genotypes in phosphorus deficient soils by plant growth-promoting rhizospheric and endophytic bacteria. Soil **Science and Plant Nutrition** 64: 719–727. https://doi.org/10.1080/00380768.2018.1510284

Glassop D, Roessner U, Bacic A, Bonnett GD (2007) Changes in the Sugarcane Metabolome with Stem Development. Are They Related to Sucrose Accumulation? **Plant and Cell Physiology** 48: 573–584. https://doi.org/10.1093/pcp/pcm027

Gomila M, Tvrzová L, Teshim A, Sedláček I, González-Escalona N, Zdráhal Z, Šedo O, González JF, Bennasar A, Moore ERB, Lalucat J, Murialdo SE (2011) *Achromobacter marplatensis* sp. nov., isolated from a pentachlorophenol-contaminated soil. International **Journal of Systematic and Evolutionary Microbiology** 61: 2231–2237. https://doi.org/10.1099/ijs.0.025304-0

Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM (2007) DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. **International Journal of Systematic and Evolutionary Microbiology** 57: 81–91. https://doi.org/10.1099/ijs.0.64483-0

Großkopf T, Soyer OS (2014) Synthetic microbial communities. **Current Opinion in Microbiology** 18: 72–77. https://doi.org/10.1016/j.mib.2014.02.002

Guo D-J, Singh RK, Singh P, Li D-P, Sharma A, Xing Y-X, Song X-P, Yang L-T, Li Y-R (2020) Complete Genome Sequence of *Enterobacter roggenkampii* ED5, a Nitrogen Fixing Plant Growth Promoting Endophytic Bacterium With Biocontrol and Stress Tolerance Properties, Isolated From Sugarcane Root. **Frontiers in Microbiology** 11: 580081. https://doi.org/10.3389/fmicb.2020.580081

Hamonts K, Trivedi P, Garg A, Janitz C, Grinyer J, Holford P, Botha FC, Anderson IC, Singh BK (2018) Field study reveals core plant microbiota and relative importance of their drivers. **Environmental Microbiology** 20: 124–140. https://doi.org/10.1111/1462-2920.14031

Haro-Moreno JM, López-Pérez M, Rodriguez-Valera F (2021) Enhanced Recovery of Microbial Genes and Genomes From a Marine Water Column Using Long-Read Metagenomics. **Frontiers in Microbiology** 12: 708782. https://doi.org/10.3389/fmicb.2021.708782

Harris DMM, Berrué F, Kerr R, Patten CL (2018) Metabolomic analysis of indolepyruvate decarboxylase pathway derivatives in the rhizobacterium *Enterobacter cloacae*. **Rhizosphere** 6: 98–111. https://doi.org/10.1016/j.rhisph.2018.06.002

Hassani MA, Durán P, Hacquard S (2018) Microbial interactions within the plant holobiont. **Microbiome** 6: 58. https://doi.org/10.1186/s40168-018-0445-0

Hungria M, Nogueira MA, Araujo RS (2013) Co-inoculation of soybeans and common beans with rhizobia and azospirilla: strategies to improve sustainability. **Biology and Fertility of Soils** 49: 791–801. https://doi.org/10.1007/s00374-012-0771-5

Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR (2015) Circlator: automated circularization of genome assemblies using long sequencing reads. Genome Biology 16: 294. https://doi.org/10.1186/s13059-015-0849-0

IMG/M Data Consortium, Nayfach S, Roux S, Seshadri R, Udwary D, Varghese N, Schulz F, Wu D, Paez-Espino D, Chen I-M, Huntemann M, Palaniappan K et al. (2020) A genomic catalog of Earth's microbiomes. **Nature Biotechnology.** https://doi.org/10.1038/s41587-020-0718-6

Janausch IG, Zientz E, Tran QH, Kröger A, Unden G (2002) C4-dicarboxylate carriers and sensors in bacteria. **Biochimica et Biophysica Acta (BBA) - Bioenergetics** 1553: 39–56. https://doi.org/10.1016/S0005-2728(01)00233-X

Jha P, Kumar A (2009) Characterization of Novel Plant Growth Promoting Endophytic Bacterium *Achromobacter xylosoxidans* from Wheat Plant. **Microbial Ecology** 58: 179–188. https://doi.org/10.1007/s00248-009-9485-0

Kang D, Jacquiod S, Herschend J, Wei S, Nesme J, Sørensen SJ (2020) Construction of Simplified Microbial Consortia to Degrade Recalcitrant Materials Based on Enrichment and Dilution-to-Extinction Cultures. **Frontiers in Microbiology** 10: 3010. https://doi.org/10.3389/fmicb.2019.03010

Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z (2019) MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. **PeerJ** 7: e7359. https://doi.org/10.7717/peerj.7359

Kartik VP, Jinal HN, Amaresan N (2021) Inoculation of Cucumber (Cucumis sativus L.) Seedlings with Salt-Tolerant Plant Growth Promoting Bacteria Improves Nutrient Uptake, Plant Attributes and Physiological Profiles. **Journal of Plant Growth Regulation** 40: 1728–1740. https://doi.org/10.1007/s00344-020-10226-w

Kaur T, Devi R, Kumar S, Sheikh I, Kour D, Yadav AN (2022) Microbial consortium with nitrogen fixing and mineral solubilizing attributes for growth of barley (*Hordeum vulgare* L.). **Heliyon** 8: e09326. https://doi.org/10.1016/j.heliyon.2022.e09326

Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, Kuhn K, Yuan J, Polevikov E, Smith TPL, Pevzner PA (2020) metaFlye: scalable long-read metagenome assembly using repeat graphs. **Nature Methods** 17:

1103–1110. https://doi.org/10.1038/s41592-020-00971-x

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM (2017) Canu: scalable and accurate long-read assembly via adaptive *k* -mer weighting and repeat separation. **Genome Research** 27: 722–736. https://doi.org/10.1101/gr.215087.116

Korir H, Mungai NW, Thuita M, Hamba Y, Masso C (2017) Co-inoculation Effect of Rhizobia and Plant Growth Promoting Rhizobacteria on Common Bean Growth in a Low Phosphorus Soil. **Frontiers in Plant Science** 08. https://doi.org/10.3389/fpls.2017.00141

Korotkov KV, Sandkvist M, Hol WGJ (2012) The type II secretion system: biogenesis, molecular architecture and mechanism. **Nature Reviews Microbiology** 10: 336–351. https://doi.org/10.1038/nrmicro2762

Kumar S, Diksha, Sindhu SS, Kumar R (2022) Biofertilizers: An ecofriendly technology for nutrient recycling and environmental sustainability. **Current Research in Microbial Sciences** 3: 100094. https://doi.org/10.1016/j.crmicr.2021.100094

Kuncharoen N, Muramatsu Y, Shibata C, Kamakura Y, Nakagawa Y, Tanasupawat S (2017) *Achromobacter aloeverae* sp. nov., isolated from the root of *Aloe vera* (L.) Burm.f. **International Journal of Systematic and Evolutionary Microbiology** 67: 37–41. https://doi.org/10.1099/ijsem.0.001566

Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. Birol I (Ed.). **Bioinformatics** 34: 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li M, Guo R, Yu F, Chen X, Zhao H, Li H, Wu J (2018) Indole-3-Acetic Acid Biosynthesis Pathways in the Plant-Beneficial Bacterium Arthrobacter pascens ZZ21. International **Journal of Molecular Sciences** 19: 443. https://doi.org/10.3390/ijms19020443

Liang J-L, Liu J, Jia P, Yang T, Zeng Q, Zhang S, Liao B, Shu W, Li J (2020) Novel phosphate-solubilizing bacteria enhance soil phosphorus cycling following ecological restoration of land degraded by mining. **The ISME Journal** 14: 1600–1613. https://doi.org/10.1038/s41396-020-0632-4

Ludueña LM, Anzuay MS, Angelini JG, McIntosh M, Becker A, Rupp O, Goesmann A, Blom J, Fabra A, Taurian T (2019) Genome sequence of the endophytic strain *Enterobacter* sp. J49, a potential biofertilizer for peanut and maize. b 111: 913–920. https://doi.org/10.1016/j.ygeno.2018.05.021

Luziatelli F, Ficca AG, Bonini P, Muleo R, Gatti L, Meneghini M, Tronati M, Melini F, Ruzzi M (2020) A Genetic and Metabolomic Perspective on the Production of Indole-3-Acetic Acid by *Pantoea agglomerans* and Use of Their Metabolites as Biostimulants in Plant Nurseries. **Frontiers in Microbiology** 11: 1475. https://doi.org/10.3389/fmicb.2020.01475

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM (2021) BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral

Genomes. Kelley J (Ed.). Molecular **Biology and Evolution** 38: 4647–4654. https://doi.org/10.1093/molbev/msab199

Marks BB, Megías M, Ollero FJ, Nogueira MA, Araujo RS, Hungria M (2015) Maize growth promotion by inoculation with *Azospirillum brasilense* and metabolites of Rhizobium tropici enriched on lipo-chitooligosaccharides (LCOs). **AMB Express** 5: 71. https://doi.org/10.1186/s13568-015-0154-z

Mohamadpoor M, Amini J, Ashengroph M, Azizi A (2022) Evaluation of biocontrol potential of *Achromobacter xylosoxidans* strain CTA8689 against common bean root rot. **Physiological and Molecular Plant Pathology** 117: 101769. https://doi.org/10.1016/j.pmpp.2021.101769

Nascimento FX, Glick BR, Rossi MJ (2021a) Multiple plant hormone catabolism activities: an adaptation to a plant‐associated lifestyle by *Achromobacter* spp. **Environmental Microbiology Reports** 13: 533–539. https://doi.org/10.1111/1758-2229.12987

