OSIEL SILVA GONÇALVES

THE KNOWN UNKNOWNS: UNDERSTANDING SLOW-GROWING BACTERIA AND THEIR PLANT INTERACTION THROUGH REVERSE ECOLOGY APPROACHES

Thesis submitted to the Agricultural Microbiology Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Doctor Scientiae*.

Adviser: Mateus Ferreira Santana

Co-advisers: Wendel Batista da Silveira
Maurício Dutra Costa

VIÇOSA - MINAS GERAIS
2023

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus

T
G635k
2023

Gonçalves, Osiel Silva, 1994-
The known unknowns: understanding slow-growing bacteria and their plant interaction through reverse ecology approaches / Osiel Silva Gonçalves. - Viçosa, MG, 2023.
1 tese eletrônica (156 f.): il. (algumas color.).
Texto em inglês
.
Inclui apêndices.
Orientador: Mateus Ferreira Santana
Tese (doutorado) - Universidade Federal de Viçosa, Departamento de Microbiologia, 2023.
Inclui bibliografia.
DOI: https://doi.org/10.47328/ufvbbt.2023.451
Modo de acesso: World Wide Web.

1. Bacteriologia - Cultura e meios de cultura; 2. Microbiologia; 3. Ecossistemas; 4. Genômica; 5. Agricultura; I. Santana, Mateus Ferreira II. Universidade Federal de Viçosa. Departamento de Microbiologia. Programa de Pós-Graduação em Microbiologia Agrícola III. Título

CDD 22. ed. 571.638293

Bibliotecário(a) responsável: BRUNA SILVA CRB-6/2552

OSIEL SILVA GONÇALVES

THE KNOWN UNKNOWNS: UNDERSTANDING SLOW-GROWING BACTERIA AND THEIR PLANT INTERACTION THROUGH REVERSE ECOLOGY APPROACHES

Thesis submitted to the Agricultural Microbiology Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Doctor Scientiae*.

APPROVED: July 18, 2023.

Assent:

Documento assinado digitalmente
gov.br
OSIEL SILVA GONCALVES
Data: 25/07/2023 11:22:37-0300
Verifique em https://validar.iti.gov.br

_____

Osiel Silva Gonçalves
Author

Documento assinado digitalmente
gov.br
MATEUS FERREIRA SANTANA
Data: 25/07/2023 11:34:56-0300
Verifique em https://validar.iti.gov.br

_____

Mateus Ferreira Santana
Adviser

Dedico essa obra àqueles que desde o princípio acreditaram que ela seria possível, minha tão amada mãe Selma, meu querido pai Carlos, meu inseparável irmão Carlos (Dudu). A vocês e por vocês.

ACKNOWLEDGEMENTS

To God and to Mikisi, for faith, strength, dedication, and courage to embark on this journey.

To my family. My mother, Selma, for being and continuing to be my strength, comfort, and for letting go of her hand so her hummingbird could fly and conquer new horizons. Despite the distance, she was present in every moment. My dear father, Carlos, who always supported my crazy ideas and never hesitated to do everything possible for my education and that of my brother, my example of honesty and a true man. And to my youngest brother, Carlos (Dudu), hard to put into words, but represents a pure feeling of goodness, love, my pride!

To Ralph, for the companionship and friendship throughout this journey, strengthening me and believing in every step I took.

To Professor Mateus Ferreira Santana, for the trust placed in me to carry out this work, dedication, and commitment to share his knowledge, and for his effective participation in this important stage of my professional development.

To my illustrious and eternal friends from LGMM. You made the weight on my heart lighter. You made every second in the laboratory worthwhile. Alexia, Bruna, Felipe, Luciano, and Mirele, Janaína, Jéssica, Giarlã, and Patrícia. You will always be the best.

This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

To the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

To the Fundação de Amparo à Pesquisa do Estado de Minas Gerais—FAPEMIG (Process 402644/2021-2) for the financial support.

To the Universidade Federal de Viçosa (UFV) for the opportunity to pursue this course.

To the administrative coordination of the Graduate Program in Agricultural Microbiology.

To the technical-administrative staff of the Instituto de Biotecnologia Aplicada à Agropecuária (BIOAGRO).

To all of you, my most heartfelt thanks!

ABSTRACT

GONÇALVES, Osiel Silva, D.Sc., Universidade Federal de Viçosa, July, 2023. **The known unknowns: understanding slow-growing bacteria and their plant interaction through reverse ecology approaches**. Adviser: Mateus Ferreira Santana. Co-advisers: Wendel Batista da Silveira and Maurício Dutra Costa.

The cultivation of bacteria that exhibit slow growth rates has posed a longstanding challenge in microbiology, resulting in a significant number of unculturable bacterial species. This phenomenon, known as the "great plate count anomaly", represents one of the oldest unresolved topics in the field. Recent advancements in cultivation techniques have shown promise in overcoming this challenge. In addition, the study of slow-growing bacteria and their interactions with plants has gained significant attention due to their ecological importance and potential applications in agriculture. Genomics techniques have provided valuable insights into the genetic characteristics underlying microbe-plant interactions. In this context, by examining the in-cultivation techniques, genomics, and computational modeling, we seek to shed light on the slow-growing bacteria and uncover their contributions to ecosystem functioning and plant interaction. Our initial understanding was obtained through the study of 92 slow-growing bacteria isolated from the Brazilian Cerrado soil during a four week-long isolation period. These bacteria can thrive in low-water conditions, promote plant growth, and belong to a novel species group. Genome analysis of five strains revealed their potential in biogeochemical cycles, plant growth promotion, and biosynthesis of secondary metabolites. Next, we conducted greenhouse experiments to assess the efficacy of these bacteria in soybean cultivation, employing a carefully designed bacterial consortium based on our comprehensive understanding of microbial-microbe and plant-microbe interactions. The results showed promising outcomes for improving soybean productivity. In the third chapter of this thesis, we employed *in-silico* modeling to design a synthetic microbial community aimed at enhancing the yield of important crop plants. By selecting six hub species with essential plant growth-promoting traits from the dominant plant species found in the Campos rupestres, we aimed to optimize the plant-microbe interactions and maximize crop productivity. Lastly, our analysis of 758 metagenome-assembled genomes shed light on the global distribution of the Acidobacteriota phylum and its interactions with plants and biogeochemical processes. This exploration revealed distinct ecological roles for individual taxonomic groups within this phylum, providing valuable insights for future research. Overall, our thesis contributes to a better understanding of slow-

growing bacteria, their ecological significance, and their potential applications in agriculture, offering insights into ecosystem functioning and plant interactions.

Keywords: Agriculture. Slow-growing bacteria. Ecosystems. Genomics. Microbe-plant interactions. Microbiology

RESUMO

GONÇALVES, Osiel Silva, D.Sc., Universidade Federal de Viçosa, julho de 2023. **Desconhecidos conhecidos: compreendendo as bactérias de crescimento lento e suas interações com as plantas através de abordagens de ecologia reversa**. Orientador: Mateus Ferreira Santana. Coorientadores: Wendel Batista da Silveira e Maurício Dutra Costa.

O cultivo de bactérias de crescimento lento tem sido um desafio persistente na área da microbiologia, resultando em muitas espécies bacterianas que não podem ser cultivadas. Esse fenômeno, conhecido como "a grande anomalia na contagem de placas", é um dos problemas mais antigos não resolvidos na microbiologia. Recentemente, foram feitos avanços nas técnicas de cultivo que têm mostrado promessa para superar esse desafio. Além disso, o estudo das bactérias de crescimento lento e suas interações com as plantas tem recebido atenção significativa devido à sua importância ecológica e às possíveis aplicações na agricultura. A genômica tem fornecido informações valiosas sobre as características genéticas que estão por trás dessas interações entre microrganismos e plantas. Nesse contexto, examinamos as técnicas de cultivo, a genômica e a modelagem computacional para esclarecer as bactérias de crescimento lento e descobrir como elas contribuem para o funcionamento dos ecossistemas e para a interação com as plantas. Nosso primeiro entendimento foi obtido por meio do estudo de 92 bactérias de crescimento lento isoladas do solo do Cerrado. Essas bactérias têm a capacidade de se desenvolver em condições com pouca água, características promissoras para promover o crescimento das plantas, e pertencem a um novo grupo de espécies. A análise do genoma de cinco linhagens revelou seu potencial em ciclos biogeoquímicos, promoção do crescimento das plantas e biossíntese de metabólitos secundários. Em seguida, realizamos experimentos em casa de vegetação para avaliar a eficácia dessas bactérias no cultivo da soja, utilizando um consórcio bacteriano cuidadosamente projetado, com base nas interações microbianas e com as plantas. Os resultados se mostraram promissores para melhorar a produtividade da soja. No terceiro capítulo da tese, utilizamos a modelagem *in-silico* para projetar uma comunidade microbiana sintética com o objetivo de aumentar o rendimento de importantes plantas agrícolas. Selecionamos seis espécies centrais com características essenciais para a promoção do crescimento das plantas, a partir das espécies dominantes encontradas nos Campos rupestres, buscando otimizar as interações entre micróbios e plantas e maximizar a produtividade de importantes culturas agrícolas. Por fim, nossa análise de 758 genomas montados a partir de metagenomas trouxe novas informações sobre a distribuição global do filo Acidobacteriota e

suas interações com plantas e processos biogeoquímicos. Essa exploração revelou papéis ecológicos distintos para grupos taxonômicos individuais dentro desse filo, fornecendo informações valiosas para pesquisas futuras. Em resumo, essa tese contribui para uma melhor compreensão das bactérias de crescimento lento, sua importância ecológica e suas possíveis aplicações na agricultura. Além disso, ela oferece informações valiosas sobre o funcionamento dos ecossistemas e as interações entre micróbios e plantas.

Palavras-chave: Agricultura. Bactérias de crescimento lento. Ecossistemas. Genômica. Interações micróbio-planta. Microbiologia.

CONTENT

General introduction

The microbial world is a vast and intricate web of life, where countless organisms coexist and interact in complex ways. Among these microorganisms, bacteria reign the most numerous and diverse biological organisms on the planet. However, despite their ubiquity, a significant portion of bacterial species remains unknown, defying attempts to cultivate them in laboratory settings. This enigmatic phenomenon, known as the "great plate count anomaly", has puzzled microbiologists for decades, representing one of the oldest unresolved topics in the field.

In recent years, remarkable progress in tackling this challenge by employing innovative strategies to cultivate previously 'unculturable' bacteria. By utilizing approaches such as oligotrophic media, extended incubation periods, and the selective isolation of slow-growing bacteria, the potential of these elusive microorganisms has been improved. This advancement not only presents a significant breakthrough in microbiology but also opens doors to understanding the vast unknown biological diversity that exists on our planet.

The studies of bacteria's interactions with plants have gained substantial attention due to their crucial role in shaping ecosystems and influencing plant health. Exploring the genetic characteristics and ecological traits of plant growth-promoting bacteria has been made possible through the application of genomics techniques. A valuable insight into the intricate dynamics of microbe-plant interactions has been done by analyzing the genetic makeup of these bacteria and deciphering their metabolic capabilities.

Reverse ecology offers a unique opportunity to unravel the genetic traits underlying slow-growing bacteria and their interactions with plants. Reverse ecology is a frontier in evolutionary systems biology that aims to obtain novel insights into an organism's ecology. This approach uses advances in systems biology and genomic metabolic modeling to predict the ecological traits of microbes and their interactions with other organisms. By examining the in-cultivation techniques, genomics, and computational modeling, we seek to shed light on the hidden world of these elusive microorganisms and uncover their vital contributions to ecosystem functioning and enhance plant growth.

This thesis includes three appendix articles that significantly contribute to the understanding of various aspects within the scope of this study. In the first article, we investigated the essential characteristics of the Acidobacteriota phylum, which carries multiple integrative and conjugative elements in the chromosome. These elements encode a diverse repertoire of genes, showcasing potential environmental functions such as heavy metal resistance, iron uptake, secondary metabolism, and antibiotic resistance. These advantageous

attributes play a pivotal role in the phylum's survival, resistance, and persistence in various environmental conditions. Moving on to the second appendix, we explore the application of reverse ecology as a powerful tool to reveal the potential metabolic interplay among *'Candidatus Liberibacter'* species, Citrus hosts, and psyllid vectors. This research sheds light on how to effectively employ reverse ecology techniques to uncover interactions between microorganisms and their hosts, providing valuable insights into managing such interplays. Finally, in the third appendix, we focus on the isolation of novel rhizobacteria species from soybean using an oligotrophic medium. We demonstrated that some of these isolates have potential to improve soybean growth under drought, and through genomic analysis, we have successfully demonstrated the presence of genes encoding plant growth-promoting proteins and secondary metabolite biosynthetic clusters in publicly available genomes for the ten genera studied. Together, these three articles substantially contribute to the knowledge and understanding of diverse microbial interactions and their implications in environmental, agricultural, and ecological contexts.

Chapter 1

Uncovering the Secrets of Slow-Growing Bacteria in Tropical Savanna Soil through Isolation and Genomic Analysis

Gonçalves, O.S, Santana, M.F. Uncovering the Secrets of Slow-Growing Bacteria in Tropical Savanna Soil through Isolation and Genomic Analysis. The manuscript was accepted for publication in the Microbial Ecology.

Uncovering the Secrets of Slow-Growing Bacteria in Tropical Savanna Soil through Isolation and Genomic Analysis

Osiel Silva Gonçalves[1], Mateus Ferreira Santana[1*]

[1]Grupo de Genômica Eco-evolutiva Microbiana, Laboratório de Genética Molecular de Microrganismos, Departamento de Microbiologia, Instituto de Biotecnologia Aplicada à Agropecuária, Universidade Federal de Viçosa, Minas Gerais, Brazil.

[*]Correspondence author: mateus.santana@ufv.br

Abstract

One gram of soil holds ten billion bacteria of thousands of different species, but most remain unknown, and one of the serious issues is intrinsic to slow-growing bacteria. Here, we report the week-long isolation of slow-growing bacteria from Brazilian Cerrado soil. During a four-week incubation, we selected a total of 92 isolates. These isolates, consisting mostly of slow-growing bacteria, could thrive in low-water conditions and possess features that promote plant growth. Our results indicate that slow-growing bacteria are genetically similar to established bacterial species but also belong to a novel group of species. The new strains identified were *Caballeronia* sp., *Neobacillus* sp., *Bradyrhizobium* sp., and high GC Gram-positive species. In addition, the growth experiment conducted under various culture media and temperature conditions revealed an extended lag phase for five strains. Genomic analysis of these five slow-growing bacteria showed their potential to participate in biogeochemical cycles, metabolize various carbohydrates, encode proteins with a role in promoting plant growth and have biosynthetic potential for secondary metabolites. Taken together, our findings reveal the untapped potential of slow-growing bacteria in tropical savanna soils.

Keywords: Cerrado; drought; plant growth-promoting bacteria; reverse ecology; uncultured bacteria.

Introduction

Bacteria, the most abundant living organisms on Earth, remain largely unexplored due to the "Great Plate Count Anomaly", which estimates that less than 1% of bacterial species have been cultivated in laboratory conditions [1]. Soil is by far the richest environment for microbial life;

it is estimated that in one gram, there are ten billion bacteria with thousands of different species, and despite extensive cultivation-independent study and advances in high-throughput sequencing technology, little is known about the majority of bacteria in soils [2]. Although the cultivation-dependent approach has some limitations, it is essential for understanding the diversity, ecology, and physiology of bacteria [3].

Many bacteria that are found in natural environments grow slowly and have been difficult to cultivate in laboratory conditions due to various factors, such as nutritional and ecological conditions [4, 5]. These bacteria are often referred to as “uncultured” or “previously uncultured” bacteria. Slow-growing bacteria are typically fastidious microorganisms, especially those that live in oligotrophic environments with slow growth rates [6]. They are present in various terrestrial and aquatic environments, and much of our current understanding of slow-growing bacteria relies on studies of human pathogenic bacteria.

In soil, slow-growing bacteria tend to be in the deeper layers of soil or are not closely associated with plant roots, where nutrients are limited, and environmental conditions are relatively stable [2]. The ability of these bacteria in the soil to adapt to changing environmental conditions is another essential characteristic. They grow slowly, and they are often more resistant to drought, low-nutrient environments, and other stressors that can be detrimental to other bacteria [7]. Altogether, they play an essential ecological role in biogeochemical cycles, soil fertility, the synthesis of new natural products, and interactions with other microbes, plants, and their surroundings [8]. However, little is known about the role of these bacteria in soil environments, especially in tropical forests.

The Cerrado is the world's most species-rich tropical savanna with an annual severe drought and is considered nutrient-poor, acidic, and aluminum-rich soil. This tropical savanna occupies one-quarter of Brazil. Interestingly, it is classified as a savanna with several physiognomies, morphoclimatic zones, and phytogeographic domains [9, 10], which are subject to variations in nutrient cycling, climatic season, and soil acidity inherent to the environment, and this reflects the assembly and structures of the microbial communities [11]. Metagenomic analyses have shown that slow-growing bacteria, such as those in the phylum Acidobacteriota, are highly abundant in Cerrado soil [12]. Therefore, this environment provides a unique opportunity to study microbes that have adapted to challenging environmental conditions.

Despite their importance, our understanding of slow-growing bacteria in soil is still limited. Much more research is needed to fully understand the functions and properties of these elusive microorganisms. Studies have shown the potential for the cultivation of previously "unculturable" bacteria from environmental samples using simple but laborious cultivation strategies[13–15]. The cultivation of "unculturable" bacteria can be improved by combining oligotrophic media, extended incubation periods, and the selection of slow-growing bacteria. Here, we reported the week-long isolation of 92 slow-growing bacteria belonging to nine genera from Cerrado sample soil and described their potential role in biogeochemical cycles and plant growth promotion potential abilities.

Materials and Methods

Soil sampling, preparation of media, and isolation procedures

Approximately 50 g of topsoil was obtained from five equidistant areas (approximately 1 meter apart) located in the Cerrado *sensu stricto* within a forest remnant in Rio Verde, Goiás, Brazil (17°47'10.8"S 50°57'59.7"W) (Fig. S1). The soil samples were then transported to the laboratory enclosed in polyethylene bags at ambient temperature and kept refrigerated at 4 °C for one week until processing. One gram of weighed sample of sieved soil was dispersed in 100 ml of sterile saline solution 0.85% after 1 ml aliquots were diluted in series. Aliquots of 100 µL of each dilution were inoculated in the VL55 medium base (composition in 1 L of distilled water: 3.9 g of MES (2-Morpholinoethanesulfonic acid), 20 mM $MgSO_4.7H_2O$, 30 mM $CaCl_2.2H_2O$, 20 mM $(NH_4)_2HPO_4$). This medium was autoclaved at 121 °C for 20 min and cooled to 56 °C. Next, 10 mL of 5% (w/v) xylan from beechwood (Fluka), 2 ml of tungstate/selenite solution, 2 mL of SL-10 trace element solution (see below), and nystatin (10 mg/L^{-1}) were supplemented in the VL55 medium base. The selenite/tungstate, vitamin, and trace element solutions were sterilized by filtration.

The tungstate/selenite solution contained (per liter of distilled water) 0.5 g of NaOH, 3 mg of $Na_2SeO_3.5H_2O$, and 4 mg of $Na_2WO_4.2H_2O$ [16]. SL-10 trace elements contained (per liter of distilled water) 10 ml of HCl [25%, 7.7 M], 1.5 g of $FeCl_2.4H_2O$, 70 mg of $ZnCl_2$, 100 mg of $MnCl_2.4H_2O$, 6 mg of H_3BO_3, 190 mg of $CoCl_2.6H_2O$, 2 mg of $CuCl_2.2H_2O$, 24 mg of $NiCl_2.6H_2O$ and 36 mg of $Na_2MoO_4.2H_2O$) [17]. The vitamin solution contained (per liter of H_2O) 2 mg (+)-biotin, 2 mg of folic acid, 10 mg of pyridoxamine hydrochloride, 5 mg of thiamine chloride, 5 mg of riboflavin, 5 mg of nicotinic acid, 5 mg of hemicalcium D-(+)-

pantothenate, 0.1 mg of cyanocobalamin, 5 mg of 4-aminobenzoate and 5 mg of DL-6,8-thiotic acid.

Plates with five replicates per sample were incubated in polyethylene bags to prevent desiccation at 25 °C in the dark for 15 weeks of incubation. To facilitate bacterial growth, the colonies were collected and transferred for purification on DMSZ *acidiphilium* medium agar (composition in 1 L of distilled water: 2 g of $(NH_4)_2SO_4$, 0.1 g of KCl, 0.5 g of K_2HPO_4, 0.5 g of $MgSO_4.7H_2O$, 0.3 g of yeast extract, 1 g of D-glucose, and 15 g of agar–agar). This step was taken to augment bacterial growth, as VL55, being an oligotrophic medium, may provide limited nutrients. The cells were grown in liquid DMSZ *acidiphilium* medium for cryostock preparation with stocks prepared with glycerol and maintained at -80 °C.

Plant growth-promoting bacterial trait analysis

Bacterial growth under reduced water availability

The slow-growing bacteria were grown in tryptic soy agar (TSA) medium (10%) with additional sorbitol at five different concentrations (0 g/L^{-1}; 85 g/L^{-1}, 285 g/L^{-1}, and 660 g/L^{-1}) to simulate water stress at 28 °C [18].

Exopolysaccharide (EPS) production

Slow-growing bacteria were inoculated onto 5 mm diameter paper discs disposed of in DSMZ *acidiphilium* medium agar. The production of EPS was checked by slime appearance and by mixing a portion of the mucoid substance in 2 mL of absolute ethanol, in which the formation of a precipitate indicated the presence of EPS [19].

Indole acetic acid (IAA) production

Aliquots of 100 µL of slow-growing bacteria were initially grown in 10 mL of TSA medium (10%) for 48 hours in the dark at 28 °C. Next, colonies were transferred to fresh plaque containing TSA medium (10%) supplemented with 5 mM L-tryptophan. Colonies were covered with a cellulose filter (0.45 µm pore size) and incubated in the dark at 28 °C. After 48 h, the membranes were washed in Salkowski reagent (50 mL of perchloric acid (35%) and 1 mL of $FeCl_3$ solution (0.5 M) for 30 min in the dark [20]. The appearance of pink to red color indicated IAA production.

Phosphate solubilization assay

Slow-growing bacteria were inoculated into tubes with 10 ml of tryptone soya broth (TSB) medium (10%) for one week at 28 °C and shaken on an orbital shaker at 180 rpm. Cells were collected by centrifugation and washed twice with 0.8% NaCl solution, and 20 µL of this suspension was spotted on the National Botanical Research Institute's phosphate growth medium (NBRIPM) containing 15 g of agar, 10 g of glucose, 5 g of $Ca_3(PO_4)^2$, 5 g of $MgCl_2 \cdot 6H_2O$, 0.25 g of $MgSO_4 \cdot 7H_2O$, 0.2 g of KCl, and 0.1 g of $(NH_4)_2SO_4$ per liter [21]. Plates were incubated for 15 days at 28 °C. Positive phosphate solubilization was confirmed by the appearance of clearing zones around the spot.

Siderophore production

Slow-growing bacteria were tested for their ability to produce siderophores in CAS medium [22]. Bacteria were collected from TBS, resuspended, and washed twice with phosphate-buffered saline pH 6.5. An aliquot of 20 µL of bacterial suspension was spotted on CAS agar plates. The production of siderophores was checked daily for color change from blue to red around each colony.

DNA Extraction and 16S rRNA Sequencing, Processing, and Analysis

Slow-growing bacteria were inoculated into tubes with 20 ml of liquid DSMZ medium for one week at 28 °C and shaken on an orbital shaker at 180 rpm. Cells were collected by centrifugation, and genomic DNA was extracted using the Wizard® Genomic DNA Purification Kit (Promega Corp., Madison WI, USA) as recommended by the manufacturer. DNA was checked for quality using a NanoDrop 2000 (Thermo Fisher Scientific, Waltham, MA, USA) and subjected to gel electrophoresis (0.8% agarose). The 16S rRNA genes were amplified using 27F and 1492R [23]. PCRs were adjusted to a volume of 25 µL containing 0.1 µM of each 16S rRNA primer, 25 ng of genomic DNA, 0.2 mM of each dNTP, and 1.25 U of Taq DNA polymerase. The regions were amplified under the following conditions: initial denaturation stage at 95 °C for 5 min, followed by 34 cycles at 94 °C for 50 sec, 60 °C for 1 min, 72 °C for 1.3 min, and a final extension at 72 °C for 8 minutes. The amplification products were analyzed by electrophoresis on a 1.5% agarose gel. The amplicon was sequenced in the ABI 3730xl System (Macrogen, Inc., Seoul, South Korea).

The sequences were trimmed and assembled using Geneious Prime 2022.0.1 (Biomatters, Inc., Boston, USA). Next, the sequences were compared to the National Center for Biotechnology Information using BLASTn [24] against the 16S ribosomal RNA sequences (Bacteria and Archaea) database. We retrieved the whole 16S rRNA sequences for each species' family taxa

and built an in-house database. These databases were aligned and subjected to phylogenetic tree inference using the neighbor-joining method in MAFFT version 7 (https://mafft.cbrc.jp/alignment/server /large.html) [26]. 16S rRNA sequences were then directly aligned against closely related strains in the reference sequence (RefSeq) database that were retrieved for phylogenetic analysis. Next, the sequences were aligned using the ClustalW algorithm [27] on MEGA11 [28]. The best-fit substitution model was calculated in MEGA11, and the phylogenetic trees were constructed using the maximum likelihood tree (1000 bootstrap replicates) and the substitution model General Time Reversible+gamma distribution with invariable sites (G+I).

DNA fingerprinting

For molecular characterization of the slow-growing bacteria, BOX regions were amplified by PCR using the BOXA1R primer (5'-CTACGGCAAGGCG ACGCTGACG-3 ') [29]. PCRs were adjusted to a volume of 25 µL containing 0.5 mM of the BOX primer, 50 ng of DNA, 250 µM of each dNTP, and 1.25 U of Taq DNA polymerase. The following setup was used: initial denaturation stage at 95 °C for 7 min, followed by 35 cycles at 95 °C for 3 s, 92 °C for 30 s, 50 °C for 1 min, 65 °C for 8 min, and a final extension at 65 °C for 8 minutes. The amplification products were analyzed by electrophoresis on a 1.5% agarose gel, the fingerprint patterns were analyzed using BioNumerics (Applied Maths Inc., Sint-Martens-Latem, Belgium) software, and the same pattern fingerprint was considered a clone.

Growth condition experiment

The growth capabilities of five bacterial strains were determined by subjecting them to diverse culture media and temperatures. Initially, the strains were streaked on R2A medium (composition in 1 L of distilled water: 0.5 g of yeast extract, 0.5 g of peptone, 0.5 g of casamino acids, 0.5 g of glucose, 0.5 g of soluble starch, 0.5 g of pyruvate, 0.3 g of K_2HPO_4, 0.05 g of $MgSO_4.7H_2O$, and 15 g of agar–agar). Subsequently, individual colonies were transferred to 50 mL tubes containing 20 mL of the following media: VL55, R2A, DMSZ *acidiphilium*, and nutrient broth (NB). The nutrient broth consisted of 0.3 g of beef extract, 0.5 g of peptone, and 0.5 g of NaCl per liter. For each strain, two tubes were prepared. These tubes were then incubated at four different temperatures: 20 °C, 25 °C, 28 °C, and 37 °C, without any agitation, over six days. We monitored the growth progress at regular intervals, initially every 12 hours and then, after three days, every 24 hours. To quantify the growth, we measured the absorbance

reading at 600 nm using the Multiskan™ GO Microplate Photometer 96-well plate reader (Thermo Scientific™, Waltham, MA, USA).

DNA Extraction and Whole-Genome Sequencing

Five slow-growing bacteria were grown in a flask containing 20 mL of liquid DSMZ medium for one week at 28 °C and shaken on an orbital shaker at 180 rpm. Cells were collected by centrifugation, and genomic DNA was extracted using the Wizard® Genomic DNA Purification Kit (Promega Corp., Madison, WI, USA), as recommended by the manufacturer. DNA was checked for quality using a NanoDrop 2000 (Thermo Fisher Scientific, Waltham, MA, USA) and subjected to gel electrophoresis (0.8% agarose). The whole genome was sequenced using the DNBseq Sequencing platform at BGI, Inc.

Genome assembly and analysis

Raw data with adapter or low-quality sequences were filtered. We first went through a series of data processing steps to remove contamination and obtain valid data. This step was completed using bynSOAPnuke software. SOAPnuke software filter parameters: "-n 0.01 -l 20 -q 0.4 -ada Mis 3 -out QualSys 1 -minReadLen 150" [30]. Sequences were assembled using a *de novo* assembler implemented as an initial assembly graph from short reads in Unicycler [31], and the assembly metrics were evaluated using QUAST v4.6 [32].

The completeness and contamination of all genomes were estimated using CheckM (v1.0.11) [33]. Taxonomic assignment of individual genomes was performed using GTDB-Tk - v1.7.0, with default settings. Open reading frames (ORFs) were predicted by Prokka v. 1.14.6 [34] and then annotated by KOfam and custom HMM profiles within METABOLIC v.4.0 [35] and eggNOG-mapper v.2.1.2 [36] with default settings. Transporters were identified by aligning genome protein sequences to reference transporter sequences of the TCDB database [37]. Carbohydrate active enzymes (CAZy) were identified by searching protein sequences against dbCAN [38] within METABOLIC v.4.0 [35]. The genomes were submitted to the National Center for Biotechnology Information (NCBI) GenBank under the bioproject number PRJNA950099. The protein sequences of the genomes were used to predict plant growth-promoting traits (PGPTs) through PLaBAse (v1.01, http://plabase.informatik.uni-tuebingen.de/pb/plabase.php) [39] and were also mined for biosynthetic gene clusters (BGCs) using antiSMASH v5.1 [40].

Results

Slow-growing bacteria isolated from Cerrado soil

Soil samples were obtained in the Cerrado *sensu stricto* in a forest remnant in Rio Verde, Goiás, Brazil (Fig. S1). The predominant climate in this region is a tropical climate, with warm temperatures in the winter and high temperatures in the summer, and two distinct seasons: a dry season (from May to October) and a rainy season (from November to April). This area has a latosol profile with an annual rainfall of 1,800 mm and an annual temperature range of 20 °C to 35 °C. We used the VL55 medium, and the incubation period was extended to 15 weeks to enhance the selection of slow-growing bacteria. Slow growers were defined as colonies that appeared after more than 7 days of incubation [14], and the growth of colonies was monitored and picked weekly. A total of 92 isolates were selected within four weeks of incubation (Fig. 1A). In addition, we noticed that the number of isolates was reduced while the incubation time was increased. At the end of the first week, 52 isolates were obtained, followed by 23 isolates in the second week, 10 in the third week, 2 in the fourth week, and 5 new isolates after more than 5 weeks of incubation (Fig. 1A). The isolates were named CER (from Cerrado) followed by a sequential number according to their isolation. During the period of isolation, a particular isolate was found to be associated with another CER isolate and was observed to inhibit its growth. This specific isolate was designated ANTI. Most of these slow-growing bacteria had formed small-sized colonies with a variety of shapes, colors (or lack thereof), and textures. Notably, after an extended period of incubation, we only observed the growth of microsized cultures in the VL55 medium, and most of these isolates were unable to thrive on fresh agar medium.

Slow-growing bacteria grow in low-water conditions and have the potential to enhance plant growth

Slow-growing bacteria are well known for their ability to tolerate stressful or starving conditions. As a result, we assessed their potential under drought conditions by testing their capacity for growth in a medium with reduced water availability. [18] Forty-seven percent (24/51) of the slow-growing bacteria isolated in the second week grew in at least two different concentrations of sorbitol (0 g L^{-1}; 85 g L^{-1}), and the same result was observed in 45% (15/33) of the slow-growing bacteria isolated after the third week. In addition, although growth was relatively low at the highest sorbitol concentration, only 19% of the slow-growing bacteria managed to grow at all sorbitol concentrations (Fig. 1B).

Next, we selected slow-growing bacteria that grew at concentrations of 85 g L^{-1}, 285 g L^{-1}, and 660 g L^{-1} sorbitol to screen for *in vitro* plant growth promotion traits, including phosphate solubilization, IAA production, EPS production, and siderophore production (Fig. 1C). Almost

all slow-growing bacteria displayed at least one characteristic for plant growth enhancement, except the isolates CER22 and CER66. The isolates CER49, CER54, CER55, CER61, and CER67 produced most of the features (Fig. 1C), and it is interesting to note that the four last isolates were isolated in the second week of incubation (Fig. 1C).

The slow-growing bacteria are phylogenetically related to known bacterial species, yet they may represent a group of new species

During this study, we collected slow-growing bacteria that were isolated in the second week and demonstrated the potential to promote plant growth. Additionally, we also gathered all slow-growing bacteria that were isolated during the second week of incubation and performed 16S rRNA sequencing. Analysis of the sequencing data revealed that the slow-growing bacteria belonged to two major taxa based on taxonomic classification: Proteobacteria (including α-proteobacteria and β-proteobacteria classes) and Terrabacteria (including Firmicutes and Actinobacteria classes) (Fig. S2). We found that *Caballeronia* sp., *Neobacillus* sp., *Bradyrhizobium* sp., and high GC gram-positive species (such as *Arthrobacter* sp., *Intrasporangium* sp., *Monashia* sp., and *Paenarthrobacter* sp.) were highly enriched among the slow-growing bacteria (Table S1).

We used the 16S rRNA gene sequences to gain insight into the phylogenetic relationship previously mentioned in Table S1 with known bacterial species. To assign the species, we first aligned the sequences against their respective family taxa (data not shown) and then against closely related strains in the reference sequence (RefSeq) database. Phylogenetic trees display the tree of *Caballeronia* sp., *Neobacillus* sp., *Bradyrhizobium* sp., and high GC gram-positive species (Fig. 2). It is noteworthy that these slow-growing bacteria appeared to cluster closely with phylogenetically related bacterial species, but they may belong to a new group of species, except for CER28, which was clustered in the same branch as *Arthrobacter alkaliphilus* with 98 bootstrap replicates. Figs. S3-S9 provide a detailed view of the evolutionary trees for each species.

Phylogenetic analysis also found a group of slow-growing bacteria that may belong to the same species and perhaps clones, genotypically indistinguishable isolates. To this end, we grouped these isolates and used the BOX-PCR fingerprint to perform molecular typing. We discovered that four isolates of *Neobacillus* sp. form a single clone, and two separate sets of clones, CER56/CER57 and CER65 through CER72, were detected in *Caballeronia* sp. (Fig. S10).

Taken together, our findings show that the collections of slow-growing bacteria retrieved contain several unique bacterial species.

Growth of bacterial strains in varied culture conditions reveals a slow growth phenotype

We sought to investigate whether five previously reported strains, known for their potential interaction with plants, exhibit slow growth due to culture media and/or temperature conditions over five days (Fig. 3A). The medium conditions ranged from oligotrophic to rich media. During the experiment, we noticed that only two strains, namely, ANTI and CER19, displayed rapid growth. They reached a density of 0.1 after 30 and 25 hours of cultivation, respectively (Fig. 3A). However, the remaining strains exhibited a prolonged lag phase in bacterial growth, with an initial logarithmic phase observed after 70 hours (Fig. 3A). Interestingly, neither changes in the media composition nor temperature variations significantly affected the growth of these strains throughout the observation period (Fig. 3A, Fig. 3B). The highest density observed was 0.2, indicating that these strains indeed exhibit slow growth. We also observed that the NB medium promoted population growth for ANTI and CER19, while CER78 thrived better in the oligotrophic VL55 medium. However, most strains had limited growth in this medium (Fig. 3C). In general, the majority of the strains exhibited better growth at temperatures of 25 °C and 28 °C, with 28 °C showing the greatest enhancement (Fig. 3D).

Insights into the taxonomy, metabolic pathways, and functional diversity of slow-growing bacterial genomes

We next gained insights into the genomes of these five strains. GTDB-Tk was used to classify the genomes taxonomically, revealing that two isolates had been assigned to species that were already known (Table S2). ANTI was the assigned species for *Paenibacillus polymyxa*, and it had a genome consisting of 26 contigs with a size of 6.2 Mb, an N50 value of 374 kb, and a GC content of 45% (Table S2, Fig. S11). The strain CER44 was assigned to *Methylobacterium brachiatum*, whose genome was composed of 6.4 Mb assembled in 57 contigs with a GC content of 69%. Additionally, we were able to map four contigs that were circularized and found to be plasmids. Although *M. brachiatum* is known to be isolated from fresh water and even human skin, this is, to the best of our knowledge, the first report of this species being isolated from soil.

At a lower taxonomic rank, we identified novel species of the genera *Nocardioides*, *Amnibacterium,* and *Bradyrhizobium*, represented by the strains CER19, CER49, and CER78,

respectively. It is worth noting that *Amnibacterium* sp. CER49 was identified as a taxonomic novelty using relative evolutionary divergence (RED). CER49 has a genome size of 3.3 Mb, which is assembled into 2 contigs and has a GC content of 73%. It belongs to a group of high-GC Gram-positive species along with CER19, whose genome size is 4.6 Mb assembled into 9 contigs with a GC content of 71% (Table S2, Fig. S11). Finally, *Bradyrhizobium* sp. CER78 has a genome size of 7.9 Mb, which is assembled into 39 contigs and has a GC content of 64%. Its closest related taxonomic classification is *Bradyrhizobium macuxiense,* which was obtained from the root nodules of *Centrolobium paraense* Tul. in the Amazon region of Brazil [41].