Nascimento FX, Urón P, Glick BR, Giachini A, Rossi MJ (2021b) Genomic Analysis of the 1-Aminocyclopropane-1-Carboxylate Deaminase-Producing *Pseudomonas thivervalensis* SC5 Reveals Its Multifaceted Roles in Soil and in Beneficial Interactions With Plants. **Frontiers in Microbiology** 12: 752288. https://doi.org/10.3389/fmicb.2021.752288

Nayfach S, Shi ZJ, Seshadri R, Pollard KS, Kyrpides NC (2019) New insights from uncultivated genomes of the global human gut microbiome. **Nature** 568: 505–510. https://doi.org/10.1038/s41586-019-1058-x

Ongena M, Jacques P (2008) *Bacillus* lipopeptides: versatile weapons for plant disease biocontrol. **Trends in Microbiology** 16: 115–125. https://doi.org/10.1016/j.tim.2007.12.009

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. **Genome Research** 25: 1043–1055. https://doi.org/10.1101/gr.186072.114

Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P (2022) GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. **Nucleic Acids Research** 50: D785–D794. https://doi.org/10.1093/nar/gkab776

Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, Hugenholtz P (2018) A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. **Nature Biotechnology** 36: 996–1004. https://doi.org/10.1038/nbt.4229

Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW (2017) Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nature Microbiology 2: 1533–1542. https://doi.org/10.1038/s41564-017-0012-7

Parray JA, Shameem N, Abd‐Allah EF, Mir MY (Eds) (2022) Core Microbiome:

Improving Crop Quality and Productivity. 1st ed. **Wiley**. https://doi.org/10.1002/9781119830795

Patz S, Gautam A, Becker M, Ruppel S, Rodríguez-Palenzuela P, Huson Dh (2021) PLaBAse: A comprehensive web resource for analyzing the plant growth-promoting potential of plant-associated bacteria. **Bioinformatics**. preprint https://doi.org/10.1101/2021.12.13.472471

Philip NV, Koteshwara A, Kiran GA, Raja S, Subrahmanyam VM, Chandrashekar HR (2020) Statistical Optimization for Coproduction of Chitinase and Beta 1, 4-Endoglucanase by Chitinolytic *Paenibacillus elgii* PB1 Having Antifungal Activity. **Applied Biochemistry and Biotechnology** 191: 135–150. https://doi.org/10.1007/s12010-020-03235-8

Pradeep S, Sarath Josh MK, Binod P, Sudha Devi R, Balachandran S, Anderson RC, Benjamin S (2015) *Achromobacter denitrificans* strain SP1 efficiently remediates di(2-ethylhexyl)phthalate. **Ecotoxicology and Environmental Safety** 112: 114–121. https://doi.org/10.1016/j.ecoenv.2014.10.035

Pruitt KD, Tatusova T, Maglott DR (2005) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. **Nucleic acids research** 33: D501–D504.

Rojas-Solís D, Zetter-Salmón E, Contreras-Pérez M, Rocha-Granados M del C, Macías-Rodríguez L, Santoyo G (2018) *Pseudomonas stutzeri* E25 and *Stenotrophomonas maltophilia* CR71 endophytes produce antifungal volatile organic compounds and exhibit additive plant growth-promoting effects. **Biocatalysis and Agricultural Biotechnology** 13: 46–52. https://doi.org/10.1016/j.bcab.2017.11.007

Romero-Perdomo F, Ocampo-Gallego J, Camelo-Rusinque M, Bonila R (2019) Plant growth promoting rhizobacteria and their potential as bioinoculants on *Pennisetum clandestinum* (Poaceae). **Revista de Biología Tropical** 67. https://doi.org/10.15517/rbt.v67i4.34029

Sánchez-Cañizares C, Jorrín B, Poole PS, Tkacz A (2017) Understanding the holobiont: the interdependence of plants and their microbiome. **Current Opinion in Microbiology** 38: 188–196. https://doi.org/10.1016/j.mib.2017.07.001

Santos A de A, Silveira JAG da, Bonifacio A, Rodrigues AC, Figueiredo M do VB (2018) Antioxidant response of cowpea co-inoculated with plant growth-promoting bacteria under salt stress. **Brazilian Journal of Microbiology** 49: 513–521. https://doi.org/10.1016/j.bjm.2017.12.003

Santos LF, Olivares FL (2021) Plant microbiome structure and benefits for sustainable agriculture. **Current Plant Biology** 26: 100198. https://doi.org/10.1016/j.cpb.2021.100198

Sarkar S, Ward K, Kamke A, Ran Q, Feehan B, Richie T, Reese N, Lee STM (2022) Perspective: Simple State Communities to Study Microbial Interactions: Examples and Future Directions. **Frontiers in Microbiology** 13: 801864. https://doi.org/10.3389/fmicb.2022.801864

Schütz A, Golbik R, Tittmann K, Svergun DI, Koch MHJ, Hübner G, König S (2003) Studies on structure-function relationships of indolepyruvate decarboxylase from *Enterobacter cloacae*, a key enzyme of the indole acetic acid pathway: Structure-function studies of *E. cloacae* IDPC. **European Journal of Biochemistry** 270: 2322–2331. https://doi.org/10.1046/j.1432-1033.2003.03602.x

Seemann T (2014) Prokka: rapid prokaryotic genome annotation. **Bioinformatics** 30: 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Selim S, Madany MMY, Mohamed Reyad A, F Alowaiesh B, Hagagy N, Al-Sanea MM, Alsharari SS, AbdElgawad H (2021) *Saccharomonospora actinobacterium* alleviates phytotoxic hazards of tungsten nanoparticles on legumes' growth and osmotic status. **Journal of Environmental Chemical Engineering** 9: 106395. https://doi.org/10.1016/j.jece.2021.106395

Setubal JC (2021) Metagenome-assembled genomes: concepts, analogies, and challenges. **Biophysical Reviews** 13: 905–909. https://doi.org/10.1007/s12551-021-00865-y

Sharma CK, Vishnoi VK, Dubey RC, Maheshwari DK (2018) A twin rhizospheric bacterial consortium induces systemic resistance to a phytopathogen *Macrophomina phaseolina* in mung bean. **Rhizosphere** 5: 71–75. https://doi.org/10.1016/j.rhisph.2018.01.001

Sharma S, Kumar S, Khajuria A, Ohri P, Kaur R, Kaur R (2020) Biocontrol potential of chitinases produced by newly isolated *Chitinophaga* sp. S167. **World Journal of Microbiology and Biotechnology** 36: 90. https://doi.org/10.1007/s11274-020-02864-9

Singh G, Mukerji KG (2006) Root Exudates as Determinant of Rhizospheric Microbial Biodiversity. In: Mukerji KG, Manoharachary C, Singh J (Eds), Microbial Activity in the **Rhizoshere**. Springer-Verlag, Berlin/Heidelberg, 39–53. https://doi.org/10.1007/3-540-29420-1_3

Singh P, Singh RK, Li H-B, Guo D-J, Sharma A, Lakshmanan P, Malviya MK, Song X-P, Solanki MK, Verma KK, Yang L-T, Li Y-R (2021) Diazotrophic Bacteria *Pantoea dispersa* and *Enterobacter asburiae* Promote Sugarcane Growth by Inducing Nitrogen Uptake and Defense-Related Gene Expression. **Frontiers in Microbiology** 11: 600417. https://doi.org/10.3389/fmicb.2020.600417

Singleton CM, Petriglieri F, Kristensen JM, Kirkegaard RH, Michaelsen TY, Andersen MH, Kondrotaite Z, Karst SM, Dueholm MS, Nielsen PH, Albertsen M (2021) Connecting structure to function with the recovery of over 1000 high-quality metagenome-assembled genomes from activated sludge using long-read sequencing. **Nature Communications** 12: 2009. https://doi.org/10.1038/s41467-021-22203-2

de Souza RSC, Armanhi JSL, Damasceno N de B, Imperial J, Arruda P (2019) Genome Sequences of a Plant Beneficial Synthetic Bacterial Community Reveal Genetic Features for Successful Plant Colonization. **Frontiers in Microbiology** 10: 1779. https://doi.org/10.3389/fmicb.2019.01779

de Souza RSC, Okura VK, Armanhi JSL, Jorrín B, Lozano N, da Silva MJ, González-Guerrero M, de Araújo LM, Verza NC, Bagheri HC, Imperial J, Arruda P (2016) Unlocking the bacterial and fungal communities assemblages of sugarcane microbiome. **Scientific Reports** 6: 28774. https://doi.org/10.1038/srep28774

Sreeja Mole S S, D. S. Vijayan, M. Anand, M. Ajona, T. Jarin (2021) Biodegradation of P-nitro phenol using a novel bacterium *Achromobacter denitrifacians* isolated from industrial effluent water. Water Science and Technology: **A Journal of the International Association on Water Pollution Research** 84: 3334–3345. https://doi.org/10.2166/wst.2021.354

Su P, Wicaksono WA, Li C, Michl K, Berg G, Wang D, Xiao Y, Huang R, Kang H, Zhang D, Cernava T, Liu Y (2022) Recovery of metagenome-assembled genomes from the phyllosphere of 110 rice genotypes. **Scientific Data** 9: 254. https://doi.org/10.1038/s41597-022-01320-7

Tao P, Li H, Yu Y, Gu J, Liu Y (2016) Ectoine and 5-hydroxyectoine accumulation in the halophile Virgibacillus halodenitrificans PDB-F2 in response to salt stress. **Applied Microbiology and Biotechnology** 100: 6779–6789. https://doi.org/10.1007/s00253-016-7549-x

Teheran-Sierra LG, Funnicelli MIG, de Carvalho LAL, Ferro MIT, Soares MA, Pinheiro DG (2021) Bacterial communities associated with sugarcane under different agricultural management exhibit a diversity of plant growth-promoting traits and evidence of synergistic effect. **Microbiological Research** 247: 126729. https://doi.org/10.1016/j.micres.2021.126729

Tittabutr P, Piromyou P, Longtonglang A, Noisa-Ngiam R, Boonkerd N, Teaumroong N (2013) Alleviation of the effect of environmental stresses using co-inoculation of mungbean by *Bradyrhizobium* and rhizobacteria containing stress-induced ACC deaminase enzyme. **Soil Science and Plant Nutrition** 59: 559–571. https://doi.org/10.1080/00380768.2013.804391

Tosi M, Mitter EK, Gaiero J, Dunfield K (2020) It takes three to tango: the importance of microbes, host plant, and soil management to elucidate manipulation strategies for the plant microbiome. **Canadian Journal of Microbiology** 66: 413–433. https://doi.org/10.1139/cjm-2020-0085

Trivedi P, Leach JE, Tringe SG, Sa T, Singh BK (2020) Plant–microbiome interactions: from community assembly to plant health**. Nature Reviews Microbiology** 18: 607–621. https://doi.org/10.1038/s41579-020-0412-1

Uritskiy GV, DiRuggiero J, Taylor J (2018) MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. **Microbiome** 6: 158. https://doi.org/10.1186/s40168-018-0541-1

Vavrek MJ (2011) fossil: palaeoecological and palaeogeographical analysis tools. **Palaeontologia Electronica** 14: 1T.

Wang M, Eyre AW, Thon MR, Oh Y, Dean RA (2020) Dynamic Changes in the Microbiome of Rice During Shoot and Root Growth Derived From Seeds. **Frontiers in Microbiology** 11: 559728. https://doi.org/10.3389/fmicb.2020.559728

Wood DE, Lu J, Langmead B (2019) Improved metagenomic analysis with Kraken 2. Genome Biology 20: 257. https://doi.org/10.1186/s13059-019-1891-0

Wu Q-S, He X-H, Zou Y-N, Liu C-Y, Xiao J, Li Y (2012) Arbuscular mycorrhizas alter root system architecture of Citrus tangerine through regulating metabolism of endogenous polyamines. **Plant Growth Regulation** 68: 27–35. https://doi.org/10.1007/s10725-012-9690-6

Wu Y-W, Simmons BA, Singer SW (2016) MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. **Bioinformatics** 32: 605–607. https://doi.org/10.1093/bioinformatics/btv638

Xie F, Jin W, Si H, Yuan Y, Tao Y, Liu J, Wang X, Yang C, Li Q, Yan X, Lin L, Jiang Q, Zhang L, Guo C, Greening C, Heller R, Guan LL, Pope PB, Tan Z, Zhu W, Wang M, Qiu Q, Li Z, Mao S (2021) An integrated gene catalog and over 10,000 metagenome-assembled genomes from the gastrointestinal microbiome of ruminants. Microbiome 9: 137. https://doi.org/10.1186/s40168-021-01078-x

Zandi P, Schnug E (2022) Reactive Oxygen Species, Antioxidant Responses and Implications from a Microbial Modulation Perspective. **Biology** 11: 155. https://doi.org/10.3390/biology11020155

Zhang B, Xu W, Ma Y, Gao X, Ming H, Jia J (2023) Effects of bioaugmentation by isolated *Achromobacter xylosoxidans* BP1 on PAHs degradation and microbial community in contaminated soil. **Journal of Environmental Management** 334: 117491. https://doi.org/10.1016/j.jenvman.2023.117491

Zhang P, Jin T, Kumar Sahu S, Xu J, Shi Q, Liu H, Wang Y (2019) The Distribution of Tryptophan-Dependent Indole-3-Acetic Acid Synthesis Pathways in Bacteria Unraveled by Large-Scale Genomic Analysis. **Molecules** 24: 1411. https://doi.org/10.3390/molecules24071411

Zhou C, Ma Z, Zhu L, Xiao X, Xie Y, Zhu J, Wang J (2016) Rhizobacterial Strain Bacillus megaterium BOFC15 Induces Cellular Polyamine Changes that Improve Plant Growth and Drought Resistance. **International Journal of Molecular Sciences** 17: 976. https://doi.org/10.3390/ijms17060976

APPENDIX

The appendices for this chapter are available on Zenodo, an open-access general-purpose repository maintained by OpenAIRE and CERN. This repository allows the deposit of research articles, datasets, research software, reports, and any other digital artifacts related to research.

Access to the appendices for this chapter can be found at the following permanent link: https://doi.org/10.5281/zenodo.10429292

Appendix A. Summary of counts of subreads and circular consensus sequencing (CCS) sequences obtained for PacBio sequencing of SMRT libraries. (EMS_1.xlsx)

Appendix B. Taxonomy assignment of MAGs at the higher taxonomic rank obtained from GTDB-tk and Kraken tools. (EMS_2.xlsx)

Appendix C. Report of the classification workflow using GTDB-tk. (EMS_3.xlsx)

Appendix D. Matrix of the KEGG Orthology (KOs) frequencies annotated by the EnrichM tool. (EMS_4.xlsx)

Appendix E. Reconstruction and completeness of KEGG modules annotated by EnrichM. The asterisks (*) in the header represent additional values obtained by the script ‘classKEGGModules.pl’ (https://github.com/dgpinheiro/bioinfoutilities) to estimate PGPTs in KEGG modules. (EMS_5.xlsx)

Appendix F. The secondary metabolite biosynthesis gene clusters (BGCs) identified with AntiSMASH. (EMS_6.xlsx)

Appendix G. The raw count of plant growth-promoting traits (PGPTs) annotations, according to KEGG Orthology (KO) predictions for MAGs. (EMS_7.xlsx)

Appendix H. The raw count of plant growth-promoting traits (PGPTs) that comprises the 39 classes (level 4 hierarchy) identified as enriched according to the results of Pearson's Chi-square test adjusted by False Discovery Rate (*q-value* ≤ 0.1). (EMS_8.xlsx)

CHAPTER 3 – Classification of genomic features of plant-associated bacteria using machine learning

ABSTRACT - Plants harbor a diverse array of microorganisms within various niches, with bacteria playing a crucial role in plant associations by adapting specific traits to thrive in their hosts. To gather insights into the genomic features associated with this lifestyle adaptation, we conducted an evaluation of 1808 bacterial genome sequences across 13 genera obtained from the GenBank database. Among these, 129 genomes were identified as plant-associated, with 190 specifically associated with roots, 1,304 isolated from non-plant sources (including humans, non-human animals, air, sediments, and aquatic environments), and 184 derived from soil samples. Analysis of the plant-associated genomes revealed a higher enrichment of orthologous gene groups (orthogroups). Employing machine learning (ML) methods to predict plant-associated genomes, the Random Forest (RF) approach outperformed Logistic Regression (LR) models, demonstrating a sensitivity value > 0.8 and specificity > 0.9. Feature selection highlighted that approximately 50% of the enriched orthogroups shared among genera were crucial for plant-associated genomes. These features represent significant functional categories with specific genes related to plant genomes.

Keywords: Comparative genomics; Random forest; Endophytic lifestyle; Host adaptation; Plant-associated environments

1. INTRODUCTION

A substantial number of microorganisms, with bacteria being the most dominant group, are closely associated with plants and many of them have been relatively well studied (Trivedi et al. 2020). The close coexistence with plants requires specific adaptive mechanisms that allow microorganisms to overcome plant defenses and inhabit plant tissues (Hassani et al. 2019). Interactions of plant-adapted microorganisms can vary, having positive (mutualistic), neutral (neutralism), or deleterious (predation) impacts on plant fitness (Thrall et al. 2007, Hassani et al. 2018).

Several studies of interactions between pathogenic bacteria and plants have clarified the responses of the plant's innate immune system, including the mechanisms of pathogen recognition, as well as the mechanisms that pathogens use to subvert the host's immune system. On the other hand, the genetic and molecular mechanisms underlying symbiotic interactions such as Legume-Rhizobium interactions were also elucidated. Furthermore, numerous studies have been conducted to isolate beneficial bacteria (PGPB/PGPR) and investigate the mechanisms involved in promoting plant growth or mitigating adverse conditions.

Other studies have employed comparative genomics to explore traits typical of lifestyles both associated and non-associated with plants in various bacterial groups. In a study conducted by Zhang et al. (2016), a comparative analysis was carried out between plant-associated and non-plant-associated *Bacillus amyloliquefaciens* and *Bacillus subtilis* species. This analysis led to the identification of genes related to the use of plant-derived polysaccharides and the synthesis of antibiotics, which were found to be more abundant in the genome pool of plant-associated strains. In the study of Poncheewin et al. (2022), they adopted a classifier of the plant-associated lifestyle of *Pseudomonas* strains using genome properties and revealed 28 discriminating features. Also, Levy et al. (2018) compared thousands of bacterial genomes and identified orthologous groups and protein domain profiles, which are characteristic of plant adaptation.

In this work, our goal was to identify functional genomic features from certain bacterial genera that have not been considered in previous studies and which are

poorly characterized in plant association. We aimed to investigate mechanisms related to adaptations for colonizing plant environments. To achieve this, we employed machine learning and feature selection approaches.

2. MATERIAL AND METHODS

2.1. Genomic data compilation and metadata

We have previously assembled and inspected the genomes of plant growth-promoting bacteria belonging to the genera *Achromobacter*, *Burkholderia*, *Chryseobacterium*, *Citrobacter*, *Enterobacter*, *Lactococcus*, *Paenibacillus*, *Pantoea*, *Pseudoxanthomonas*, *Rhizobium*, *Sphingobacterium*, *Sphingobium*, *Stenotrophomonas* (Unpublished) within bacterial communities associated with sugarcane (Teheran-Sierra et al. 2021).