The analysis of the genomes revealed that a significant proportion of the genes, ranging from 17 to 20%, were classified as having an unknown function (N, NC) by the EggNOG analysis, suggesting that there is still much to be discovered and understood about the functional diversity of these genomes (Fig. S12). Additionally, the analysis also revealed that genes classified under C (energy production and conversion), E (amino acid metabolism and transport), K (transcription), M (cell wall/membrane/envelope biogenesis), and P (inorganic ion transport and metabolism) were prevalent in these genomes (Fig. S12A). We observed that the classification of genes was consistent with the genome size, meaning that larger genomes tended to have a higher number of genes assigned to Clusters of Orthologous Genes (COG) functional categories (Fig. S12B). However, we also noted that despite the relatively small genome size of *Amnibacterium* sp. CER49, nearly half of its genes were not assigned to COG functional categories. In contrast, *M. brachiatum* strain CER44 has a comparatively high degree of functional diversity despite its genome size.

The metabolic capability of slow-growing bacteria was predicted based on their genome using METABOLIC [35]. Many genes involved in central metabolic pathways, such as glycolysis, gluconeogenesis, the pentose phosphate pathway, and the citrate cycle, were present in the genomes analyzed (Fig. 4, Table S3). However, we observed that the number of genes involved in these pathways varied among the species. In particular, we found that the high GC Gram-positive species *Nocardioides* sp. CER19 and *Amnibacterium* sp. CER49 had similar metabolic profiles (Fig. 4). Notably, both species lacked genes involved in the conversion of fructose 6-phosphate to ribose 5-phosphate, which is a characteristic feature of the pentose phosphate pathway. We mapped genes involved in the biosynthesis and metabolism of essential and aromatic amino acids and found that all genomes analyzed contained genes related to the biosynthesis of tryptophan and phenylalanine (Fig. S13). Additionally, an examination of genes associated with different metabolic pathways revealed that numerous genomes possessed the

capability to synthesize and metabolize diverse carbohydrates, including inositol, ascorbate, galactose, and D-galacturonate (Fig. S13). We observed that *M. brachiatum* strain CER44 had an enriched number of genes associated with the ethylmalonyl pathway, which is involved in the assimilation of methylamine.

The genomes were searched for genes encoding carbohydrate-degrading enzymes described in the Carbohydrate-Active enzymes (CAZymes) and peptidases (Fig. S14, Table S4, Table S5). Multiple CAZymes were found encoded in the genomes, with the *P. polymyxa* strain ANTI being the most prolific, producing 150 enzymes. The most frequently found CAZymes were glycoside hydrolases of the families GH013, GH109, and GH001 (Fig. 5, Fig. S14A). Together, these glycoside hydrolases are responsible for hydrolyzing glycosidic bonds in starch, degrading xylan, and hydrolyzing glycosidic bonds in various types of carbohydrates, including cellulose, chitin, and xylan [42]. Additionally, our analysis revealed that *Bradyrhizobium* sp. CER78 had the highest number of peptidase genes, with S33 and C26 being enriched in all the genomes (Fig. S14B). An analysis of transporters revealed that among the five families enriched in the genomes of slow-growing bacteria, the 3A family was the most prevalent (Fig. S15A, Table S6). These transporters utilize ATP-binding cassette (ABC) mechanisms to transport a variety of molecules such as sugars, amino acids, and ions. Notably, the 9B family, which is involved in the transport of small molecules such as drugs and toxins, was also found to be significant (Fig. S15A). This family is important for bacterial antibiotic resistance and survival in toxic environments. We also examined phosphate transporters and discovered two primary transporters in the genomes. The PiT family was found to be the most prevalent among all the genomes (Fig. S15B).

Exploring the importance of slow-growing bacteria in biogeochemical cycles

We also used METABOLIC analysis to gain insights into the role of the five slow-growing bacteria in carbon, nitrogen, sulfur, and other cycles. Our findings indicate that all the bacteria have the potential to participate in organic carbon oxidation to ethanol, as well as the fermentation of acetate to ethanol, while some may also have a role in acetate oxidation (Fig. 6A, Table S7). Additionally, only *Bradyrhizobium* sp. CER78 was found to play a role in methanotrophy and carbon fixation, while *P. polymyxa* strain ANTI was the only strain capable of nitrogen fixation (Fig. 5, Fig. 6B). ANTI, along with *Nocardioides* sp. CER19 also showed potential for nitrate reduction and a role in nitrite ammonification (Fig. 5). ANTI was found to have a genomic arsenal in the sulfur cycle, being responsible for sulfur and sulfide oxidation

and reduction (Fig. 6C). All five strains were found to be able to reduce iron, but only three strains may participate in the oxidation step (Fig. 6D).

In addition, some of these bacteria, such as *M. brachiatum* strain CER44 and *Bradyrhizobium* sp. CER78, have the potential to be involved in urea utilization and chloride reduction, with the latter strains also being responsible for nitrite hydration. The distribution of 105 genes important for biogeochemical processes, which were identified in the genomes, can be found in Table S7. Altogether, these findings provide valuable insight into the potential roles of the five slow-growing bacteria in biogeochemical cycles and their impact on the environment.

Uncovering the biosynthetic potential of slow-growing bacteria for secondary metabolites and their role in promoting plant growth

Slow-growing bacteria have attracted research attention due to the possibility of encoding biosynthetic pathways to synthesize secondary metabolites that may not be found in fast-growing bacteria, including antibiotics, phytohormones, siderophores, and other plant-growth-promoting compounds. We analyzed the genomes of several bacterial strains for genes related to promoting plant growth, focusing on 31,500 genes classified into eight categories based on PLaBAse [39]. These categories included biofertilization, phytohormone/plant signal production, plant immune response stimulation, colonizing plant system, competitive exclusion, bioremediation, and stress control/biocontrol (Fig. 7A, Table S8). We observed that the number of genes mapped varied by genome size, with *Bradyrhizobium* sp. CER78 has the most genes (3,389). We then identified the genes with direct effects on plants, finding that nitrogen acquisition was a common role among most genomes. These genes include those involved in N acquisition/regulation, glutamate transport and metabolism, denitrification, urea usage, and atmospheric nitrogen fixation (Table S8). However, this category had a smaller number of genes compared to other categories. We also found that most genomes encode genes responsible for phosphate solubilization and neutralizing abiotic stress (Fig. 7A), including genes responsible for neutralizing salinity stress, such as the proline metabolism cluster *argABCDEHJO*, which can protect cells from dehydration, oxidative stress, and other stressors [43]. Heat shock proteins and genes involved in glycine-betaine also play a role in osmotic stress. Some genomes, such as CER44 and CER49, have an entire cluster of tryptophan biosynthesis (*trpABCDEFGS*), which is a precursor to indole-acetic acid (Fig. 5). In addition to indirect genes involved in the colonization of plant roots, such as chemotaxis and motility

406 genes for all strains, there are also genes involved in plant-derived carbohydrate transport,
407 including glycerol, cellobiose, maltose, sorbitol, and others (Table S8).

408 Our analysis revealed that the slow-growing bacteria contain a total of 46 biosynthetic gene
409 clusters (BCGs), which are mainly responsible for encoding nonribosomal peptide synthase
410 (NRPS) by *P. polymyxa* strain ANTI (Fig. 7B). These BCGs include well-known clusters for
411 synthesizing antimicrobial compounds such as fusaricidin B, paenibacterin, polymyxin, and
412 anabaenopeptin. Additionally, CER44 and other genomes were found to contain BCGs for
413 siderophore production, while most genomes contain BCG regions for terpene biosynthesis,
414 particularly carotenoids (Table S9). Notably, CER49 contains a cluster for synthesizing
415 alkylresorcinol, a compound with potent antimicrobial activity against a variety of bacteria and
416 fungi.

417 **Discussion**

418 Soil harbors one of the most diverse bacterial communities on the planet. It is believed that one
419 gram of soil contains up to one billion bacterial cells, each of which includes tens of thousands
420 of different species [44]. Despite the overwhelming diversity of bacteria, most of the species
421 remain largely unknown, and one of the serious issues is intrinsic to a group of bacteria that
422 have slow growth rates, classified as slow-growing bacteria. Here, we reported the isolation of
423 this remarkable group of bacteria from the soil of the Cerrado, the world's most species-rich
424 tropical savanna with an annual severe drought that provides a unique opportunity to explore
425 microbes adapted to extreme environmental conditions. Over four weeks, we isolated 92 slow-
426 growing bacteria. To do this, we used VL55 oligotrophic medium [45]. In addition to simple
427 cultivation process adjustments, such as the use of a polymeric growth substrate (xylan),
428 oligotrophic media, and a longer incubation period, VL55 has been used to isolate globally
429 widespread yet uncultured phylogenetically novel soil bacteria [45–47]. Interestingly, our
430 analysis also revealed that while some of these microbes were able to form colonies on agar
431 after several weeks of incubation, others only thrived when placed on fresh agar. Specifically,
432 some of the slow-growing bacteria present in our study were able to be domesticated on new
433 fresh media. However, after an extended period of incubation, most of these isolates were
434 unable to grow on any medium, indicating that they may be extremely recalcitrant bacteria.
435 This domestication process has been proven to make it more challenging to isolate unculturable
436 and slow-growing bacteria [48].

We showed that 42% of the slow-growing bacteria isolated in the Cerrado can grow in a medium with reduced water availability. This ability might be an adaptation to surviving the prolonged water shortages that are frequent in this environment. Hence, we tested whether these slow-growing bacteria produce features for enhancing plant growth. Almost all slow-growing bacteria tested positive for at least one feature: the production of IAA, EPS, siderophores, and phosphate solubilization. These features produced by microbes have been documented to play a major role in maintaining the health of the plant host [49, 50]. Plant growth-promoting bacteria (PGPB) surrounding their roots, as well as positive gene selection, have recently been linked to key beneficial mechanisms for plant survival in the Atacama Desert [51]. Moreover, the PGPB isolated from Brazilian cactus exhibited various traits that could enhance plant development under drought conditions. [18]. Overall, these findings shed light on our slow-growing bacteria, suggesting that they have untapped potential for improving plant development under drought conditions.

Next, to gain insight into the species, we sequenced almost full-length 16S rRNA genes. These slow-growing bacteria were found to be part of two main taxa and 19 distinct genera, including *Arthrobacter* sp., *Bacillus* sp., *Bosea* sp., *Bradyrhizobium* sp., *Caballeronia* sp., *Cupriavidus* sp., *Dermacoccus* sp., *Intrasporangium* sp., *Massilia* sp., Mesorhizobium sp., *Methylobacterium* sp., *Monashia* sp., *Neobacillus* sp., *Nocardioides* sp., *Paenarthrobacter* sp., *Pseudomonas* sp., *Rhizobium* sp. They were discovered to be phylogenetically linked to existing bacterial species but comprise a group of novel species. Although we observed certain species that were relatively ubiquitous in the isolation, such as *Bacillus*, we discovered that this strain did form a distinctive clade in phylogenetic trees along with species in the family, indicating that this strain may be unique, which also includes the characteristic of slow growth.

Caballeronia sp., a genus of bacteria from the family *Burkholderiaceae* isolated from soil, as well as plant-associated bacteria such as endophytes and legume nodulation, was one of the main enriched species in our isolation [52]. This species has been reported to perform biological nitrogen fixation and promote plant growth on extremely nutrient-poor soils [53, 54]. Five slow-growing bacteria are phylogenetically close to *Bradyrhizobium* spp., a genus of slow-growing bacteria, gram-negative, nitrogen-fixing α-Proteobacterium, that includes *B. japonicum* that forms root nodules on soybeans [55]. In addition, Actinobacteria comprise most of our slow-growing bacteria in the Terrabacteria group. This phylum is ubiquitous, abundant, and the most diverse group of bacteria in the soil that plays roles in vital ecological processes, regulation of biogeochemical cycles, decomposition of biopolymers, exopolysaccharide secretion, and plant

growth promotion [56]. Additionally, five slow-growing bacteria belong to *Arthrobacter* spp., a bacterium commonly found in soil that exhibits an intriguing cell division known as "snapping division" or reversion in which the exterior bacterial cell wall ruptures at a joint, being rods during exponential growth and cocci in their stationary phase [57]. One of this species' most well-known roles is the bioremediation of hexavalent chromium and 4-chlorophenol levels in contaminated soil by *A. crystallopoietes* [58]. It has also been characterized as a plant growth-promoting rhizobacterium and endophytic with the potential to control fungal and bacterial pathogens [59–61].

We have identified a group of slow-growing bacteria known as *Neobacillus* species. Members of this genus were originally classified as *Bacillus*; however, comparative genomic investigations of *Bacillus* species indicated that isolates from these genera shared several molecular synapomorphies, resulting in the formation of a new genus [62]. *Neobacilli* can be found in a wide range of environments, including soil, human origin, and plant roots, as well as endophytes, which have been reported to enhance plant growth [63, 64]. To the best of our knowledge, in this study, most of these species were the first documented slow-growing bacteria.

Five isolates were selected for genome sequencing, which provided valuable information to gain insights into their taxonomy, metabolic capabilities, and functional diversity. Genomic analysis showed that some isolates belonged to known species, while others represented novel species. *Amnibacterium* sp. CER49 was identified as a taxonomic novelty using relative evolutionary divergence (RED). This metric is used to measure the evolutionary distance between two organisms based on their genetic differences and estimate the evolutionary divergence between organisms that is independent of the time since they last shared a common ancestor [65] The bacterium in question is classified under the Actinomycetota phylum, and its genus refers to a rod-shaped bacterium commonly found in rivers. Hence, *Amnibacterium* sp. CER49 could potentially represent a distinct species within this group.

These genomes provided essential information about the gene's functional categories, showing that 17-20% of the genes are still largely unknown. Despite consistency with genome size, *M. brachiatum* strain CER44 displays a comparatively high degree of functional diversity. *Methylobacterium* species are pink-colored, gram-negative, aerobic, nonspore-forming, facultative methylotrophic bacteria commonly found in various environments, such as soil, water, and plants [66]. This genus has gained attention as a novel plant growth-promoting bacteria due to its ability to fix nitrogen, nodulate the host plant, and produce the plant hormone

cytokinin and the enzymes pectinase and cellulase, thereby promoting plant growth through increased nitrogen availability and systemic resistance induction [67].

Our analysis indicated that many genes involved in central metabolic pathways were present in the genomes of the five slow-growing bacteria. However, we also observed that these microbes exhibit slow growth rates when isolated under laboratory conditions, which raises the question of why this was the case. One potential explanation is that it may be intrinsic to the environment in which these microbes were originally found - an oligotrophic soil that is limited in terms of available nutrients, which could slow down growth rates.

Genome analysis has also provided an insightful view into the potential roles of the five slow-growing bacteria in carbon, nitrogen, sulfur, and other biogeochemical cycles. The results indicate that all five strains could participate in organic carbon oxidation to ethanol and fermentation of acetate to ethanol, with some strains also capable of acetate oxidation. Furthermore, only *Bradyrhizobium* sp. CER78 was found to be involved in methanotrophy and carbon fixation, while *P. polymyxa* strain ANTI was the only strain capable of nitrogen fixation. Overall, these findings offer valuable insights into the potential roles of the five slow-growing bacteria in biogeochemical cycles and their impact as individual organisms on communities. This prediction is linked to the biosynthetic potential of slow-growing bacteria for secondary metabolites and their ability to promote plant growth. Our research demonstrates that microbe genomes contain genes responsible for biofertilization, production of phytohormones/plant signals, stimulation of plant immune responses, colonization of plant systems, competitive exclusion, bioremediation, and stress control/biocontrol. Furthermore, the analysis exposed a range of unique BGCs for secondary metabolites that could be further explored in future studies. It is necessary to supplement more complex nutrients in culture media to support the growth of recalcitrant isolates such as *Amnibacterium* sp. CER49. In summary, genome analysis is an effective tool for comprehending the growth and functions of microbes in the environment. By providing insights into the genetic mechanisms that enable microbial growth and function, genome analysis can facilitate the development of strategies for managing microbial populations and harnessing their potential for various applications.

Conclusions

In this study, we isolated 92 slow-growing bacteria from tropical savanna soil, providing a unique opportunity to study microbes that have adapted to challenging environmental conditions, such as drought and acidity. Despite taking several weeks to grow, these bacteria

were successfully cultivated under laboratory conditions. Our analysis revealed their potential to enhance plant growth under drought conditions, and genomic analysis provided insights into their taxonomy, metabolic capabilities, and functional diversity. This information is crucial for developing strategies to assist the future direction of microbial growth and applications. Overall, with the development of new techniques for studying soil microbiology, there is optimism that we will be able to uncover the secrets of slow-growing bacteria and utilize their power for sustainable agriculture and environmental management.

Contributions

OSG.: investigation, visualization, writing—original draft, and writing—review and editing; MFS.: conceptualization, funding acquisition, resources, supervision, validation, writing—review and editing.

Acknowledgments

This work was financially supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq (Process APQ-02381-21), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior/Programa de Excelência Acadêmica-Finance Code 001 (CAPES ProEx grant 23038.019105/2016-86) and Fundação de Amparo à Pesquisa do Estado de Minas Gerais—FAPEMIG (Process 402644/2021-2). The authors expressed their gratitude to the technical support team of the Cluster at Universidade Federal de Viçosa.

Competing Interests

The authors have no relevant financial or nonfinancial interests to disclose.

Data availability

The names of the repositories/repositories and accession number(s) can be found in the Supplementary Material for this manuscript. The genomes were submitted to the National Center for Biotechnology Information (NCBI) GenBank under the Bioproject number PRJNA950099.