To improve comprehension of the relationship of these bacteria with plant-associated environments, we retrieved the sequences of bacterial genomes within these genera with the status of “complete genome” or “chromosome” available in the NCBI Genome database[8] until December 2021.

The genomes were classified according to the isolation sites, distinguishing between plant-associated (PA) - which includes any plant niche, including roots - or root-associated (RA) - which genomes derived solely from endophytic or epiphytic compartments of roots (rhizoplane), and non-plant-associated (NPA) or soil-associated (SA) through a manual curation process. This process included scanning of metadata in the following catalogs: NCBI BioSample (Barrett et al. 2012), Integrated Microbial Genomes (IMG) (Markowitz et al. 2012), Leibniz-Institut - Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ)[9], American Type Culture Collection (ATCC) (Benton et al. 2021), Czech Collection of Microorganisms (CCM)[10], Global Catalogue of Microorganisms (GCM) (Wu et al. 2018), Japan Collection of Microorganisms (JCM)[11], World Data Centre for Microorganisms (WDCM) (Wu et al. 2017), and the related scientific literature.

[8]http://www.ncbi.nlm.nih.gov/genome
[9] https://www.dsmz.de
[10] https://ccm.sci.muni.cz
[11] https://jcm.brc.riken.jp

The bacterial genome fastas (n=1,807) were downloaded from the GenBank database at the National Center for Biotechnology Information (NCBI), and the sequences were then subjected to gene prediction using Prokka (Seemann 2014).

2.2. Construction of the phylogenetic tree

The phylogenomic analysis was performed from the alignment of 124 single-copy marker genes, which were identified with the BUSCO tool (v5.2.2) (Manni et al. 2021), using the bacterial dataset (bacteria_odb10/ 2020- 03-06). The phylogenetic tree was inferred by the maximum likelihood method (ML) from the BUSCO output by using the BUSCO_Phylogenomics.py[12] script to reconstruct the species phylogeny from 1807 bacterial genomes.

2.3. Inference of orthologous gene groups

To infer similarities between protein sequences we used an all-gainst-all pairwise BLASTp alignments among the genomes within each taxonomic group (genus level). The BLASTp alignments were performed using the algorithm implemented in DIAMOND (v 2.0.8) (Buchfink et al. 2015) with the following options: *--ultra-sensitive --query-cover* 90 *--subject-cover* 90 *--top* 5 *--evalue* 0.001. Subsequently, the Orthofinder (2.5.4) tool (Emms and Kelly 2019) was employed with default parameters to infer the orthologous gene groups (orthogroups). In the second step, we retrieved a representative sequence (the longest sequence) of each orthogroup from each taxonomic group and performed a new orthology inference using the same approach, this time, among the taxonomic groups.

2.4. Functional annotation of the representative orthogroups

The representative orthogroups' sequences were annotated using eggNOG-mapper (v2.1.6) (Huerta-Cepas et al. 2017) configured to use DIAMOND-based BLASTp searches against the EggNOG database (v5.0)

[12] https://github.com/jamiemcg/BUSCO_phylogenomics

(Huerta-Cepas et al. 2019) and also the following parameters: *--target_orthologs* one2one, *--sensmode* ultra-sensitive, *--pident* 30, *--query_cover* 90, *--evalue* 0.00001, *--tax_scope* bacteria, *--tax_scope_mode* inner_narrowest *--go_evidence* all.

2.5. Enrichment analysis of orthogroups according to lifestyle

The identification of plant-associated (PA), non-plant-associated (NPA), root-associated (RA), and soil-associated (SA) genes was carried out based on the existence of significantly enriched protein clusters (orthogroups) in the gene proteins set from these defined groups of bacterias (PA, NPA, RA or SA). The enrichment analyses were based on three approaches: the hypergeometric test (Hyperg), PhyloGLM (Ives and Garland 2010), and Scoary (Brynildsrud et al. 2016), similar to those adopted by Levy et al. (2018). For the hypergeometric test, two versions (hypergbin, hypergcn) were used. For both gene copy-number and gene presence/absence data, the *P-values* were corrected by Benjamini–Hochberg False Discovery Rate (FDR) with a threshold of $P < 0.05$. In the PhyloGLM approach, the copy number (phyloglmcn) or presence/absence data (phyloglmbin) of each gene was used as the only independent variable. The positive estimates in the PhyloGLM and $P < 0.05$ indicate enrichment for PA/RA, while negative estimates and $P < 0.05$ exhibit enrichment for NPA/SA. The Scoary uses exclusively the gene presence/absence dataset. The orthogroup was considered significant only if (1) it had a *q-value* (an FDR-adjusted *P-value*) less than 0.05 for Fisher's exact test, (2) the 'worst' *P-value* from the pairwise comparison algorithm was < 0.05, and (3) the empirical (permutation-based test) *P-value* was < 0.05. Odds ratios greater or equal than 1 indicated PA or RA and less than 1 indicated NPA or SA enriched genes.

2.6. Classification of plant-associated (PA) genomes and features (genes) selection

To predict the lifestyle of plant-associated bacteria, we employed two machine learning approaches using the content of bacterial genomes as input, *i.e.* the gene feature tables, considering the copy-number and the presence/absence datasets. One approach is to use logistic regression, a traditional statistical technique, and the

other is to use a random forest classifier, a supervised machine learning algorithm. The evaluation was performed using as input the genome datasets organized by genus. The datasets were partitioned into 70% for training (model building step) and the remaining (30%) for testing (validation step).

The analyses were performed in the statistical program R version 4.1.2 (R Core Team 2021), the logistic regression (LR) was performed using *glmnet* package (Friedman et al. 2021) zeroing ridge-regression penalty parameter (α = 0), and the random forest (RF) was performed using *randomForest* package based on Breiman (2001) algorithm with an increase in the parameter related to the number of trees (ntree = 1000) (Liaw and Wiener 2002). The performance was evaluated using the following metrics: accuracy, error rate, precision, sensitivity, F-score, and Receiver Operating Characteristic (ROC) curves.

The variable importance (varImp) from each classification analysis using the two approaches with all datasets was computed using the function “varImp” of the *caret* package in R (Kuhn 2008). The varImp is a measure of the importance of variables (genes/orthologous) in a machine learning model. It indicates how much each variable contributes to the model's ability to make accurate predictions. This measure is useful for selecting the most relevant features for a given model.

2.7. Enrichment of functional categories of Clusters of Orthologous Groups (COG)

The orthogroups shared among genome groups were subjected to enrichment analysis using Fisher's exact test in R to identify any significant associations. We evaluated the functional categories enriched in the comparison between plant-associated (PA) and non-plant-associated (NPA), as well as between root-associated (RA) and soil-associated genomes.

3. RESULTS

3.1. Bacterial genome groups and phylogenomic analysis

A total of 1,807 sequences of bacterial genomes were retrieved from the Genbank database (Appendix A). Among these, 129 were annotated as plant-associated (PA), 190 as root-associated (RA). Additionally, there were 1,304 non-plant-associated bacterial genomes (NPA) sourced from various environments, such as humans, non-human animals, air, sediments, and aquatic environments, and 184 genomes derived from soil-associated (SA) (Appendix B). The phylogenomic analysis reveals the distribution of taxonomic groups: Alcaligenaceae (*Achromobacter*), Burkholderiacea (*Burkholderia*), Enterobacteriaceae (*Citrobacter* and *Enterobacter*), Erwiniaceae (*Pantoea*), Xanthomonadaceae (*Pseudoxanthomonas*, *Stenotrophomonas*), Rhizobiaceae (*Rhizobium*), and Sphingomonadaceae (*Sphingobium*), all belonging to the phylum Pseudomonadota in addition to Streptococcaceae (*Lactococcus*) and Paenibacillaceae (*Paenibacillus*) representatives of the phylum Bacillota, and Weeksellaceae (*Chryseobacterium*) and Sphingobacteriaceae (*Sphingobacterium*), representatives of the phylum Bacteroidota (Figure 1, Appendix B).

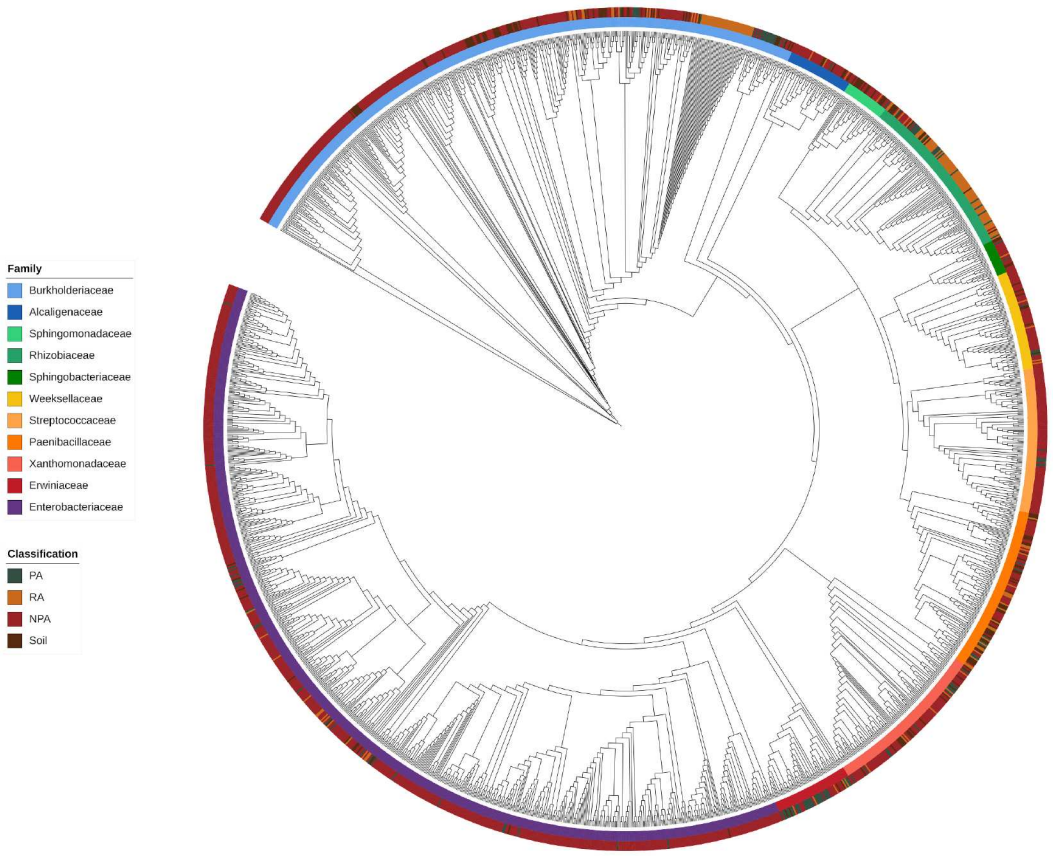

Figure 1. The maximum-likelihood phylogenetic tree was constructed from the concatenated alignment of 124 single-copy marker genes. The inner ring displays the taxonomic group, with colors indicating family level. The outer ring displays the isolation source for each genome, distinguishing among plant-associated (PA) genomes, root-associated (RA), non-plant-associated (NPA), and soil (SA) genomes.

3.2. Orthologous gene enrichment

The results of genus-specific protein clustering analysis, revealed a total of 635,466 orthogroups (Table 1). From the orthogroups' enrichment analysis, we identified the significantly enriched orthogroups related to a plant-associated lifestyle based on any of the employed statistical approaches. The lists of statistically significant PA, RA, SA, and NPA orthogroups for each genus-specific dataset are laid out in Appendix C.

Table 1. Numbers of the identified orthogroups (OGs) into each genus-specific dataset, and also, the enriched genes for each genus according to the considered environments.

Genus	OGs	Enriched OGs					
		NPA	NPA/SA	PA	PA/RA	RA	SA
Achromobacter	22,389	0	0	1	0	0	0
Burkholderia	104,439	1,437	1,761	2,908	3,406	307	645
Chryseobacterium	40,969	0	0	1	0	0	0
Citrobacter	55,059	2	0	0	0	0	0
Enterobacter	87,512	505	0	347	0	0	0
Lactococcus	19,781	13	0	22	0	0	0
Paenibacillus	89,675	32	0	540	0	0	0
Pantoea	25,202	3	0	3	0	0	0
Pseudoxanthomonas	11,079	0	0	0	0	0	0
Rhizobium	79,328	1,387	4	1,122	367	63	7
Sphingobacterium	32,816	0	0	2	0	0	0
Sphingobium	28,622	0	0	0	0	0	0
Stenotrophomonas	38,582	0	0	11	0	0	0

The enrichment analysis detected few orthogroups significantly enriched in the following genus-specific dataset: *Achromobacter* (n=1), *Chryseobacterium* (n=1), *Citrobacter* (n=2), *Pantoea* (n=3), *Pseudoxanthomonas* (n=0), *Sphingobium* (0), with exception of *Citrobacter* and *Pantoea*, the others did not proceed in the subsequent analyzes (Table 1). All these datasets needed more representatives for an appropriate enrichment analysis. Maybe in the future, we will have more useful data for them.

We identified a total of 501,446 orthogroups shared among the genomes. In our results, we observed that the vast majority of these orthogroups were exclusive to a specific genus 92.42% (463,441), while only 7.58% (38,005) were shared among more than two genera groups. Many of these were shared between the *Burkholderia* and *Rhizobium* genera.

Approximately 55% (276,630) of the total orthogroups had a match in at least one of the databases integrated into eggNOG, including COGs (236,990), ECs (47,502), KEGG (107,905), GOs (26,840), CAZy (3,750), and PFAMs (215,846). However, a substantial number of orthologs remain unidentified.

3.3. Performance evaluation of the machine learning approaches

Overall, the random forest approach outperformed the logistic regression models in predicting plant-associated genomes, with a sensitivity value greater than 0.8 and specificity greater than 0.9. An exception was observed for the presence/absence dataset of *Stenotrophomonas* (PA vs. NPA), which showed a sensitivity of 0.375 (Appendix D). The analysis results of the gene count datasets of the group *Sphingobacterium* exhibited a sensitivity of 0.667 (PA vs. NPA), whereas for *Pantoea* (PA vs. NPA) and *Rhizobium* (RA vs. SA) values of specificity were the smallest 0.353 and 0.667, respectively (Appendix D). To visualize these results, we provide ROC curves and their corresponding Area under the ROC Curve (AUC) values for the two employed approaches (Figure 2).

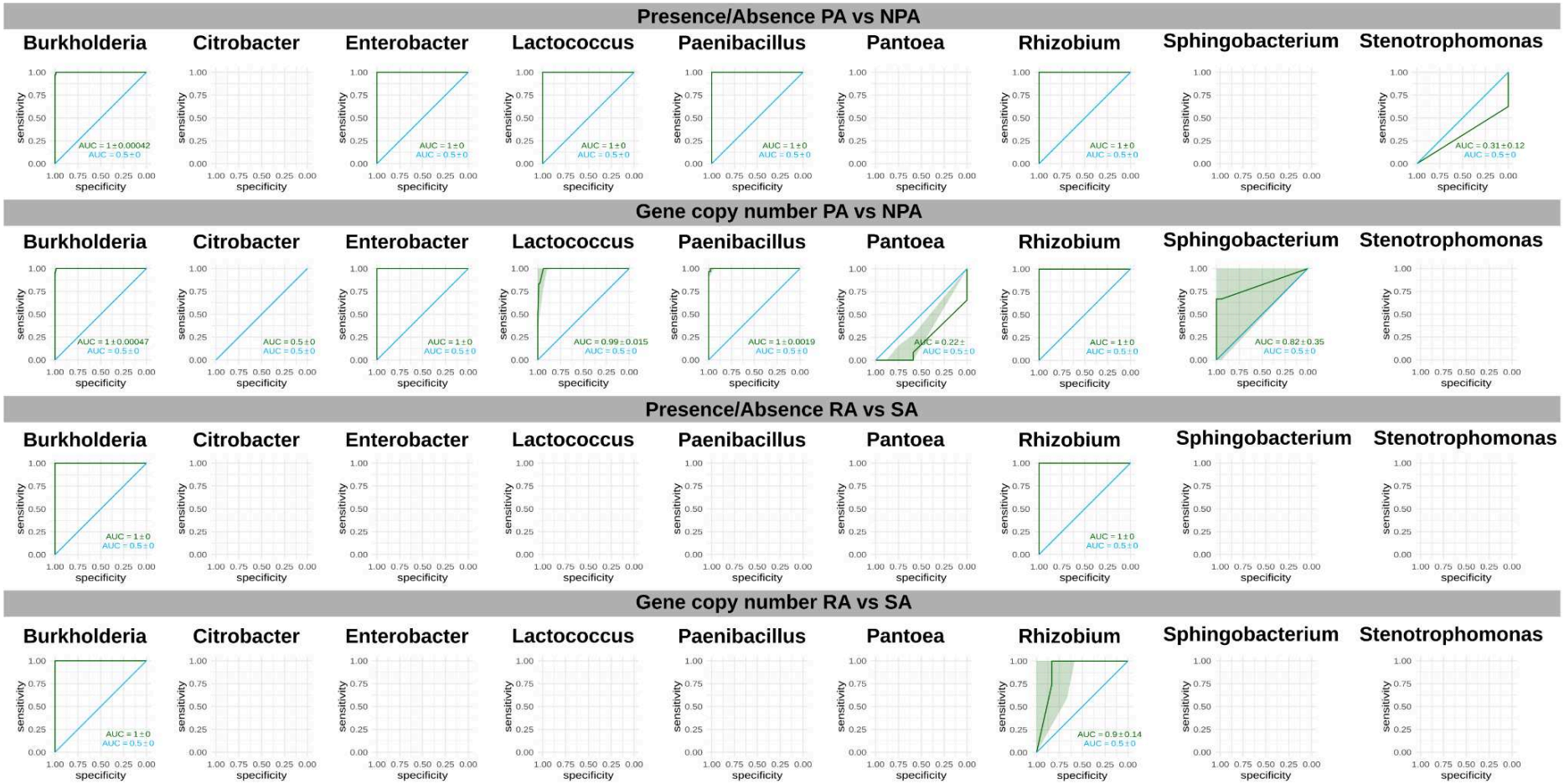

Figure 2. Performance evaluation of random forest and logistic regression classifiers via ROC curves (the better model is the one with a larger area under the curve).

3.4. Functional categories regarding plant-associated bacteria gene content

The most important features (orthogroups) for each genus-specific group, considering each dataset type (presence/absence or gene count number), were obtained from random forest models. These features were then selected and associated with the orthogroups shared among the genus datasets. This combined information was considered to search for the most important genes related to plant association lifestyle independently of the genus. In total, 13,821 orthogroups showed great importance in at least one of the conditions, of which 30% did not have importance, while plant-associated genomes accumulated approximately 49%, and non-plant-associated genomes had 21%.

To identify which functions may be more prominent in plant-associated genomes, we performed an enrichment of the function profile based on Clusters of Orthologous Group (COG) and considered only features with varImp > 1 and which did not present inconsistent annotation, this means that the orthogroups were derived only from plant-associated genomes or were specific from non-plant-associated genomes. We had 1,883 orthogroups related to plant-associated (PA or RA) and 664 to non-plant-associated (NPA or SA).