References

1. Staley JT, Konopka A (1985) Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. Annu Rev Microbiol 39:321–346. https://doi.org/10.1146/annurev.mi.39.100185.001541

2. Fierer N (2017) Embracing the unknown: disentangling the complexities of the soil microbiome. Nat Rev Microbiol 15:579–590. https://doi.org/10.1038/nrmicro.2017.87

3. Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev 59:143–169. https://doi.org/10.1128/mr.59.1.143-169.1995

4. Stewart EJ (2012) Growing unculturable bacteria. J Bacteriol 194:4151–4160. https://doi.org/10.1128/JB.00345-12

5. Rappé MS, Giovannoni SJ (2003) The Uncultured Microbial Majority. Annu Rev Microbiol 57:369–394. https://doi.org/10.1146/annurev.micro.57.030502.090759

6. Stevenson BS, Eichorst SA, Wertz JT, et al (2004) New strategies for cultivation and detection of previously uncultured microbes. Appl Environ Microbiol 70:4748–4755. https://doi.org/10.1128/AEM.70.8.4748-4755.2004

7. Schimel J, Schaeffer S (2012) Microbial control over carbon cycling in soil. Front Microbiol 3:

8. Navarrete AA, Tsai SM, Mendes LW, et al (2015) Soil microbiome responses to the short-term effects of Amazonian deforestation. Mol Ecol 24:2433–2448./https://doi.org/10.1111/mec.13172

9. Coutinho LM (2006) O conceito de bioma. Acta Bot Brasilica 20:13–23

10. KLINK CA, MACHADO RB (2005) Conservation of the Brazilian Cerrado. Conservation Biology 19:707–713. https://doi.org/10.1111/j.1523-1739.2005. 00702.x

11. Procópio L, Barreto C (2021) The soil microbiomes of the Brazilian Cerrado. J Soils Sediments 21:2327–2342. https://doi.org/10.1007/s11368-021-02936-9

12. Araujo JF, de Castro AP, Costa MMC, et al (2012) Characterization of Soil Bacterial Assemblies in Brazilian Savanna-Like Vegetation Reveals Acidobacteria Dominance. Microb Ecol 64:760–770. https://doi.org/10.1007/s00248-012-0057-3

13. Janssen PH, Yates PS, Grinton BE, et al (2002) Improved culturability of soil bacteria and isolation in pure culture of novel members of the divisions Acidobacteria, Actinobacteria, Proteobacteria, and Verrucomicrobia. Appl Environ Microbiol 68:2391–2396. https://doi.org/10.1128/AEM.68.5.2391-2396.2002

14. Kato S, Yamagishi A, Daimon S, et al (2018) Isolation of Previously Uncultured Slow-Growing Bacteria by Using a Simple Modification in the Preparation of Agar Media. Appl Environ Microbiol 84:e00807-18. https://doi.org/10.1128/AEM.00807-18

15. Gonçalves OS, Souza TS, Gonçalves G de C, et al (2023) Harnessing Novel Soil Bacteria for Beneficial Interactions with Soybean. Microorganisms 11:. https://doi.org/10.3390/microorganisms11020300

16. Tschech A, Pfennig N (1984) Growth yield increase linked to caffeate reduction in *Acetobacterium woodii*. Arch Microbiol 137:163–167. https://doi.org/10.1007/BF00414460

17. Widdel F, Kohring G-W, Mayer F (1983) Studies on dissimilatory sulfate-reducing bacteria that decompose fatty acids. Arch Microbiol 134:286–294. https://doi.org/10.1007/BF00407804

18. Kavamura VN, Santos SN, Silva JL da, et al (2013) Screening of Brazilian cacti rhizobacteria for plant growth promotion under drought. Microbiol Res 168:183–191. https://doi.org/https://doi.org/10.1016/j.micres.2012.12.002

19. Paulo EM, Vasconcelos MP, Oliveira IS, et al (2012) An alternative method for screening lactic acid bacteria for the production of exopolysaccharides with rapid confirmation. Food Science and Technology 32:710–714. https://doi.org/10.1590/S0101-20612012005000094

20. Gordon SA, Weber RP (1951) Colorimetric estimation of indoleacetic acid. Plant Physiol 26:192–195. https://doi.org/10.1104/pp.26.1.192

21. Nautiyal CS (1999) An efficient microbiological growth medium for screening phosphate solubilizing microorganisms. FEMS Microbiol Lett 170:265–270. https://doi.org/10.1111/j.1574-6968.1999.tb13383.x

22. Schwyn B, Neilands JB (1987) Universal chemical assay for the detection and determination of siderophores. Anal Biochem 160:47–56. https://doi.org/10.1016/0003-2697(87)90612-9

23. Heuer H, Krsek M, Baker P, et al (1997) Analysis of actinomycete communities by specific amplification of genes encoding 16S rRNA and gel-electrophoretic separation in denaturing gradients. Appl Environ Microbiol 63:3233–3241

24. Altschul SF, Madden TL, Schäffer AA, et al (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402. https://doi.org/10.1093/nar/25.17.3389

25. Menzel P, Ng KL, Krogh A (2016) Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun 7:11257. https://doi.org/10.1038/ncomms11257

26. Katoh K, Rozewicki J, Yamada KD (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform 20:1160–1166. https://doi.org/10.1093/bib/bbx108

27. Larkin MA, Wilm A, Higgins DG, et al (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948. https://doi.org/10.1093/bioinformatics/btm404

28. Tamura K, Stecher G, Kumar S (2021) MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol 38:3022–3027. https://doi.org/10.1093/molbev/msab120

29. Koeuth T, Versalovic J, Lupski JR (1995) Differential subsequence conservation of interspersed repetitive *Streptococcus pneumoniae* BOX elements in diverse bacteria. Genome Res 5:408–418. https://doi.org/10.1101/gr.5.4.408

30. Chan Y, Chen Y, Shi C, et al (2017) SOAPnuke: A MapReduce Acceleration supported Software for integrated Quality Control and Preprocessing of High-Throughput Sequencing Data. Gigascience 7: https://doi.org/10.1093/gigascience/gix120

31. Wick RR, Judd LM, Gorrie CL, Holt KE (2017) Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595-

32. Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086

33. Parks D, Imelfort M, Skennerton C, et al (2015) CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25. https://doi.org/10.1101/gr.186072.114

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

35. Zhou Z, Tran PQ, Breister AM, et al (2022) METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. Microbiome 10:33. https://doi.org/10.1186/s40168-021-01213-8

36. Cantalapiedra CP, Hernández-Plaza A, Letunic I, et al (2021) eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Mol Biol Evol 38:5825–5829. https://doi.org/10.1093/molbev/msab293

37. Saier Jr MH, Reddy VS, Moreno-Hagelsieb G, et al (2021) The Transporter Classification Database (TCDB): 2021 update. Nucleic Acids Res 49:D461–D467. https://doi.org/10.1093/nar/gkaa1004

38. Zhang H, Yohe T, Huang L, et al (2018) dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 46:W95–W101. https://doi.org/10.1093/nar/gky418

39. Patz S, Gautam A, Becker M, et al (2021) PLaBAse: A comprehensive web resource for analyzing the plant growth-promoting potential of plant-associated bacteria. bioRxiv 2021.12.13.472471. https://doi.org/10.1101/2021.12.13.472471

40. Blin K, Shaw S, Steinke K, et al (2019) antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 47:W81–W87. https://doi.org/10.1093/nar/gkz310

41. Michel DC, Passos SR, Simões-Araujo JL, et al (2017) *Bradyrhizobium centrolobii* and *Bradyrhizobium macuxiense* sp. nov. isolated from *Centrolobium paraense* grown in soil of Amazonia, Brazil. Arch Microbiol 199:657–664. https://doi.org/10.1007/s00203-017-1340-y

42. Lombard V, Golaconda Ramulu H, Drula E, et al (2014) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490–D495. https://doi.org/10.1093/nar/gkt1178

43. Szabados L, Savouré A (2010) Proline: a multifunctional amino acid. Trends Plant Sci 15:89–97. https://doi.org/10.1016/j.tplants.2009.11.009

44. Roesch LFW, Fulthorpe RR, Riva A, et al (2007) Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J 1:283–290. https://doi.org/10.1038/ismej.2007.53

45. Sait M, Hugenholtz P, Janssen PH (2002) Cultivation of globally distributed soil bacteria from phylogenetic lineages previously only detected in cultivation-independent surveys. Environ Microbiol 4:654–666. https://doi.org/ 10.1046/j.1462-2920.2002.00352.x

46. Davis KER, Joseph SJ, Janssen PH (2005) Effects of growth medium, inoculum size, and incubation time on culturability and isolation of soil bacteria. Appl Environ Microbiol 71:826–834. https://doi.org/10.1128/AEM.71.2.826-834.2005

47. Sait M, Davis KER, Janssen PH (2006) Effect of pH on isolation and distribution of members of subdivision 1 of the phylum Acidobacteria occurring in soil. Appl Environ Microbiol 72:1852–1857. https://doi.org/10.1128/AEM.72.3.1852-1857.2006

48. Lewis K, Epstein S, D'Onofrio A, Ling LL (2010) Uncultured microorganisms as a source of secondary metabolites. J Antibiot (Tokyo) 63:468–476. https://doi.org/10.1038/ja.2010.87

49. Souza R, Ambrosini A, Passaglia LMP, et al (2015) Plant growth-promoting bacteria as inoculants in agricultural soils. Genet Mol Biol 38:401–419. https://doi.org/10.1590/S1415-475738420150053

50. Backer R, Rokem JS, Ilangumaran G, et al (2018) Plant Growth-Promoting Rhizobacteria: Context, Mechanisms of Action, and Roadmap to Commercialization of Biostimulants for Sustainable Agriculture. Front Plant Sci 9:

51. Eshel G, Araus V, Undurraga S, et al (2021) Plant ecological genomics at the limits of life in the Atacama Desert. Proceedings of the National Academy of Sciences 118:e2101177118. https://doi.org/10.1073/pnas.2101177118

52. Dobritsa AP, Samadpour M (2016) Transfer of eleven species of the genus *Burkholderia* to the genus *Paraburkholderia* and proposal of *Caballeronia* gen. nov. to accommodate twelve species of the genera *Burkholderia* and *Paraburkholderia*. Int J Syst Evol Microbiol 66:2836–2846. https://doi.org/10.1099/ijsem.0.001065

53. Puri A, Padda KP, Chanway CP (2020) Can naturally occurring endophytic nitrogen-fixing bacteria of hybrid white spruce sustain boreal forest tree growth on extremely nutrient-poor soils? Soil Biol Biochem 140:107642. https://doi.org/https://doi.org/10.1016/j.soilbio.2019.107642

54. Puri A, Padda KP, Chanway CP (2020) Sustaining the growth of Pinaceae trees under nutrient-limited edaphic conditions via plant-beneficial bacteria. PLoS One 15:e0238055-

55. Jordan DC (1982) NOTES: Transfer of *Rhizobium japonicum* Buchanan 1980 to *Bradyrhizobium* gen. nov., a Genus of Slow-Growing, Root Nodule Bacteria from Leguminous Plants. Int J Syst Evol Microbiol 32:136–139. https://doi.org/10.1099/00207713-32-1-136

56. Barka EA, Vatsa P, Sanchez L, et al (2015) Taxonomy, Physiology, and Natural Products of Actinobacteria. Microbiol Mol Biol Rev 80:1–43. https://doi.org/10.1128/MMBR.00019-15

57. Jones D, Keddie RM (2006) The Genus Arthrobacter. In: Dworkin M, Falkow S, Rosenberg E, et al (eds) The Prokaryotes: Volume 3: Archaea. Bacteria: Firmicutes, Actinomycetes. Springer New York, New York, NY, pp 945–960

58. Westerberg K, Elväng AM, Stackebrandt E, Jansson JK (2000) Arthrobacter chlorophenolicus sp. nov., a new species capable of degrading high concentrations of 4-chlorophenol. Int J Syst Evol Microbiol 50:2083–2092. https://doi.org/10.1099/00207713-50-6-2083

59. Aviles-Garcia ME, Flores-Cortez I, Hernández-Soberano C, et al (2016) La rizobacteria promotora del crecimiento vegetal *Arthrobacter agilis* UMCV2 coloniza endofíticamente a Medicago truncatula. Rev Argent Microbiol 48:342–346. https://doi.org/https://doi.org/10.1016/j.ram.2016.07.004

60. Xu X, Xu M, Zhao Q, et al (2018) Complete Genome Sequence of Cd(II)-Resistant *Arthrobacter* sp. PGP41, a Plant Growth-Promoting Bacterial with Potential in Microbe-Assisted Phytoremediation. Curr Microbiol 75:1231–1239. https://doi.org/10.1007/s00284-018-1515-z

61. Sun Y, Sun P, Xue J, et al (2022) *Arthrobacter wenxiniae* sp. nov., a novel plant growth-promoting rhizobacteria species harboring a carotenoids biosynthetic gene cluster. Antonie Van Leeuwenhoek. https://doi.org/10.1007/s10482-021-01701-9

62. Patel S, Gupta RS (2020) A phylogenomic and comparative genomic framework for resolving the polyphyly of the genus *Bacillus*: Proposal for six new genera of *Bacillus* species, *Peribacillus* gen. nov., *Cytobacillus* gen. nov., *Mesobacillus* gen. nov., *Neobacillus* gen. nov., *Metabacillus* gen. nov. and *Alkalihalobacillus* gen. nov. Int J Syst Evol Microbiol 70:406–438. https://doi.org/10.1099/ijsem.0.003775

63. Yadav S, Kaushik R, Saxena AK, Arora DK (2011) Diversity and phylogeny of plant growth-promoting bacilli from moderately acidic soil. J Basic Microbiol 51:98–106. https://doi.org/https://doi.org/10.1002/jobm.201000098

64. Hernández-Pacheco CE, Orozco-Mosqueda M del C, Flores A, et al (2021) Tissue-specific diversity of bacterial endophytes in Mexican husk tomato plants (*Physalis ixocarpa* Brot. ex Horm. and screening for their multiple plant growth-promoting activities. Curr Res Microb Sci 2:100028. https://doi.org/10.1016/j.crmicr.2021.100028

65. Parks DH, Chuvochina M, Waite DW, et al (2018) A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol 36:996–1004. https://doi.org/10.1038/nbt.4229

66. Green PN, Ardley JK (2018) Review of the genus *Methylobacterium* and closely related organisms: a proposal that some *Methylobacterium* species be reclassified into a new genus, *Methylorubrum* gen. nov. Int J Syst Evol Microbiol 68:2727–2748. https://doi.org/10.1099/ijsem.0.002856

67. Grossi CEM, Fantino E, Serral F, et al (2020) *Methylobacterium* sp. 2A Is a Plant Growth-Promoting Rhizobacteria That Has the Potential to Improve Potato Crop Yield Under Adverse Conditions. Front Plant Sci 11.

Figures

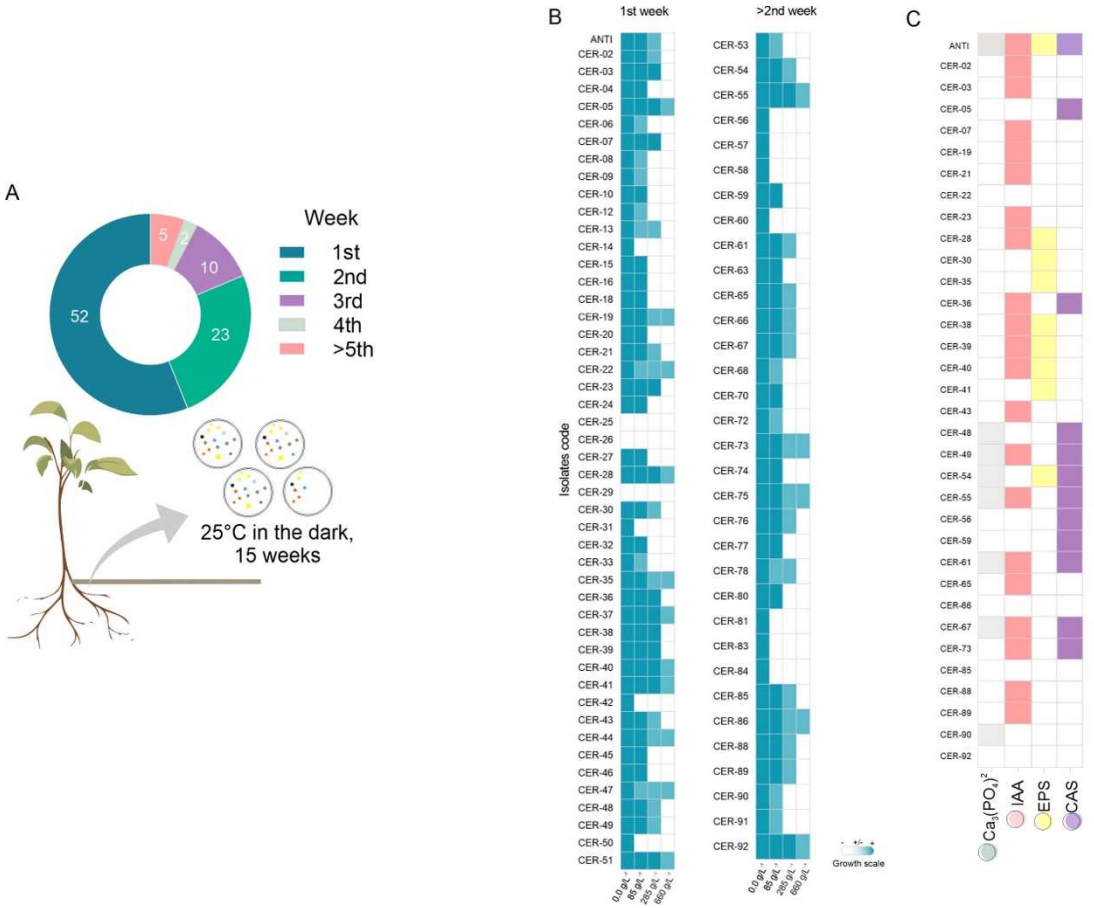


Figure 1. Experimental strategy and *in vitro* features for plant growth promotion were tested. (A) Sample collection strategy displaying total incubation time and the number of isolates per week. (B) The ability of bacteria to grow in a medium with reduced water availability containing different sorbitol in increasing concentrations. The colored squads indicate the growth of each isolate. The scale used to indicate growth levels is as follows: (-) no growth, (+/-) moderate growth, and (+) significant growth. (C) The production of features for plants promotes growth. Left to right: phosphate solubilization $(Ca_3(PO_4)^2$ in gray; synthesis of indole acetic acid (IAA) in pink; exopolysaccharide (EPS) production in yellow; siderophore production in CAS medium in purple. The colored squads indicate the growth of each isolate.