The Figure 3 and Figure 4 display the differences in the abundance of orthogroups assigned to COG according to lifestyle. A total of 5 categories were significantly abundant for plant-associated genomes ($P < 0.05$): “Amino acid transport and metabolism” (E), “Lipid transport and metabolism” (I), “Secondary metabolites biosynthesis, transport and catabolism” (Q), “Transcription” (K), and “Signal transduction mechanisms” (T). Whereas the non-plant-associated genomes present enrichment for the following COG categories: “Nucleotide transport and metabolism” (F), “Replication, recombination and repair” (L), “Cell motility” (N), “Function unknown” (S) and “Intracellular trafficking, secretion, and vesicular transport” (Figure 3.a and 3.b). In the comparison of COG functional category abundance between root-associated and soil-associated genomes, only “Replication, recombination and repair” (L) category showed significant enrichment considering the gene copy number dataset for soil bacterial genomes (Figure 4.a).

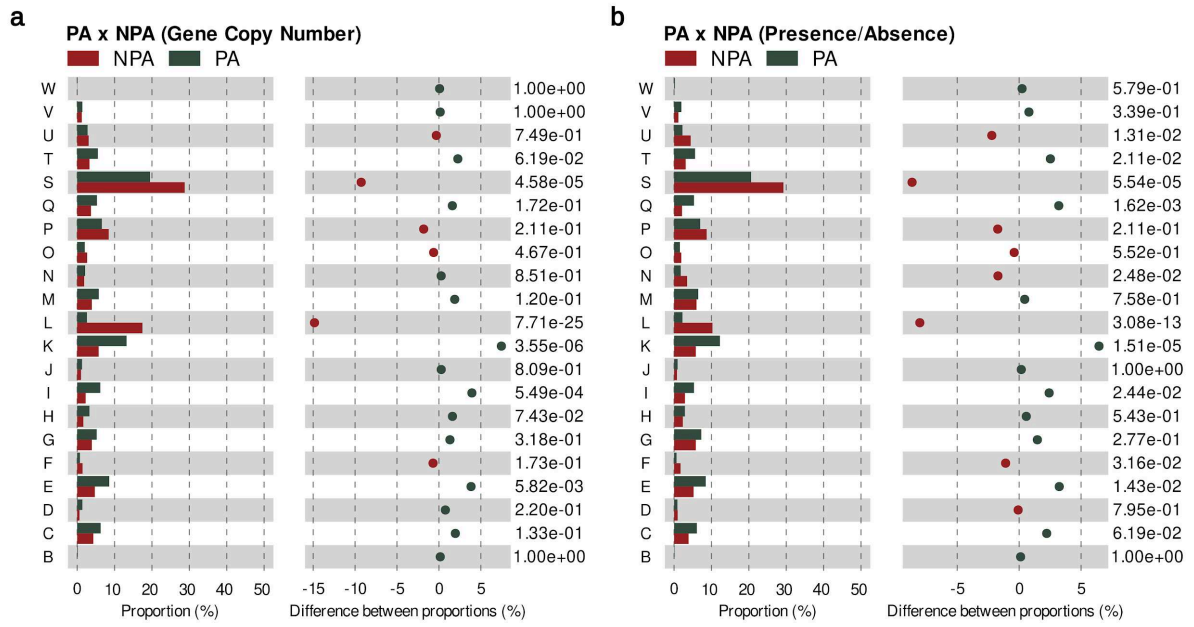

Figure 3. Differential abundance of orthogroups shared among the genera (varImp > 1), assigned COGs functional categories in plant-associated and non-plant-associated.

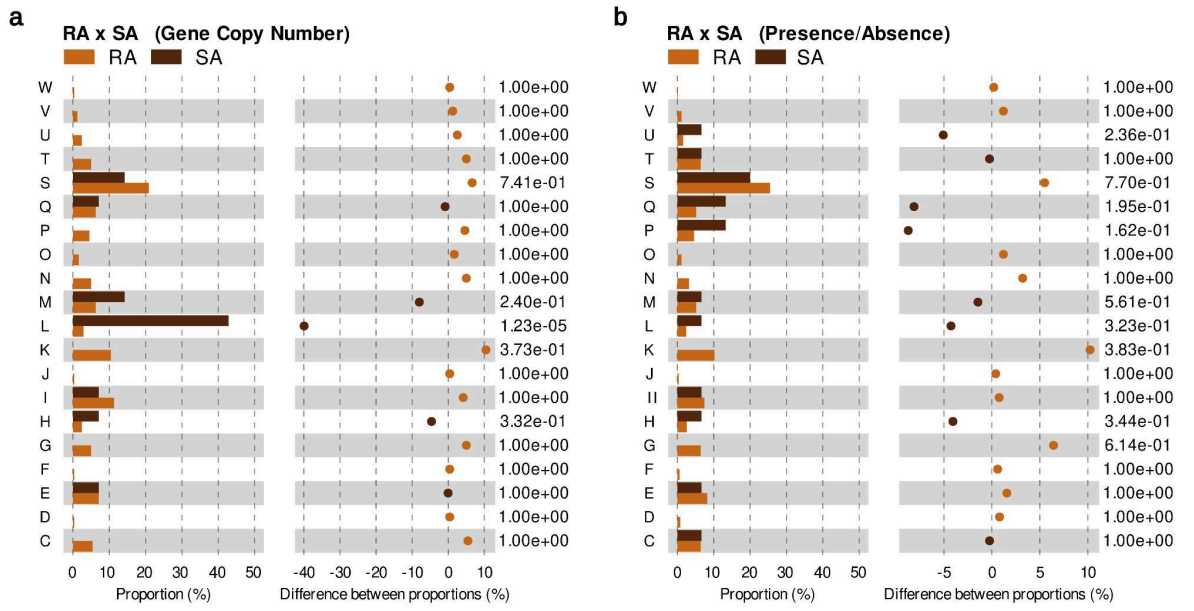

Figure 4. Differential abundance of orthogroups shared among the genera (varImp > 1), assigned COGs functional categories in root-associated and soil-associated genomes.

4. DISCUSSION

This study aimed to identify genomic features, specifically genes organized in orthogroups, that are associated with plant-related environments. We adopted machine learning approaches, hypothesizing that these features would recur across the genomes of various taxonomic groups of bacteria known to inhabit such environments. Our approach seeks to distinguish between the genomic characteristics of bacteria associated with plants and those not directly related to plant environments, reflecting the unique adaptations of plant-associated bacterial groups.

Our results indicated that there was either no detection or minimal detection of orthogroup enrichment related to bacterial genera *Achromobacter*, *Chryseobacterium*, *Citrobacter*, *Pantoea*, *Pseudoxanthomonas*, and *Sphingobium*.

However, these members have been reported for their potential in agriculture as plant growth-promoting and/or biological control agents (Abdel-Rahman et al. 2017, Bhise et al. 2017, Singh et al. 2021, Ajmal et al. 2022, Huda et al. 2022). On the other hand, a large number of shared orthogroups between *Burkholderia* and *Rhizobium* genera can indicate function similarities and are often associated with beneficial plant association (Díaz-Valle et al. 2019, Cui et al. 2020, Heo et al. 2022).

Certainly, the stringent criteria to consider valid alignments, influenced the obtention of shared orthogroups between genera, generating a large number of unique orthogroups. The intention was to ensure that more concise orthologous groups were obtained. Enrichment analysis approaches have been commonly used to identify trends of over-represented features in large-scale biological datasets (Cai et al. 2018, Levy et al. 2018, de Souza et al. 2019), contributing to our understanding of biological patterns.

Machine learning has been commonly used to predict environmental or host phenotypes, classify microbial features, study interactions between microbiome components, and monitor changes in microbiome composition (Hernández Medina et al. 2022). This has improved the efficiency of discovering important functional features of microorganisms associated with plants, including predicting plant-associated bacterial lifestyles and interaction factors (Martínez-García et al.

2016, Biggs et al. 2021, te Molder et al. 2021, Poncheewin et al. 2022). In our study, we used logistic regression and random forest techniques, and both proved to be very useful in identifying discriminant features even with a very small amount of data (representative genomes for each class, i.e. lifestyle). The random forest performed slightly better based on AUC values in the analyses of our small datasets. Taking all of this into consideration, we acknowledge that there is still the possibility to obtain even more information on this topic using our approach by adjusting parameters and thresholds, such as the alignment stringency for organizing orthogroups. Another aspect of extreme relevance to the success of the approach is obtaining a greater quantity of representative genomes.

Overall, our results using these approaches provide valuable insights into functional categories associated with the ability to colonize environments associated with plants, particularly focusing on plant-associated (PA) and root-associated (RA) enriched categories. These findings shed light on the molecular mechanisms underlying the interactions between microorganisms and plants, highlighting their significance for plant health and growth.

For the functional category termed “Amino acids transport and metabolism” (E) some studies reported as of strong importance in plant-microbe interactions (Cai et al. 2018, Cheng et al. 2019, Fabian et al. 2021). De Souza et al. (2019) reported that amino acid transporters are related to plant–microbe interaction and might be linked to the robust colonization lifestyle.

Another category overrepresented is the “Secondary metabolites biosynthesis, transport and catabolism” (Q), which includes biosynthesis of non-ribosomal peptides (NPR), siderophores, and signaling molecules (autoinducers) involved in key steps of the quorum-sensing process (Appendix E). Bacterial secondary metabolites serve as a significant reservoir of antimicrobials and other bioactive compounds (Chevrette et al. 2022). These compounds play crucial roles in microbial interactions by mediating rhizosphere competition (Andrić et al. 2022, Bhat et al. 2022) and responding to environmental changes (Kim et al. 2011). The production of these metabolites is often regulated, in part, through processes such as quorum sensing, emphasizing their role in bacterial communication and adaptation.

In the category 'Lipid transport and metabolism' (I), which encompasses processes related to the breakdown of fatty acids, lipid synthesis, and transport, we identified genes associated with the biosynthesis of phytoene, an isoprenoid that serves as a precursor for various carotenoids. Additionally, we found genes encoding fatty acid desaturase enzymes, which play a crucial role in the synthesis of unsaturated fatty acids. Lipides play several roles in plant-microbe association, some of their key functions include stress tolerance (Rojas-Solis et al. 2020), inducing systemic resistance: unsaturated fatty acids produced by PGPB can trigger systemic resistance in plants (Macabuhay et al. 2022), biofilm formation: lipids are components in a biofilm matrix, and generally act as biosurfactant providing dispersal and bioavailability of hydrophobic substances, and bacterial attachment, poorly explored in plant-microbe association (Bogino et al. 2013).

The COGs belonging to “Signal transduction mechanisms” (T) are primarily associated with sensor proteins that participate in various cellular processes, including chemotaxis proteins.

The “Transcription (K)” category can be related to the regulation of quorum sensing and motility. Additionally, this category includes proteins with domains for binding to a variety of ligands, such as a Solute-binding protein (SBP) predicted to bind specifically only to amino acids (SBP_bac_3) (Ortega et al. 2022). "The LysR, a transcriptional regulator (TR) found within this category, represents the most abundant type of TR in prokaryotes. It mediates a cell's response to changes in its environment by regulating a diverse set of genes, including those involved in virulence, metabolism, quorum sensing, and motility (Maddocks and Oyston 2008).

5. CONCLUSION

In our work, using the available datasets, we observed that random forest (RF) slightly outperformed logistic regression, particularly in terms of Area Under the ROC Curve. However, further efforts are needed, such as the algorithm parameters tuning, and thresholds reviewing, to improve the accuracy of our classifiers’ models. In addition, we also need more representative genomes to increase the number of instances in genus-specific datasets and to avoid the 'curse of dimensionality,'

enabling us to extract more informative features. Even so, overall, this study provides a significant set of functional categories related to plant-associated bacterial genomes, shedding light on genomic features that reveal bacterial adaptations to inhabit plants. These categories include “Amino Acid Transport and Metabolism (E)”, “Lipid Transport and Metabolism (I)”, “Secondary Metabolites Biosynthesis, Transport and Catabolism (Q)”, “Transcription (K)”, and “Signal Transduction Mechanisms (T)”. A focused effort on the specific characterization of the results for each genus is crucial to further extend our knowledge about the specific strategies and mechanisms used for each genus to interact with plant hosts.

6. REFERENCES

Abdel-Rahman HM, Salem AA, Moustafa MMA, El-Garhy HAS (2017) A novice *Achromobacter* sp. EMCC1936 strain acts as a plant-growth-promoting agent. **Acta Physiologiae Plantarum** 39: 61. https://doi.org/10.1007/s11738-017-2360-6

Ajmal AW, Yasmin H, Hassan MN, Khan N, Jan BL, Mumtaz S (2022) Heavy Metal–Resistant Plant Growth–Promoting *Citrobacter werkmanii* Strain WWN1 and *Enterobacter cloacae* Strain JWM6 Enhance Wheat (*Triticum aestivum* L.) Growth by Modulating Physiological Attributes and Some Key Antioxidants Under Multi-Metal Stress. **Frontiers in Microbiology** 13: 815704. https://doi.org/10.3389/fmicb.2022.815704

Andrić S, Rigolet A, Argüelles Arias A, Steels S, Hoff G, Balleux G, Ongena L, Höfte M, Meyer T, Ongena M (2022) Plant-associated *Bacillus* mobilizes its secondary metabolites upon perception of the siderophore pyochelin produced by a *Pseudomonas* competitor. **The ISME Journal.** https://doi.org/10.1038/s41396-022-01337-1

Barrett T, Clark K, Gevorgyan R, Gorelenkov V, Gribov E, Karsch-Mizrachi I, Kimelman M, Pruitt KD, Resenchuk S, Tatusova T, Yaschenko E, Ostell J (2012) BioProject and BioSample databases at NCBI: facilitating capture and organization of metadata. **Nucleic Acids Research** 40: D57–D63. https://doi.org/10.1093/nar/gkr1163

Benton B, King S, Greenfield SR, Puthuveetil N, Reese AL, Duncan J, Marlow R, Tabron C, Pierola AE, Yarmosh Jr DA, others (2021) The ATCC Genome Portal: microbial genome reference standards with data provenance. **Microbiology Resource Announcements** 10: e00818-21.

Bhat BA, Tariq L, Nissar S, Islam ST, Islam SU, Mangral Z, Ilyas N, Sayyed RZ, Muthusamy G, Kim W, Dar TUH (2022) The role of plant‐associated rhizobacteria in plant growth, biocontrol and abiotic stress management. **Journal of Applied Microbiology** 133: 2717–2741. https://doi.org/10.1111/jam.15796

Bhise KK, Bhagwat PK, Dandge PB (2017) Synergistic effect of *Chryseobacterium gleum* sp. SUK with ACC deaminase activity in alleviation of salt stress and plant growth promotion in *Triticum aestivum* L. **3 Biotech** 7: 105. https://doi.org/10.1007/s13205-017-0739-0

Biggs MB, Craig K, Gachango E, Ingham D, Twizeyimana M (2021) Genomics- and Machine Learning-Accelerated Discovery of Biocontrol Bacteria. **Phytobiomes Journal** 5: 452–463. https://doi.org/10.1094/PBIOMES-01-21-0003-R

Bogino P, Oliva M, Sorroche F, Giordano W (2013) The Role of Bacterial Biofilms and Surface Components in Plant-Bacterial Associations. **International Journal of Molecular Sciences** 14: 15838–15859. https://doi.org/10.3390/ijms140815838

Breiman L (2001) Random Forests. Machine Learning 45: 5–32. https://doi.org/10.1023/A:1010933404324

Brynildsrud O, Bohlin J, Scheffer L, Eldholm V (2016) Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. **Genome Biology** 17: 238. https://doi.org/10.1186/s13059-016-1108-8

Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. **Nature Methods** 12: 59–60. https://doi.org/10.1038/nmeth.3176

Cai H, Bai Y, Guo C (2018) Comparative genomics of 151 plant-associated bacteria reveal putative mechanisms underlying specific interactions between bacteria and plant hosts. **Genes & Genomics** 40: 857–864. https://doi.org/10.1007/s13258-018-0693-1

Cheng YT, Zhang L, He SY (2019) Plant-Microbe Interactions Facing Environmental Challenge. **Cell Host & Microbe** 26: 183–192. https://doi.org/10.1016/j.chom.2019.07.009

Chevrette MG, Thomas CS, Hurley A, Rosario-Meléndez N, Sankaran K, Tu Y, Hall A, Magesh S, Handelsman J (2022) Microbiome composition modulates secondary metabolism in a multispecies bacterial community. **Proceedings of the National Academy of Sciences** 119: e2212930119. https://doi.org/10.1073/pnas.2212930119

Cui G, Yin K, Lin N, Liang M, Huang C, Chang C, Xi P, Deng Y (2020) *Burkholderia gladioli* CGB10: A Novel Strain Biocontrolling the Sugarcane Smut Disease. **Microorganisms** 8: 1943. https://doi.org/10.3390/microorganisms8121943

Díaz-Valle A, López-Calleja AC, Alvarez-Venegas R (2019) Enhancement of Pathogen Resistance in Common Bean Plants by Inoculation With Rhizobium etli. *Frontiers in Plant Science* 10: 1317. https://doi.org/10.3389/fpls.2019.01317

Emms DM, Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. **Genome biology** 20: 1–14.

Fabian BK, Foster C, Asher AJ, Elbourne LDH, Cain AK, Hassan KA, Tetu SG, Paulsen IT (2021) Elucidating Essential Genes in Plant-Associated *Pseudomonas protegens* Pf-5 Using Transposon Insertion Sequencing. Galperin MY (Ed.). **Journal of Bacteriology** 203. https://doi.org/10.1128/JB.00432-20

Friedman J, Hastie T, Tibshirani R, Narasimhan B, Tay K, Simon N, Qian J (2021) Package 'glmnet.' CRAN R Repositary.

Hassani MA, Durán P, Hacquard S (2018) Microbial interactions within the plant holobiont. **Microbiome** 6: 58. https://doi.org/10.1186/s40168-018-0445-0

Hassani MA, Özkurt E, Seybold H, Dagan T, Stukenbrock EH (2019) Interactions and Coadaptation in Plant Metaorganisms. **Annual Review of Phytopathology** 57: 483–503. https://doi.org/10.1146/annurev-phyto-082718-100008

Heo AY, Koo YM, Choi HW (2022) Biological Control Activity of Plant Growth Promoting Rhizobacteria *Burkholderia contaminans* AY001 against Tomato Fusarium Wilt and Bacterial Speck Diseases. **Biology** 11: 619. https://doi.org/10.3390/biology11040619

Hernández Medina R, Kutuzova S, Nielsen KN, Johansen J, Hansen LH, Nielsen

M, Rasmussen S (2022) Machine learning and deep learning applications in microbiome research. **ISME Communications** 2: 98. https://doi.org/10.1038/s43705-022-00182-9

Huda N ul, Tanvir R, Badar J, Ali I, Rehman Y (2022) Arsenic-Resistant Plant Growth Promoting *Pseudoxanthomonas mexicana* S254 and *Stenotrophomonas maltophilia* S255 Isolated from Agriculture Soil Contaminated by Industrial Effluent. **Sustainability** 14: 10697. https://doi.org/10.3390/su141710697

Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P (2017) Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. **Molecular Biology and Evolution** 34: 2115–2122. https://doi.org/10.1093/molbev/msx148

Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P (2019) eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. **Nucleic Acids Research** 47: D309–D314. https://doi.org/10.1093/nar/gky1085

Ives AR, Garland T (2010) Phylogenetic Logistic Regression for Binary Dependent Variables. **Systematic Biology** 59: 9–26. https://doi.org/10.1093/sysbio/syp074

Kim YC, Leveau J, McSpadden Gardener BB, Pierson EA, Pierson LS, Ryu C-M (2011) The Multifactorial Basis for Plant Health Promotion by Plant-Associated Bacteria. **Applied and Environmental Microbiology** 77: 1548–1555. https://doi.org/10.1128/AEM.01867-10

Kuhn M (2008) Building predictive models in R using the caret package. **Journal of statistical software** 28: 1–26.