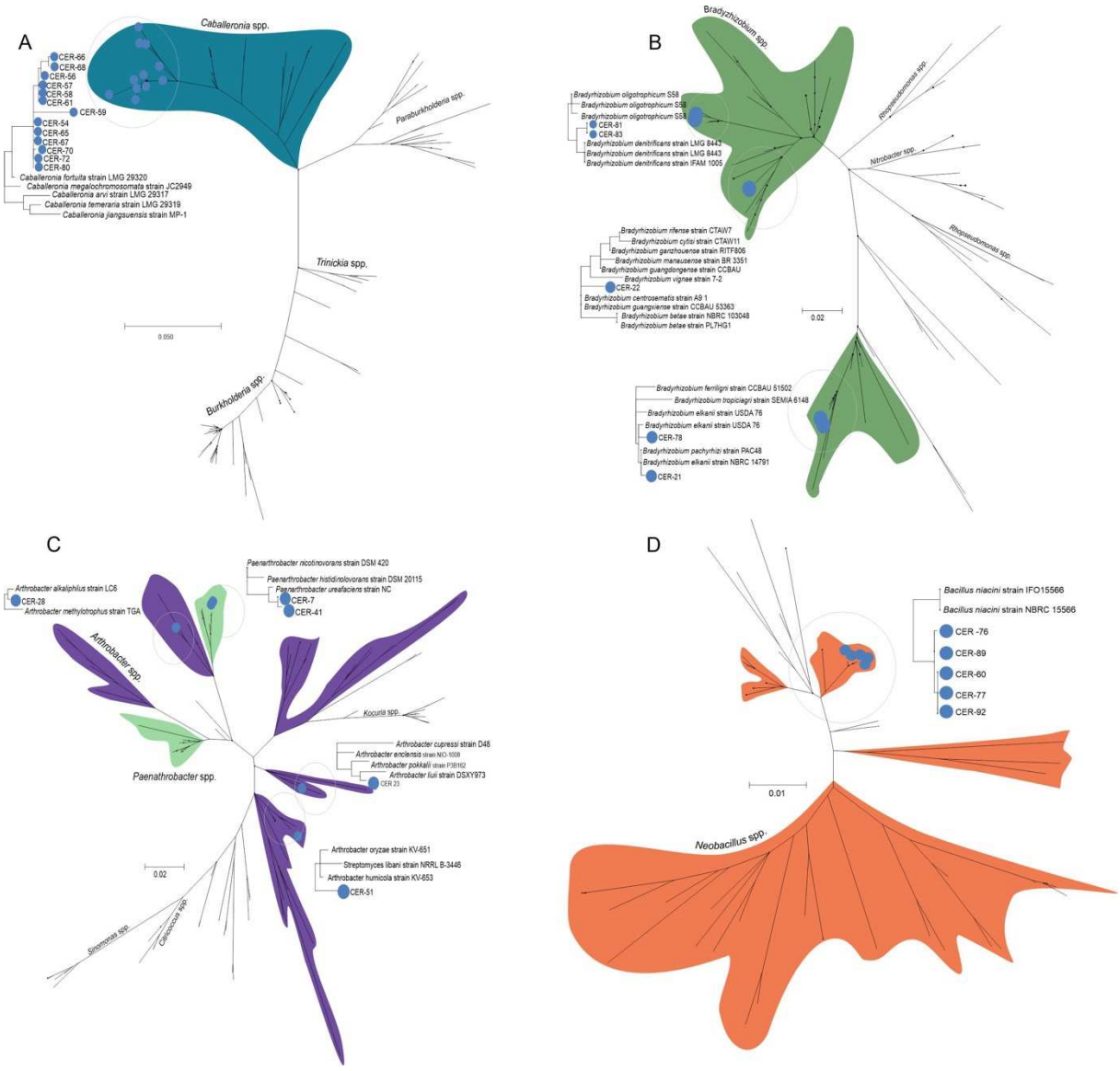


Figure 2. Evolutionary relationships among the slow-growing bacteria and related known species. (A) Likelihood phylogeny tree of *Caballeronia* species. (B) Likelihood phylogeny tree of *Bradyrhizobium* species. (C) Likelihood phylogeny tree of *Arthrobacter/Paenarthrobacter* species. (D) Likelihood phylogeny tree of *Neobacillus* species. Colored shades denote the same species cluster, while blue dots indicate slow-growing bacterial isolate codes. The evolutionary history was inferred using the maximum likelihood tree (1000 bootstrap replicates) and the substitution model General Time Revesible+gamma distribution with invariable sites (G+I). The scale bar at the bottom indicates the number of differences in base composition among sequences.

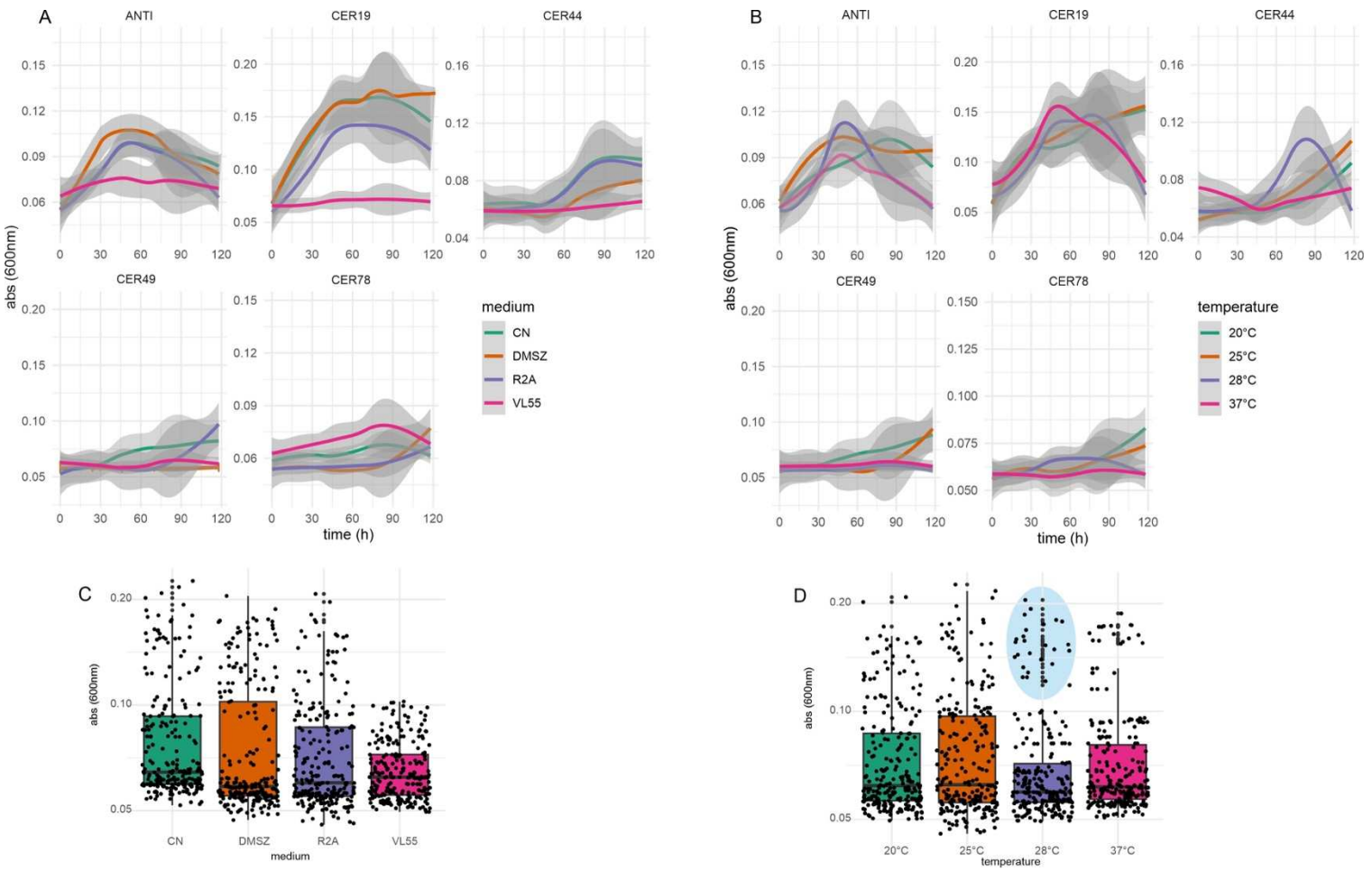


Figure 3. Standard growth curve of slow-growing bacteria. (A) Growth curves of five strains of slow-growing bacteria over a six-day period using different culture media. (B) The growth curve of the same five strains over six days but under varying temperature conditions. The x-axis represents time in hours, while the y-axis represents absorbance at 600 nm (nanometer). The line curve represents the average absorbance of two measurements, and the shaded regions indicate the 95% confidence interval. (C) The distribution of absorbance in the different media. (D) The distribution of absorbance under different temperature conditions. The blue shading indicates values above the average absorbance

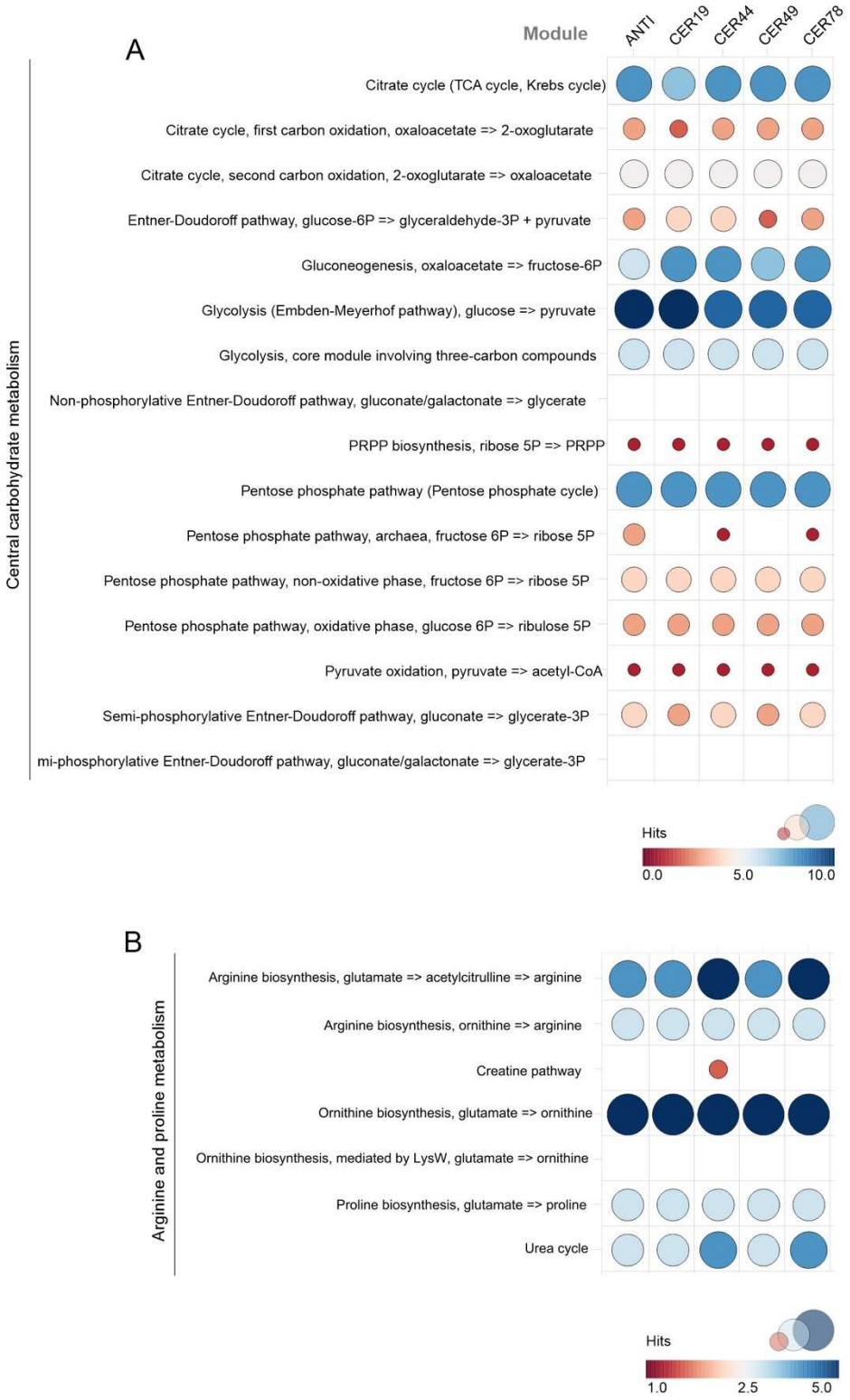


Figure 4. Distribution of metabolic marker genes in the genomes of slow-growing bacteria, focusing on those involved in central metabolic pathways (A) and essential amino acid biosynthesis (B). The size and color of each circle correspond to the number of hits for the respective gene. For more detailed information about these pathways and a full description of the marker genes present in the genomes, see Table S6.

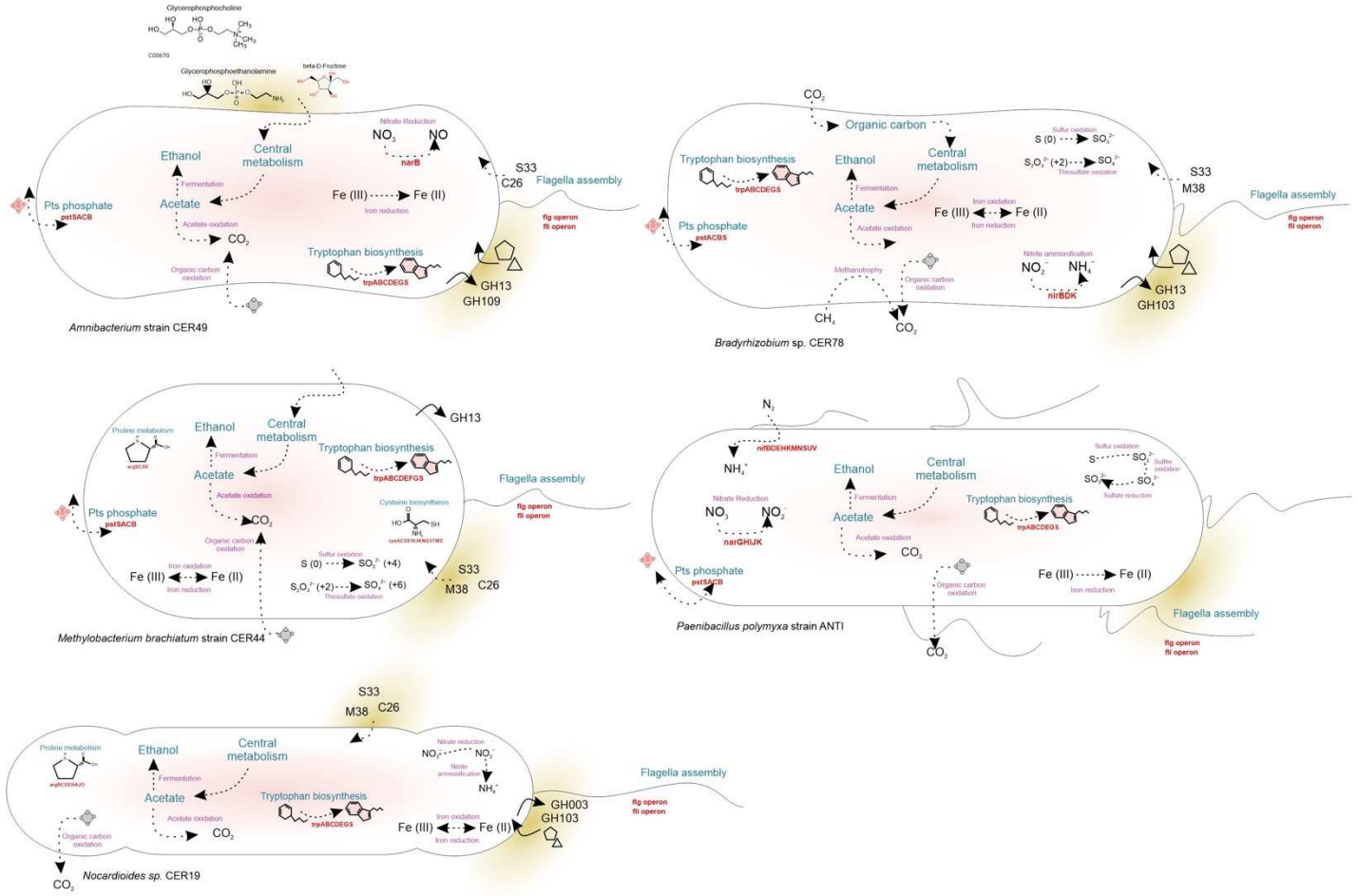


Figure 5. Genome-based predictions of the potential metabolic role of the five-novel slow-growing bacterial species. The metabolic features of the novel bacteria are highlighted. Central metabolism includes glycolysis, gluconeogenesis, and the TCA cycle at some degree of completion among genomes (see Supplementary Table 6). Blue indicates the main metabolic function, and purple denotes the specific step in a biochemical reaction. Red color is used to indicate the main operon involved in the function. Outer curve lines are extracellular CAZyme and peptidase IDs.

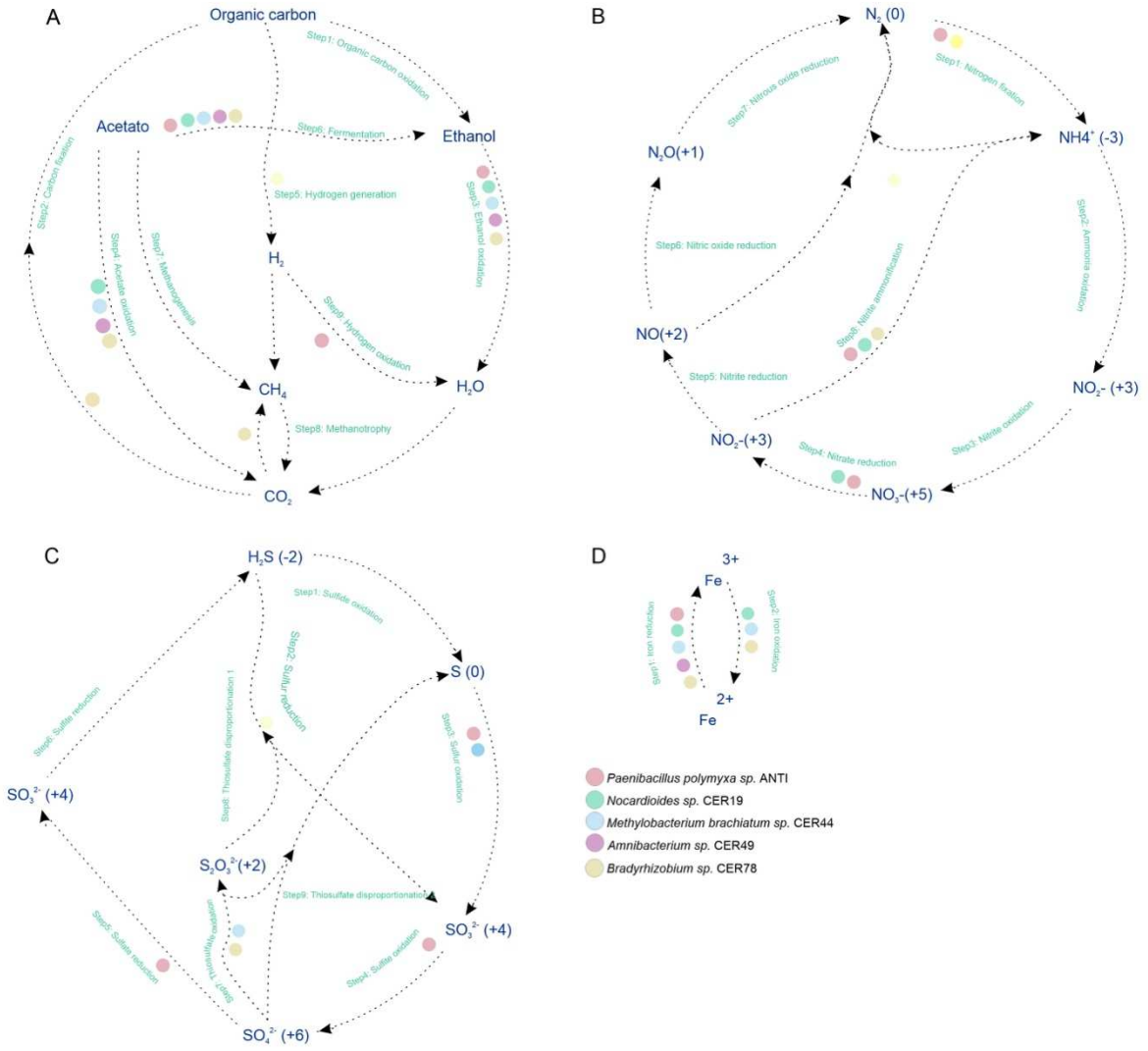


Figure 6. Potential biogeochemical cycling processes mapped in the genomes of slow-growing bacteria. Genomes are color-coded in each step within carbon (A), nitrogen (B), sulfur (C), and ferrous (D) cycles. Each arrow in the figure represents a single transformation or step within a cycle.

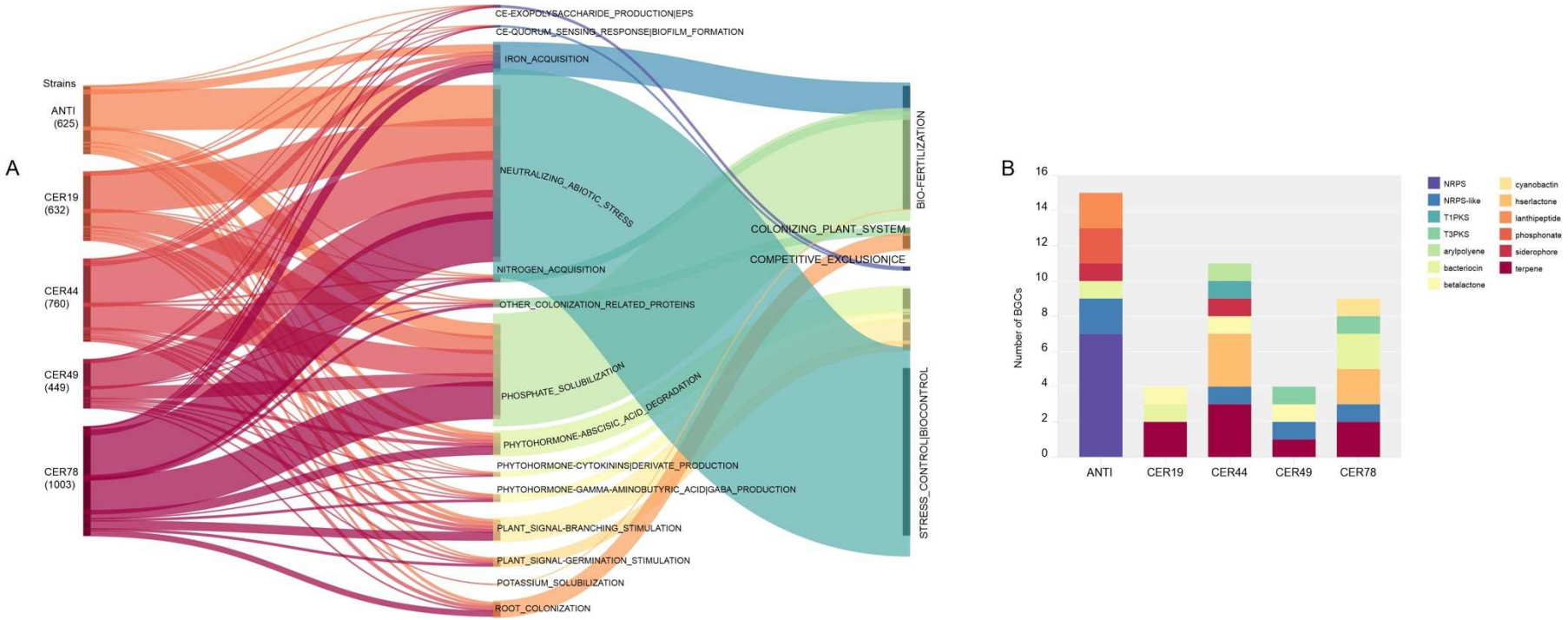


Figure 7. The biosynthetic potential of slow-growing bacteria for secondary metabolites and their role in promoting plant growth. (A) Sankey graph showing plant growth-promoting genes involved in biofertilization, colonization of the plant system, competitive exclusion, and stress control biocontrol. The number of genes found in total is indicated in parentheses under the strain code. (B) The number and class of BGCs in each species' genome.

Chapter 2

The Reverse Ecology-Based Approach to Design a Bacterial Consortium as Soybean Bioinoculant

Gonçalves, O.S., Veloso, T.G.R., Fernandes, A.S., Santana, M.F. The Reverse Ecology-Based Approach to Design a Bacterial Consortium as Soybean Bioinoculant. The manuscript will be submitted as a short communication to Microbiological Research.

A Reverse Ecology-Based Approach to Design a Bacterial Consortium as Soybean Bioinoculant

Osiel S. Gonçalves[1], Tomás G. R. Veloso[1], Alexia S. Fernandes[1], Mateus F. Santana[1*]

[1]Grupo de Genômica Eco-evolutiva Microbiana, Laboratório de Genética Molecular de Microrganismos, Departamento de Microbiologia, Instituto de Biotecnologia Aplicada à Agropecuária, Universidade Federal de Viçosa, Minas Gerais, Brazil.

*Correspondence authors: mateus.santana@ufv.br

Highlights

- A microbial consortium was designed and evaluated as soybean inoculants using reverse ecology.
- The consortium, composed of three soil bacteria, significantly improved soybean growth.
- The study provides a novel insight into the design of microbial consortia using a computational approach.

Abstract

Traditionally, bioinoculants for plant growth have relied on the selection of efficient microbes from the soil that exhibit potential traits for enhancing plant growth. However, this selection often overlooks microbe-microbe and microbe-plant interactions. In this study, we applied a reverse ecology framework to design and evaluate the composition of a bacterial consortium for soybean. Our analysis showed the efficacy of our microbial consortium composed of three soil bacteria strains, including *Paenibacillus polymyxa*, *Methylobacterium brachiatum*, and *Enterobacter*. By leveraging a comprehensive understanding of microbial-microbe interactions and plant-microbe interactions, we demonstrated through computational and *in vitro* analysis that the selected microbial strains lacked strong competitive abilities and metabolic compatibility with the host plant. In addition, we have successfully validated the effectiveness of our consortium in a greenhouse experiment. Taken together, this study provides a novel insight into the design of a microbial consortium using a computational approach and offers promising prospects for the development of bioinoculants with enhanced performance in promoting plant growth and productivity.

Keywords: bioprospection; drought; PGPB; plant-microbe interaction; soil bacteria.

Introduction

Traditionally, bioinoculants for plant growth have relied on the selection of efficient microbes from the soil that exhibit potential traits for enhancing plant growth (Kumar et al., 2022; Singh et al., 2021) These traits encompass both direct and indirect effects on plants, including the production of phytohormones, nitrogen fixation capabilities, organic and inorganic phosphate solubilization, protection against plant pathogens through the production of metabolite compounds, as well as the induction of plant resistance and the release of volatile organic compounds (Compant et al., 2010; Glick, 2012; Gonçalves et al., 2023; Kavamura et al., 2013; Souza et al., 2015). The conventional approach of using bioinoculants from soil has been a significant advancement in the field, but it has limitations. It requires extensive time and effort for isolation and testing, often overlooking microbe-microbe and microbe-plant interactions. Therefore, it is crucial to consider these limitations and move towards a more comprehensive and holistic approach to harnessing the potential of bioinoculants.

High-throughput sequencing has revolutionized genomics by enabling rapid and cost-effective analysis of DNA sequences(Reuter et al., 2015) . This technology has greatly enhanced our understanding of genomes and their potential applications in various fields, including the harnessing of novel plant-growth bacteria (Bruto et al., 2014; Gonçalves et al., 2023; Imam et al., 2016). It allows for the efficient sequencing of bacterial genomes, providing valuable insights into their genetic composition, functional traits, and metabolic capabilities. In general, this centric-genomic approach is rooted in the concept of reverse ecology. Reverse ecology represents a frontier in evolutionary systems biology, offering critical information about an organism's metabolic capabilities and shedding new light on its ecological dynamics(Levy and Borenstein, 2012).

Reverse ecology relies on several computational approaches that to translate high-throughput genetic data into large-scale ecological data, which potentially turns ecology into a high-throughput field (Levy and Borenstein, 2012). It was developed by demonstrating the concept that a microorganism's metabolic activity is directly linked to the biochemical environment (Freilich et al., 2010, 2009; Levy and Borenstein, 2013; Parter et al., 2007). In addition, reverse ecology approaches have been used in different fields, including microbial-plant interactions to uncover metabolic environments (Karpinets et al., 2014; Ofaim et al., 2017); in human health, to quantify microbes' biosynthetic capabilities across the human oral microbiome (Bernstein et

al., 2019), to identify immune-beneficial infant gut bacteria by mining their metabolism for prebiotic feeds (Michelini et al., 2018), and more recently to uncover the potential metabolic interplay among plant bacteria pathogen and its hosts (Gonçalves et al., 2022).

We have conducted a recent study in which we isolated a remarkable collection of 92 isolates from the tropical savannah Cerrado, primarily consisting of slow-growing bacteria (Gonçalves and Santana, 2023). Our findings demonstrated that certain strains exhibited the ability to thrive in low water conditions and possess distinctive features that may support plant growth. To further explore these promising characteristics, we applied the concept of reverse ecology to design and evaluate the composition of a bacterial consortium. This consortium was carefully assembled, incorporating three specific strains known to foster syngenetic interactions with soybean plants, while ensuring the inclusion of complementary strains that enhance their collective benefits.

Material and Methods

Bacterial strains and Growth conditions

In our study, we carefully chose eight bacterial strains that were previously isolated from the tropical savannah Cerrado. These strains included *Paenibacillus polymyxa* strain ANTI, *Nocardioides* strain CER19, and *Arthrobacter* sp. CER28, *Methylobacterium brachiatum* strain CER44, *Amnibacterium* strain CER49, *Enterobacter* sp. CER55, *Mesorhizobium* sp. CER75, and *Bradyrhizobium* strain CER78. The strains were cultivated in a liquid R2A medium for reactivation. The composition of the medium per liter of distilled water included 0.5 g of yeast extract, 0.5 g of peptone, 0.5 g of casamino acids, 0.5 g of glucose, 0.5 g of soluble starch, 0.5 g of pyruvate, 0.3 g of K_2HPO_4, and 0.05 g of $MgSO_4.7H_2O$. The cultivation was performed at 28 °C and 180 rpm for one week. Afterward, streaking onto solid R2A medium.

Plant Growth Promotion in Greenhouse Experiment

Soybean cultivar Brevant® DS5916IPRO seeds underwent surface sterilization before being inoculated with eight different strains. The inoculation process involved mixing the seeds with the inoculum for one hour, with a concentration of 10^7 to 10^8 CFU mL^{-1} (measured at DO550 = 0.1 or 0.5). To establish a control treatment, the seeds were mixed with a phosphate-buffered saline (PBS) 1X (137 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4, and 1.8 mM KH2PO4) pH 7. The seedlings were grown in plastic trays filled with Tropstrato®, and were exposed to natural sunlight in a greenhouse. The average daytime temperature ranged from 12 to 33 °C. Throughout the growth period until the emergence of the first trifoliate (stage V1) leaves, the

soybean plants were watered daily with the same volume of water. Subsequently, a water restriction treatment was implemented using PEG (Polyethylene Glycol) at a concentration of 10%. Evaluation of the plants was conducted two weeks after the initiation of water restrictions. Parameters such as leaf area, number of nodes, shoot and root lengths, as well as shoot and root dry biomass, were measured and recorded. For the greenhouse experiments, a completely randomized design was employed. The collected data were subjected to one-way ANOVA, and the Skott–Knott clustering algorithm was applied using the R package *Easyanova* to analyze the results.

Microbe-Microbe and Plant-Microbe interactions using Reverse Ecology

The genome sequences of strains ANTI and CER44 were obtained from the National Center for Biotechnology Information (NCBI) with the access numbers JARVWT000000000 and JARVWR000000000, respectively. For CER55, we used the closest placement reference genome of *Enterobacter kobei* strain DSM 13645 under RefSeq assembly accession GCF_001729765.1. To annotate the protein-level assemblies of these genomes, we employed KofamKOALA (Aramaki et al., 2020) with the following parameters: '--e-value 0.00001'. The resulting KEGG orthologs (KO) from these annotations were then utilized to determine microbe-microbe interactions, specifically focusing on competition and complementarity indices. For calculating these indices, we used the Cooperation Index package within RevEcoR (Cao et al., 2016). To further analyze the potential ecological host-microbe interactions, we created a matrix containing substrate and product information for each species based on the RevEcoR analysis. Subsequently, we employed NetCooperate (Levy et al., 2015) to gain insights into these interactions. Specifically, for pairwise interactions between hosts and microbes, we measured the biosynthetic support score (BSS) and the metabolic complementarity index (MCI). The compounds involved were annotated in KEGG compounds, including their biological roles, as well as in the Phytochemical Compounds Database and the MetaCyc Metabolic Pathway Database (Caspi et al., 2020).

Co-culture experiment

To facilitate pairwise interactions between strains ANTI, CER44, and CER55, they were streaked onto solid R2A medium in four quadrants, with each strain occupying one quadrant. The plates were then incubated at 28 °C for one week, and the growth of the strains was monitored daily. To investigate the potential for growth inhibition by volatile organic compounds, the plates were sealed with parafilm. This sealing mechanism aimed to prevent the

escape of any volatile compounds that might be produced by the strains during their growth on the medium.

Greenhouse Experiment to Validate the RevEco Approach for Plant Growth Promotion

Soybean cultivar BMX® LANÇA 58I60 RSF IPRO seeds were surface-sterilized and inoculated with ANTI, CER44, and CER55 by mixing for 1 h in the inoculum ($10^7 \sim 10^8$ colony-forming units (CFU) mL^{-1} (DO550 = 0.5 for the first two strains and 0.1 for CER44). As a control treatment, the seeds were mixed with sterilized PBS 1X. Additionally, a consortium (CSC) was created by mixing 10^2 CFU mL^{-1} of CER55, 10^3 mL^{-1} of ANTI, and 10^3 mL^{-1} of CER44. The seedlings were then grown in individual pots filled with agricultural soil, and placed in a greenhouse that received natural sunlight, with an average daytime temperature ranging from 12 to 25 °C. Daily watering of the soybean plants was carried out until the end of stage V2. Measurements were taken for shoot and root lengths, as well as fresh and dry biomass. The greenhouse experiments were conducted using a completely randomized design. The collected data were subjected to one-way ANOVA, and the Skott–Knott clustering algorithm was applied using the R package *Easyanova* to analyze the results.

Results and Discussions

Selection of strains with the potential to enhance soybean growth in a greenhouse experiment.

We conducted an initial evaluation to assess the effectiveness of eight strains in promoting soybean growth under both water and drought conditions. An interaction between the factors of water condition and type of inoculum was observed for the leaf area of the soybean. This means that the choice of inoculum to prevent leaf area loss depends on the water condition (p-value = 0.035216). In the well-watered treatment, the leaf area remained consistent regardless of the inoculum used (p-value = 0.072). However, under PEG-induced drought, the selection of inoculum becomes significant (p-value = 0.016), and it is advisable to avoid using the ANTI and CER49 (p-value = 0.016) (Fig. 1).

Significant effects of the inoculum on the dry weight of the root system were found (p-value = 0.014). The CNT, CER44, and CER28 resulted in the greatest weight gains for the root system, regardless of the water treatment conditions. No significant differences were observed in root length among the inoculums in well-watered treatment. However, in the presence of PEG-induced drought deficit, plants treated with CER28, CER44, CER49, CER55, and CER75 showed the highest root growth compared to other treatments (Fig. 1). Additionally, the dry

weight of the shoot also exhibited a significant effect of the inoculum (p-value = 0.014), and an interaction between the water condition and the type of inoculum used (p-value = 0.00348). The Scott-Knott test revealed that there were no differences between the inoculums under the condition of water stress induced by PEG. However, in the absence of stress, the control treatments, ANTI, CER28, and CER44 showed the highest dry weights (p-value = 0.00057). Plants inoculated with CER19, CER28, CER44, CER49, CER55, CER75, and had significantly higher dry weights (p-value = 0.0023) compared to soybean inoculated with ANTI, CER76, and non-inoculated (CNT) (Fig. 1).

The results of this experiment revealed a complex and intricate interaction between the water condition and the type of inoculum for each factor. Overall, most strains did not exhibit significant improvements in soybean growth under well-watered conditions compared to the control treatment. However, in contrast, most of the strains demonstrated significant enhancements in both root and shoot length of soybean plants under PEG-induced drought conditions. It is well-documented that during water shortages, plants undergo shifts in root system development to enhance their access to water. In this study, the observed increase in root length supports this notion. Additionally, it has been suggested that some plants respond to drought by stimulating or maintaining root growth while reducing shoot growth, the "balanced growth" hypothesis (Bloom et al., 1985). Interestingly, in contrast to this hypothesis, plants inoculated with the strain exhibited increased shoot growth. This could be attributed to the strains' ability to produce auxins, which are known to promote plant stem development (Gallavotti, 2013).

Designing a microbial consortium using reverse ecology

We next selected specific bacterial strains to create a soybean bioinoculant using a reverse ecology framework. These strains include *M. brachiatum* strain CER44, which demonstrated a significant increase in soybean growth. Additionally, we incorporated *Enterobacter* sp. CER55 is known for its remarkable capacity to produce phytohormones and solubilize phosphate, as well as *P. polymyxa* strain ANTI, an outstanding biocontrol agent and plant growth-promoting rhizobacterium. This specie has a broad host range and exhibits essential characteristics such as nitrogen fixation, phosphate solubilization, siderophore production, and the production of volatile organic compounds (Pandey et al., 2023; Salme et al., 2005; Singh and Wesemael, 2022; Soni et al., 2021).