Levy A, Salas Gonzalez I, Mittelviefhaus M, Clingenpeel S, Herrera Paredes S, Miao J, Wang K, Devescovi G, Stillman K, Monteiro F et al. (2018) Genomic features of bacterial adaptation to plants. **Nature Genetics** 50: 138–150. https://doi.org/10.1038/s41588-017-0012-9

Liaw A, Wiener M (2002) Classification and Regression by randomForest. **R News** 2: 18–22.

Macabuhay A, Arsova B, Walker R, Johnson A, Watt M, Roessner U (2022) Modulators or facilitators? Roles of lipids in plant root–microbe interactions. **Trends in Plant Science** 27: 180–190. https://doi.org/10.1016/j.tplants.2021.08.004

Maddocks SE, Oyston PCF (2008) Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. **Microbiology** 154: 3609–3623. https://doi.org/10.1099/mic.0.2008/022772-0

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM (2021) BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Kelley J (Ed.). **Molecular Biology and Evolution** 38: 4647–4654. https://doi.org/10.1093/molbev/msab199

Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC (2012) IMG: the integrated microbial genomes database and comparative analysis system. **Nucleic Acids Research** 40: D115–D122. https://doi.org/10.1093/nar/gkr1044

Martínez-García PM, López-Solanilla E, Ramos C, Rodríguez-Palenzuela P (2016) Prediction of bacterial associations with plants using a supervised machine-learning approach: Prediction of bacterial associations with plants. **Environmental Microbiology** 18: 4847–4861. https://doi.org/10.1111/1462-2920.13389

te Molder D, Poncheewin W, Schaap PJ, Koehorst JJ (2021) Machine learning approaches to predict the Plant-associated phenotype of *Xanthomonas* strains. **BMC Genomics** 22: 848. https://doi.org/10.1186/s12864-021-08093-0

Ortega Á, Matilla MA, Krell T (2022) The Repertoire of Solute-Binding Proteins of Model Bacteria Reveals Large Differences in Number, Type, and Ligand Range. Gao B (Ed.). **Microbiology Spectrum** 10: e02054-22. https://doi.org/10.1128/spectrum.02054-22

Poncheewin W, van Diepeningen AD, van der Lee TAJ, Suarez-Diez M, Schaap PJ (2022) Classification of the plant-associated lifestyle of *Pseudomonas* strains using genome properties and machine learning. **Scientific Reports** 12: 10857. https://doi.org/10.1038/s41598-022-14913-4

R Core Team (2021) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available from: https://www.R-project.org/.

Rojas-Solis D, Vences-Guzmán MA, Sohlenkamp C, Santoyo G (2020) Bacillus toyonensis COPE52 Modifies Lipid and Fatty Acid Composition, Exhibits Antifungal Activity, and Stimulates Growth of Tomato Plants Under Saline Conditions. **Current Microbiology** 77: 2735–2744. https://doi.org/10.1007/s00284-020-02069-1

Seemann T (2014) Prokka: rapid prokaryotic genome annotation. **Bioinformatics** 30: 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Singh P, Singh RK, Li H-B, Guo D-J, Sharma A, Lakshmanan P, Malviya MK, Song X-P, Solanki MK, Verma KK, Yang L-T, Li Y-R (2021) Diazotrophic Bacteria *Pantoea dispersa* and *Enterobacter asburiae* Promote Sugarcane Growth by Inducing Nitrogen Uptake and Defense-Related Gene Expression. **Frontiers in Microbiology** 11: 600417. https://doi.org/10.3389/fmicb.2020.600417

de Souza RSC, Armanhi JSL, Damasceno N de B, Imperial J, Arruda P (2019) Genome Sequences of a Plant Beneficial Synthetic Bacterial Community Reveal Genetic Features for Successful Plant Colonization. **Frontiers in Microbiology** 10: 1779. https://doi.org/10.3389/fmicb.2019.01779

Teheran-Sierra LG, Funnicelli MIG, de Carvalho LAL, Ferro MIT, Soares MA, Pinheiro DG (2021) Bacterial communities associated with sugarcane under different agricultural management exhibit a diversity of plant growth-promoting traits and evidence of synergistic effect. **Microbiological Research** 247: 126729. https://doi.org/10.1016/j.micres.2021.126729

Thrall PH, Hochberg ME, Burdon JJ, Bever JD (2007) Coevolution of symbiotic mutualists and parasites in a community context. **Trends in Ecology & Evolution** 22: 120–126. https://doi.org/10.1016/j.tree.2006.11.007

Trivedi P, Leach JE, Tringe SG, Sa T, Singh BK (2020) Plant–microbiome interactions: from community assembly to plant health. **Nature Reviews Microbiology** 18: 607–621. https://doi.org/10.1038/s41579-020-0412-1

Wu L, Sun Q, Desmeth P, Sugawara H, Xu Z, McCluskey K, Smith D, Alexander V, Lima N, Ohkuma M, Robert V, Zhou Y, Li J, Fan G, Ingsriswang S, Ozerskaya S, Ma J (2017) World data centre for microorganisms: an information infrastructure to explore and utilize preserved microbial strains worldwide. **Nucleic Acids Research** 45: D611–D618. https://doi.org/10.1093/nar/gkw903

Wu L, McCluskey K, Desmeth P, Liu S, Hideaki S, Yin Y, Moriya O, Itoh T, Kim CY, Lee J-S, et al. (2018) The global catalogue of microorganisms 10K type strain sequencing project: closing the genomic gaps for the validly published prokaryotic and fungi species. **GigaScience** 7. https://doi.org/10.1093/gigascience/giy026

Zhang N, Yang D, Kendall JRA, Borriss R, Druzhinina IS, Kubicek CP, Shen Q, Zhang R (2016) Comparative Genomic Analysis of Bacillus amyloliquefaciens and *Bacillus subtilis* Reveals Evolutional Traits for Adaptation to Plant-Associated Habitats. **Frontiers in Microbiology** 7. https://doi.org/10.3389/fmicb.2016.02039

APPENDIX

The appendices for this chapter are available on Zenodo, an open-access general-purpose repository maintained by OpenAIRE and CERN. This repository allows the deposit of research articles, datasets, research software, reports, and any other digital artifacts related to research.

Access to the appendices for this chapter can be found at the following permanent link: https://doi.org/10.5281/zenodo.10447438

Appendix A- List of all bacterial genomes used in orthologous genes clustering in the feature extraction step and in the further steps to build and test classifiers' models. The list includes the isolation source information and the related category for the genome classification and features selection purposes.

Appendix B - Distribution of genomes by phylum, family, and genus among the categories defined according to bacteria lifestyle association.

Appendix C- Enriched orthogroups by genus according to each enrichment test (Material and Methods). Values for each test are "Y" (enriched), "N" (not enriched), or "Untested" (clusters were untested when there was insufficient phylogenetic signal, they were too small or were found in all genomes).

Appendix D- Classification performance of random forest and logistic regression techniques applied to genus-specific datasets of genomic features (orthogroups) using both matrices from gene count number and presence/absence values. Sensitivity is a measure of how well a test identifies true positives; Specificity: is a measure of how well a test or model avoids false positives; Positive Predictive Value (Pos. Pred. Value): The probability that a positive prediction is correct; Negative Predictive Value (Neg. Pred. Value): The probability that a negative prediction is correct; Precision: The accuracy of positive predictions; Recall (Sensitivity): The ability to find all relevant cases; F1 Score: A combined measure of precision and recall; Prevalence: The proportion of positive cases in the total; Detection Rate: The proportion of true positive cases identified; Detection Prevalence: The proportion of positive predictions; Balanced Accuracy: An average of sensitivity and specificity;

Area Under the Curve (AUC): The overall performance of the model in distinguishing between positive and negative cases.

Appendix E- Orthogroups assigned with predicted COGs as an important feature for classifying plant-associated genomes. COG categories: A - RNA processing and modification; B - Chromatin structure and dynamics; C - Energy production and conversion; D - Cell cycle control, cell division, chromosome partitioning; E - Amino acid transport and metabolism; F - Nucleotide transport and metabolism; G - Carbohydrate transport and metabolism; H - Coenzyme transport and metabolism; I - Lipid transport and metabolism; J - Translation, ribosomal structure and biogenesis; K - Transcription; L - Replication, recombination and repair; M - Cell wall/membrane/envelope biogenesis; N - Cell motility; O - Posttranslational modification, protein turnover, chaperones; P - Inorganic ion transport and metabolism; Q - Secondary metabolites biosynthesis, transport and catabolism; R - General function prediction only; S - Function unknown; T - Signal transduction mechanisms; U - Intracellular trafficking, secretion, and vesicular transport; V - Defense mechanisms; W - Extracellular structures; X - Mobilome: prophages, transposons; Y - Nuclear structure; Z - Cytoskeleton.