We observed that the strains in our study did not exhibit significant differences in their competitiveness index. The highest score observed was 0.6, which occurred in the pairwise interaction between CER55 and CER44. This suggests that these two strains have similar metabolic capabilities, as indicated by their complementarity index of 0.1 (Fig. 2A). This similarity implies that competition between these strains may occur when resources are scarce. However, despite their metabolic similarities, CER55 and CER44 exhibit different growth characteristics. CER55 is considered a copiotroph, characterized by rapid growth, typically within 24 hours, in nutrient-rich culture media. On the other hand, CER44 is classified as an oligotroph, exhibiting slower growth that is better suited for low-nutrient culture media (Gonçalves and Santana, 2023; Koch, 2001). Interestingly, our co-culture assay demonstrated that these strains could grow together without any inhibition caused by secondary metabolites or volatile organic compounds (Fig. 2B). This finding is particularly noteworthy, considering that *P. polymyxa* strain ANTI possesses numerous biosynthetic gene clusters encoding antimicrobial potential (Gonçalves and Santana, 2023).

The Biosynthetic Support Score (BSS) exhibited a range of 0.4 to 0.8 across microbe-microbe interactions (Fig. 2C). When considering the strains as an individual metabolic consortium (CSC), the BSS score was found to be similar to that of soybean strains, as was the case with the Complementarity Index (MCI) (Fig. 2C). These two metrics rely on the host species' ability to fulfill the nutritional requirements of a parasitic or commensal species and support each other through biosynthetic complementarity (Levy et al., 2015). In the interaction among the strains, we identified 277 BSS compounds (106 unique compounds), with most of these compounds belonging to the lipid, carbohydrate, and amino acid classes (Fig. 2D). A similar result was observed for lipid-associated BSS compounds in the plant-microbe interaction (Supplementary Fig S1).

Lipids play a crucial role in rhizosphere interactions, facilitating communication between plant roots, and microbes, and even modulating the plant's defense responses upon interaction with beneficial or phytopathogenic microorganisms (Macabuhay et al., 2022). This suggests that these strains possess metabolic capabilities to modulate synergetic interactions with plants through similar complementarity. An interesting finding was the identification of a supportive interaction involving the compound mannitol between CER55 and ANTI. Mannitol, a sugar alcohol, serves as a carbon and energy source for *M. brachiatum* strain CER44. The genus *Methylobacterium* is commonly associated with the leaf surfaces of plants, known as the phyllosphere. Mannitol is abundant on plant leaf surfaces as a by-product of pectin metabolism

during cell wall synthesis (Abanda-Nkpwatt et al., 2006; MacDonald and Fall, 1993; Nemecek-Marshall et al., 1995).

The validation of the consortium in the greenhouse experiment

In summary, our reverse ecology framework has yielded supportive results for our microbial consortium, leveraging comprehensive knowledge of microbe-microbe interactions and plant-microbe interactions. To further validate the potential of our reverse ecology consortium, we conducted a greenhouse experiment to assess its ability to enhance plant growth compared to individual strains. By employing this approach, we aim to demonstrate the collective benefits and synergistic effects that can arise from the interactions within our consortium.

In contrast to the first experiment, the results of our second experiment showed that all individual strains had a positive impact on soybean growth. We attribute this outcome to the characteristics of the soil used in this experiment. The commercial substrate, as used in the first greenhouse experiment, provided a nutrient-rich environment for the plants, potentially influencing the synergistic interaction between the plant and the microorganisms. This phenomenon has been well-documented in the case of mycorrhizal fungi (Huey et al., 2020; Mujica et al., 2016) or symbiotic association with *rhizobium* (de Castro Pires et al., 2018). Also, it is noteworthy that nitrogen nodules were found in all plants across the different treatments.

Significant effects of the inoculum were observed in terms of shoot length (Fig. 3A), root length (Fig. 3B), as well as fresh and dry weight. Interestingly, we noted that the CSC (consortium) treatment showed significant effects compared to the control plants, but not when compared to the individual strains. However, on average, the CSC treatment exhibited slightly better results than the individual strains. This observation is also evident in Fig 3C, where the plant architecture in the CSC treatment stands out compared to the other treatments. It is worth mentioning that the presence of ANTI, one of the strains in the consortium, did not significantly improve the root growth of soybeans and occasionally even had a slight negative impact on the performance of the consortium. This suggests that the absence of ANTI might enhance the overall effectiveness of the consortium.

Employing a microbial consortium offers several advantages over single-strain inoculants. The inclusion of diverse species within the consortium enables a wider range of functions and benefits for plants (Bradáčová et al., 2019). Furthermore, soil habitats are dynamic systems, and the majority of microbes present have evolved strategies to cope with changing

environmental conditions (Jansson and Hofmockel, 2020). Therefore, species with adaptive functional traits, such as the ability to exhibit r and k microbe strategies, such as employed in our consortia, can respond to various climate changes, and facilitate spatial-temporal coordination of functions that ultimately benefit plants.

Conclusions

Our reverse ecology framework has provided valuable insights into the microbe-microbe and plant-microbe interactions within our microbial consortium. Through a greenhouse experiment, we evaluated the potential of our consortium to enhance plant growth compared to individual strains. The results demonstrated significant effects of the inoculum on shoot length, root length, as well as fresh and dry weight of the plants. Interestingly, the consortium treatment exhibited notable effects when compared to the control plants, but not when compared to the individual strains. Further investigation is warranted to explore the underlying mechanisms and interactions within the consortium. Fine-tuning the composition of the consortium by considering the presence or absence of specific strains, such as ANTI, could potentially optimize its performance and yield even more beneficial effects on plant growth. In summary, this study has provided novel insights into the design of a microbial consortium using a computational approach, focusing on understanding microbe-microbe and microbe-plant interactions comprehensively. By leveraging computational tools and knowledge of these interactions, we have gained a new perspective on the construction of microbial consortia.

CRediT authorship contribution statement

Osiel S. Goncalves: Writing – original draft, Conceptualization, Data Curation, Investigation, Project administration, Visualization, Writing – review & editing. Tomás G. R. Veloso: data processing, Writing – review & editing. Alexia S. Fernandes: Data Curation, Investigation Mateus F. Santana: Conceptualization, Supervision, Validation, Funding acquisition, Project administration, Writing - review & editing.

Declaration of Competing Interest

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

Acknowledgments

This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq (Process APQ-02381-21), Coordenação de Aperfeiçoamento de Pessoal de

Nível Superior/Programa de Excelência Acadêmica-Finance Code 001 (CAPES ProEx grant 23038.019105/2016-86) and Fundação de Amparo à Pesquisa do Estado de Minas Gerais—FAPEMIG (Process 402644/2021-2) for the financial support.

References

Abanda-Nkpwatt, D., Müsch, M., Tschiersch, J., Boettner, M., Schwab, W., 2006. Molecular interaction between *Methylobacterium extorquens* and seedlings: growth promotion, methanol consumption, and localization of the methanol emission site. J Exp Bot 57, 4025–4032. https://doi.org/10.1093/jxb/erl173

Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., Ogata, H., 2020. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics 36, 2251–2252. https://doi.org/10.1093/bioinformatics/btz859

Bernstein, D.B., Dewhirst, F.E., Segrè, D., 2019. Metabolic network percolation quantifies biosynthetic capabilities across the human oral microbiome. Elife 8, 1–33. https://doi.org/10.7554/eLife.39733.001

Bloom, A.J., Chapin, F.S., Mooney, H.A., 1985. Resource Limitation in Plants-An Economic Analogy. Annu Rev Ecol Syst 16, 363–392. https://doi.org/10.1146/annurev.es.1 6.110185.002051

Bradáčová, K., Florea, A., Bar-Tal, A., Minz, D., Yermiyahu, U., Shawahna, R., Kraut-Cohen, J., Zolti, A., Erel, R., Dietel, K., Weinmann, M., Zimmermann, B., Berger, N., Ludewig, U., Neumann, G., Gheorghe, P., 2019. Microbial Consortia versus Single-Strain Inoculants: An Advantage in PGPM-Assisted Tomato Production? Agronomy 9, 105. https:// doi.org/10.3390/agronomy9020105

Bruto, M., Prigent-Combaret, C., Muller, D., Moënne-Loccoz, Y., 2014. Analysis of genes contributing to plant-beneficial functions in plant growth-promoting rhizobacteria and related Proteobacteria. Sci Rep 4, 6261. https://doi.org/10.1038/srep06261

Cao, Y., Wang, Y., Zheng, X., Li, F., Bo, X., 2016. RevEcoR: an R package for the reverse ecology analysis of microbiomes. BMC Bioinformatics 17, 294. https://doi.org/10.1186/s12859-016-1088-4

Caspi, R., Billington, R., Keseler, I.M., Kothari, A., Krummenacker, M., Midford, P.E., Ong, W.K., Paley, S., Subhraveti, P., Karp, P.D., 2020. The MetaCyc database of metabolic pathways and enzymes - a 2019 update. Nucleic Acids Res 48, D445–D453. https://doi.org/ 10.1093/nar/gkz862

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 Biol Biochem 42, 669–678. https://doi.org/10.1016/j.soilbio.2009.11.024

de Castro Pires, R., dos Reis Junior, F.B., Zilli, J.E., Fischer, D., Hofmann, A., James, E.K., Simon, M.F., 2018. Soil characteristics determine the rhizobia in association with different species of Mimosa in central Brazil. Plant Soil 423, 411–428. https://doi.org/10.1007/s11104-017-3521-5

Freilich, S., Kreimer, A., Borenstein, E., Yosef, N., Sharan, R., Gophna, U., Ruppin, E., 2009. Metabolic-network-driven analysis of bacterial ecological strategies. Genome Biol 10, R61. https://doi.org/10.1186/gb-2009-10-6-r61

Freilich, S., Kreimer, A., Meilijson, I., Gophna, U., Sharan, R., Ruppin, E., 2010. The large-scale organization of the bacterial network of ecological co-occurrence interactions. Nucleic Acids Res 38, 3857–3868. https://doi.org/10.1093/nar/gkq118

Gallavotti, A., 2013. The role of auxin in shaping shoot architecture. J Exp Bot 64, 2593–2608. https://doi.org/10.1093/jxb/ert141

Glick, B.R., 2012. Plant Growth-Promoting Bacteria: Mechanisms and Applications. Scientifica (Cairo) 2012, 963401. https://doi.org/10.6064/2012/963401

Gonçalves, O.S., Souza, T.S., Gonçalves, G. de C., Fernandes, A.S., Veloso, T.G.R., Tupy, S.M., Garcia, E.A., Santana, M.F., 2023. Harnessing Novel Soil Bacteria for Beneficial Interactions with Soybean. Microorganisms 11. https://doi.org/10.3390/microorganisms11020300

Gonçalves Silva, O., Bonandi Barreiros, R., Martins Tupy, S., Ferreira Santana, M., 2022. A reverse-ecology framework to uncover the potential metabolic interplay among '*Candidatus Liberibacter'* species, citrus hosts and psyllid vector. Gene 837, 146679. https://doi.org/https://doi.org/10.1016/j.gene.2022.146679

Huey, C.J., Gopinath, S.C.B., Uda, M.N.A., Zulhaimi, H.I., Jaafar, M.N., Kasim, F.H., Yaakub, A.R.W., 2020. Mycorrhiza: a natural resource assists plant growth under varied soil conditions. 3 Biotech 10, 204. https://doi.org/10.1007/s13205-020-02188-3

Imam, J., Singh, P.K., Shukla, P., 2016. Plant Microbe Interactions in Post Genomic Era: Perspectives and Applications. Front Microbiol 7, 1488. https://doi.org/ 10.3389/fmicb.2016.01488

Jansson, J.K., Hofmockel, K.S., 2020. Soil microbiomes and climate change. Nat Rev Microbiol 18, 35–46. https://doi.org/10.1038/s41579-019-0265-7

Karpinets, T. V, Park, B.H., Syed, M.H., Klotz, M.G., Uberbacher, E.C., 2014. Metabolic Environments and Genomic Features Associated with Pathogenic and Mutualistic Interactions Between Bacteria and Plants. Molecular Plant-Microbe Interactions® 27, 664–677. https://doi.org/10.1094/MPMI-12-13-0368-R

Kavamura, V.N., Santos, S.N., Silva, J.L. da, Parma, M.M., Ávila, L.A., Visconti, A., Zucchi, T.D., Taketani, R.G., Andreote, F.D., Melo, I.S. de, 2013. Screening of Brazilian cacti rhizobacteria for plant growth promotion under drought. Microbiol Res 168, 183–191. https://doi.org/10.1016/j.micres.2012.12.002

Koch, A.L., 2001. Oligotrophs versus copiotrophs. BioEssays 23, 657–661. https:// doi.org/10.1002/bies.1091

Kumar, P., Singh, S., Pranaw, K., Kumar, S., Singh, B., Poria, V., 2022. Bioinoculants as mitigators of multiple stresses: A ray of hope for agriculture in the darkness of climate change. Heliyon 8, e11269. https://doi.org/10.1016/j.heliyon.2022.e11269

Levy, R., Borenstein, E., 2013. Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. Proceedings of the National Academy of Sciences 110, 12804 LP – 12809. https://doi.org/10.1073/pnas.1300926110

Levy, R., Borenstein, E., 2012. Reverse Ecology: from systems to environments and back. Adv Exp Med Biol 751, 329–345. https://doi.org/10.1007/978-1-4614-3567-9_15

Levy, R., Carr, R., Kreimer, A., Freilich, S., Borenstein, E., 2015. NetCooperate: a network-based tool for inferring host-microbe and microbe-microbe cooperation. BMC Bioinformatics 16, 164. https://doi.org/10.1186/s12859-015-0588-y

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 Plant Sci 27, 180–190. https://doi.org/10.1016/j.tplants.2021.08.004

MacDonald, R.C., Fall, R., 1993. Detection of substantial emissions of methanol from plants to the atmosphere. Atmospheric Environment. Part A. General Topics 27, 1709–1713. https://doi.org/https://doi.org/10.1016/0960-1686(93)90233-O

Michelini, S., Balakrishnan, B., Parolo, S., Matone, A., Mullaney, J.A., Young, W., Gasser, O., Wall, C., Priami, C., Lombardo, R., Kussmann, M., 2018. A reverse metabolic approach to weaning: In silico identification of immune-beneficial infant gut bacteria, mining their metabolism for prebiotic feeds and sourcing these feeds in the natural product space. Microbiome 6, 1–18. https://doi.org/10.1186/s40168-018-0545-x

Mujica, M.I., Saez, N., Cisternas, M., Manzano, M., Armesto, J.J., Pérez, F., 2016. Relationship between soil nutrients and mycorrhizal associations of two *Bipinnula* species (*Orchidaceae*) from central Chile. Ann Bot 118, 149–158. https://doi.org/10.1093/aob/mcw082

Nemecek-Marshall, M., MacDonald, R.C., Franzen, J.J., Wojciechowski, C.L., Fall, R., 1995. Methanol emission from leaves (enzymatic detection of gas-phase methanol and relation of methanol fluxes to stomatal conductance and leaf development). Plant Physiol 108, 1359–1368.

Ofaim, S., Ofek-Lalzar, M., Sela, N., Jinag, J., Kashi, Y., Minz, D., Freilich, S., 2017. Analysis of Microbial Functions in the Rhizosphere Using a Metabolic-Network Based Framework for Metagenomics Interpretation. Frontiers in Microbiology.

Pandey, A.K., Barbetti, M.J., Lamichhane, J.R., 2023. *Paenibacillus polymyxa*. Trends Microbiol 31, 657–659. https://doi.org/10.1016/j.tim.2022.11.010

Parter, M., Kashtan, N., Alon, U., 2007. Environmental variability and modularity of bacterial metabolic networks. BMC Evol Biol 7, 169. https://doi.org/10.1186/1471-2148-7-169

Reuter, J.A., Spacek, D. V, Snyder, M.P., 2015. High-Throughput Sequencing Technologies. Mol Cell 58, 586–597. https://doi.org/https://doi.org/10.1016/j.molcel.2015.05.004

Salme, T., Nina, G., H, W.E.G., 2005. *Paenibacillus polymyxa* Invades Plant Roots and Forms Biofilms. Appl Environ Microbiol 71, 7292–7300. https://doi.org/10.1128/AEM.71.11. 7292-7300.2005

Singh, M., Bhasin, S., Madan, N., Suyal, D.C., Soni, R., Singh, D., 2021. Bioinoculants for Agricultural Sustainability, in: Soni, R., Suyal, D.C., Bhargava, P., Goel, R. (Eds.), Microbiological Activity for Soil and Plant Health Management. Springer Singapore, Singapore, pp. 629–641. https://doi.org/10.1007/978-981-16-2922-8_25

Singh, R.R., Wesemael, W.M.L., 2022. Endophytic *Paenibacillus polymyxa* LMG27872 inhibits *Meloidogyne incognita* parasitism, promoting tomato growth through a dose-dependent effect. Front Plant Sci 13.

Soni, R., Rawal, K., Keharia, H., 2021. Genomics assisted functional characterization of *Paenibacillus polymyxa* HK4 as a biocontrol and plant growth promoting bacterium. Microbiol Res 248, 126734. https://doi.org/10.1016/j.micres.2021.126734

Souza, R., Ambrosini, A., Passaglia, L.M.P., de Souza, R., Ambrosini, A., Passaglia, L.M.P., 2015. Plant growth-promoting bacteria as inoculants in agricultural soils. Genet Mol Biol 38, 401–419. https://doi.org/10.1590/S1415-475738420150053

Figures

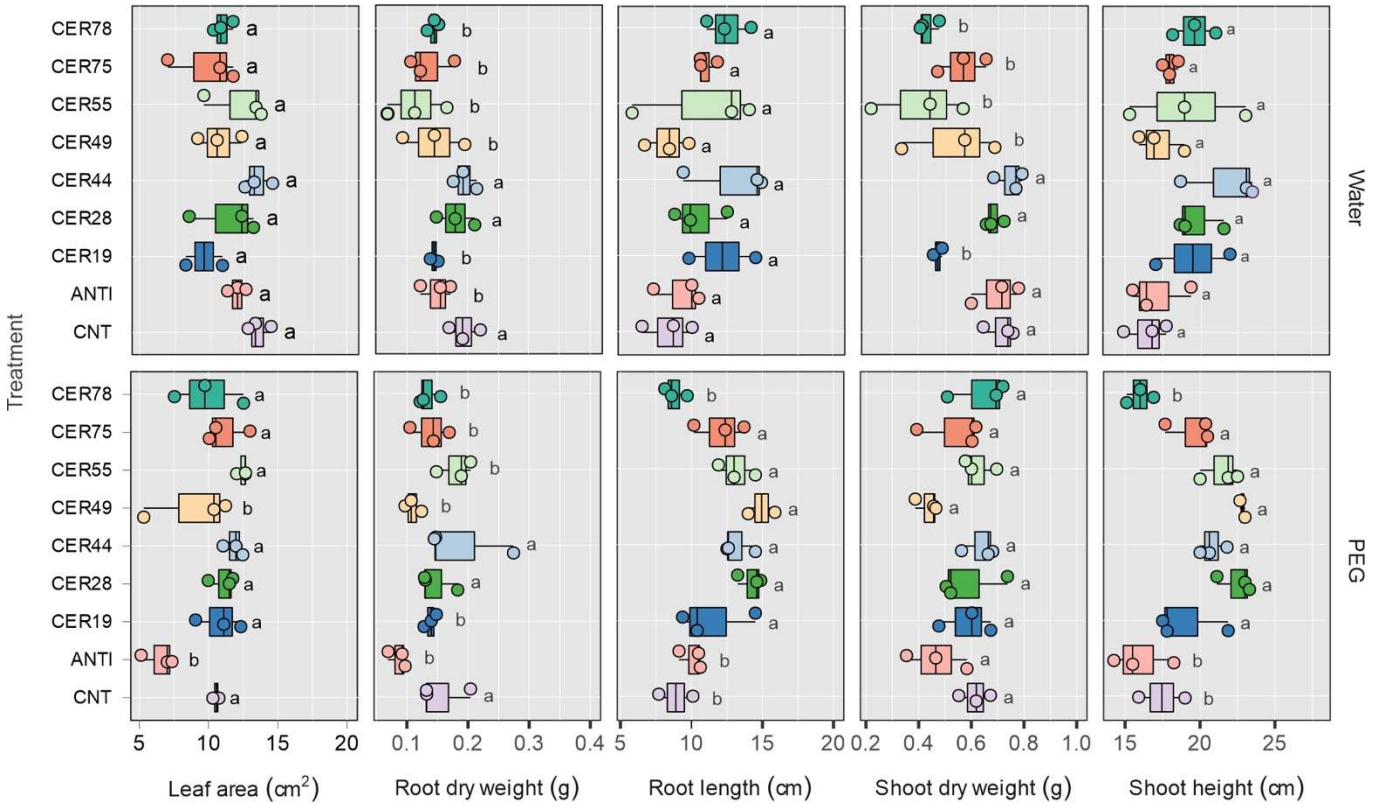


Figure 1. **Soybean growth under water and PEG-induced drought condition in a greenhouse experiment**. Phenotypic traits were measured in soybean plants under water conditions (upper graph) and PEG-induced drought (lower graph) and inoculation with eight bacterial strains. Treatments that shared the same lowercase letter within each water deficit condition did not exhibit statistically significant differences according to the Skott-Knott test.

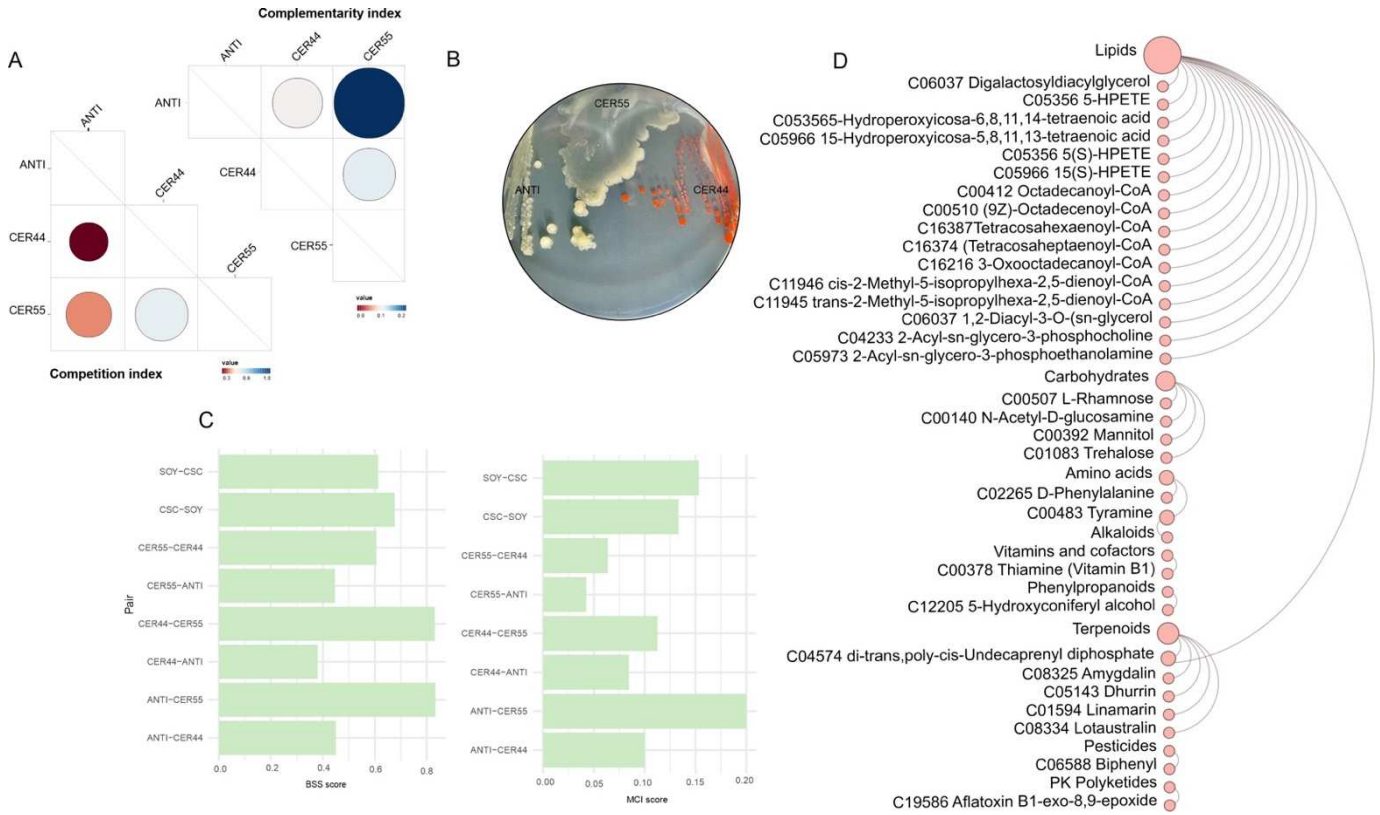


Figure 2. A reverse ecology approach to depict the genomic potential of microbe-microbe and plant-microbe interactions. A) Matrix plot of the competition and the complementary indices among three species of the microbial consortium. The competition index ranges from 0 (indicating the lowest competition) to 1 (representing the highest competition index). The size and color of each circle correspond to the competition and complementary index. B) Co-culture microbe assay depicts in vitro interaction of the three species in the microbial consortium. C) The distribution of the biosynthetic support score (BSS) and metabolic complementary (MCI) for the microbe and plant pairwise interactions. The Y-axes represent the pairwise interactions and indicate that the first microbe/plant is metabolically supported or complemented by the second microbe/plant. D) Detailed information regarding the biosynthetic compounds found in the interaction between the bacterial strains. Circles represent the proportion of each class, while arcs depict the set of compounds within each class.

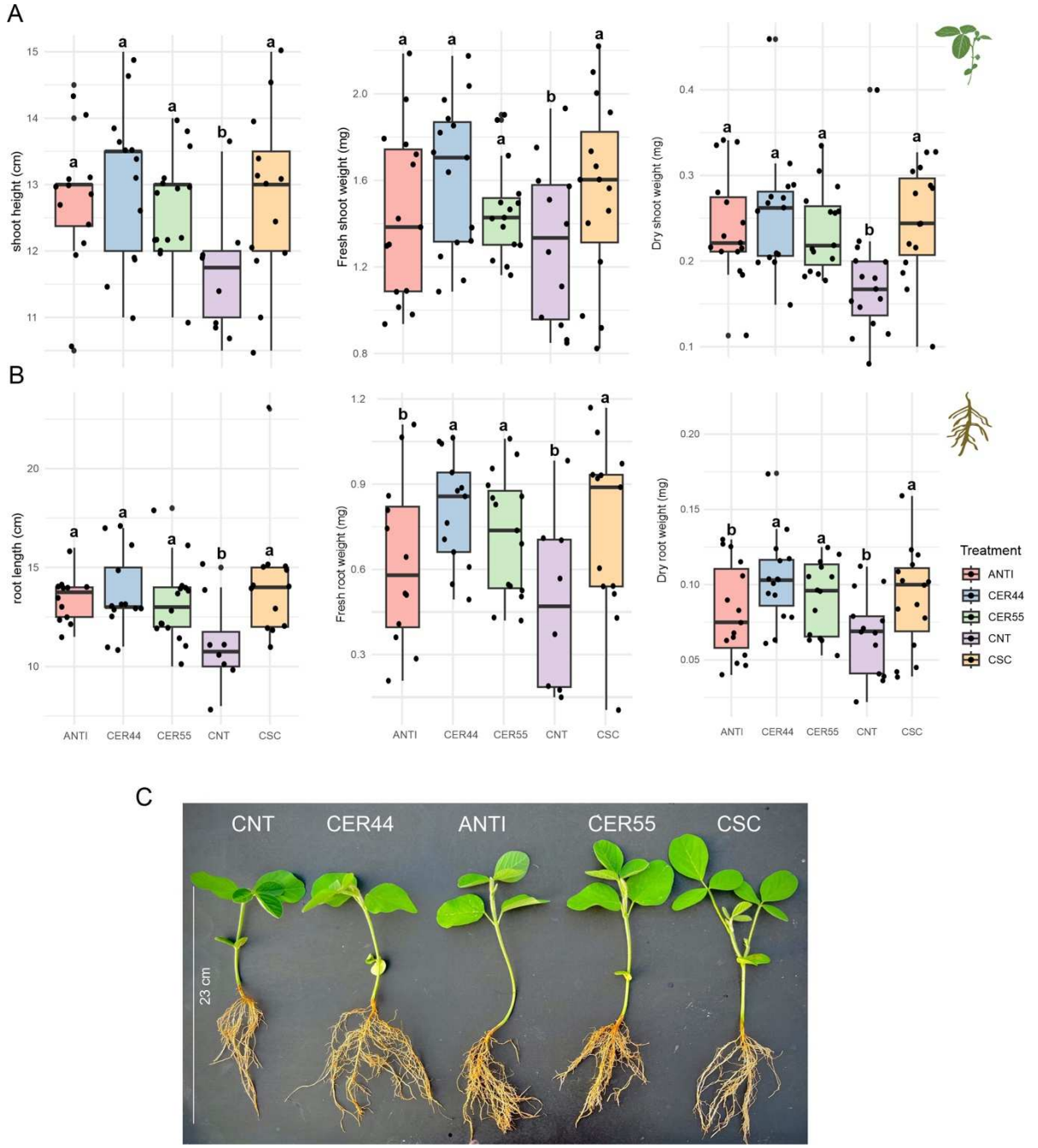


Figure 3. Validation of microbial consortium designed by reverse ecology approach to enhance soybean growth in a greenhouse experiment. Phenotypic traits of shoot (A) and root (B) were measured in soybean plants under four treatment conditions: without inoculation (Control, CNT), individually inoculated with three strains, and inoculated with the consortium (CSC). Treatments that shared the same lowercase letter within each water deficit condition did not show statistically significant differences according to the Skott-Knott test. C, Visual comparison of soybean plant across different treatments after the end of V2. The left side scale indicates the actual size of the plants in centimeters.

Chapter 3

Designing a Synthetic Microbial Community through Genome Metabolic Modeling to enhance Plant-Microbe Interactions

Gonçalves, O.S., Creevey C.J., Santana, M.F. Designing a Synthetic Microbial Community through Genome Metabolic Modeling to enhance Plant-Microbe Interactions. The manuscript will be submitted to Environmental Microbiome.

Designing a Synthetic Microbial Community through Genome Metabolic Modeling to enhance Plant-Microbe Interaction

Osiel S. Gonçalves[1], Christopher J. Creevey[2], Mateus F. Santana[1*]

[1]Grupo de Genômica Eco-evolutiva Microbiana, Laboratório de Genética Molecular de Microrganismos, Departamento de Microbiologia, Instituto de Biotecnologia Aplicada à Agropecuária, Universidade Federal de Viçosa, Minas Gerais, Brazil.

[2]School of Biological Sciences, Institute for Global Food Security, Queen's University Belfast, Belfast, BT9 5DL, UK.

*Correspondence authors: mateus.santana@ufv.br

Abstract

Background

Manipulating the rhizosphere microbial community through beneficial microorganism inoculation has gained interest to improve crop productivity and stress resistance. Synthetic microbial communities, known as SynCom, mimic natural microbial compositions while reducing the number of components. However, achieving this goal requires a comprehensive understanding of natural microbial communities and a careful selection of compatible microorganisms with colonization traits, which still pose challenges. In this study, we employed an *in-silico* approach using genome metabolic modeling to design a synthetic microbial community aimed at improving the yield of important crop plants.

Results

We used a targeted approach to select a minimal community (MinCom) that encompassed essential compounds for microbial metabolism and compounds relevant to plant interactions. This resulted in a reduction of the initial community size by approximately 4.5-fold. Importantly, the MinCom retained crucial genes associated with essential plant growth-promoting traits, such as iron acquisition, EPS production, potassium solubilization, nitrogen fixation, GABA production, and IAA-related tryptophan metabolism. Furthermore, our selection process for the SymCom, based on a comprehensive understanding of microbe-

microbe-plant interactions, yielded a set of six hub species that displayed notable taxonomic novelty, including members of the Eremiobacterota and Verrucomicrobiota phyla.

Conclusion

Overall, our study contributes to the growing body of research on synthetic microbial communities and their potential to enhance agricultural practices. The insights gained from our in-silico approach and the selection of hub species pave the way for further investigations into the development of tailored microbial communities that can optimize crop productivity and improve stress resilience in agricultural systems.

Keywords: Microbe-microbe-plant interactions; Plant growth-promoting bacteria; Reverse-ecology; SynComs.

Background

Plants and microbial communities have a complex and interdependent relationship, involving various mechanisms that influence ecological interactions. One crucial aspect is the release of photosynthates by plants belowground through mucilage and exudates that serve as energy sources for specific microbial groups [1–3]. In return, certain microbial taxa can positively impact plant growth and protect against biotic and abiotic stresses [4]. They achieve this through the synthesis of phytohormones, nutrient acquisition, and by engaging in antagonistic interactions with plant pathogens [5,6]. Consequently, it is intriguing to consider that plants experiencing drought can dynamically influence the eco-evolutionary dynamics of the microbial community and thereby affect plant performance [7].

In recent years, there has been growing interest in harnessing the potential of microbial communities in the rhizosphere to improve crop productivity and enhance stress resistance [8–10]. One approach gaining attention is the manipulation of the rhizosphere microbial community through the inoculation of beneficial microorganisms. This strategy aims to establish synthetic microbial communities (SynCom) that can positively influence plant health and yield. SynCom aim to mimic the original microbial composition by reducing the number of components in the community while preserving the essential characteristics of their natural counterparts [11].

The designing of SynCom that promote plant growth and help plants withstand various stresses, such as drought, salinity, and pathogen attacks, relies upon a careful selection and compatibility of specific microorganisms into the community that possesses traits for robust colonization,

prevalence throughout plant development and specific beneficial functions for plants [12–16]. Additionally, factors such as microbial diversity, community dynamics, and the stability of the introduced microorganisms need to be considered to ensure long-term effectiveness [16,17].

Decades of extensive cultivation-independent study, as well as recent high-throughput sequencing technology, have radically altered our perspectives on the diversity of microbial life [18,19]. These approaches capture the breadth of bacterial and archaeal genomic diversity across Earth's biomes and provide a resource that underscores the value of genome-centric approaches for revealing genomic properties of uncultivated microorganisms that affect ecosystem processes [20]. However, there are still significant limitations in answering the fundamental ecological and evolutionary questions surrounding natural microbial communities. In addition, genome-resolved metagenomics represents a valuable strategy for acquiring insights into the genomic properties, metabolic capabilities, and functional potential of individual microorganisms within complex microbial communities [21–24]. This enables a deeper understanding of their ecological roles, interactions, and contributions to various environmental processes.

The Cerrado, known for its arid conditions and nutrient-poor soils, harbors a diverse array of resilient plants with an associated microbiome [25–27]. A promising group of microbes that exhibit unique traits beneficial for plants by isolating and characterizing these bacteria. Therefore, harnessing novel soil bacteria from harsh environments like the Brazilian Cerrado holds immense potential for agricultural practices. These novel plant-promoting bacteria can enhance the growth, productivity, and stress tolerance of important crop plants such as soybean, maize, sorghum, and sugarcane. Notably, their ability to mitigate drought stress in these crops is of utmost importance, considering the increasing frequency of drought events caused by climate change [28,29].

In this study, our aims were threefold: (i) to reconstruct genome-scale metabolic networks (GSMNs) to gain insight into the metabolic complementarity between bacterial species and host crop plants; (ii) to define a minimal microbiota community; and (iii) to select hub-species that preserve the essential plant growth-promoting traits (PGPTs). To this end, we integrated several *in silico* approaches, employing multi-genome metabolic modeling of 270 previously described metagenome-assembled genomes (MAGs) from Campos rupestres, a grassland ecosystem located in Brazil.

Materials and methods

Genome data

We retrieved 270 MAGs belonging to two dominant plant species *Vellozia epidendroide*s and *Barbacenia macrantha* found in the Campos rupestres [30]. Annotated MAGs were downloaded from the National Center for Biotechnology Information (NCBI) under the BioProject PRJNA522264. MAGs were filtered based on co-assembly type to prevent data redundancy. MAGs details are provided in Supplementary Table S1.

We applied CheckM v1.0.13 [31] with the tree_qa command to extract a set of 43 single-copy, protein-coding marker genes. These marker genes were employed to evaluate phylogenetic markers within a dataset consisting of 270 assembled metagenomic bins. The concatenated protein alignments of the 43 universal marker genes, obtained from CheckM, were then used to reconstruct a maximum-likelihood phylogenetic tree using IQ-TREE v1.6.11 [32]. During the reconstruction process, specific parameters ('*-m TEST -bb 1000'*) were implemented to ensure the generation of an accurate tree. Subsequently, the phylogenetic tree was uploaded to iTOL [33], where it underwent visual annotation, including color coding and the application of heatmaps.

Reconstruction of the genome-scale metabolic networks (GSMNs)

The metabolic networks of each MAG were analyzed using genome-scale metabolic models reconstructed through an automated command-line version of PathwayTools [34]. The entire analysis was conducted using the metage2metabo (m2m) tool suite [35]. To begin the reconstruction process, we utilized the *mpwt* (multiprocessing pathway tools) program from the m2m tool with the "--path" flag to create a PathoLogic environment for each genome in gbff format. Subsequently, we performed the automatic reconstruction of non-curated metabolic networks using the *m2m recon* command line.

For the analysis of metabolic producibility and calculation of cooperation potential, we employed the *m2m iscope* command line for individual potential and the *m2m cscope* command line for collective metabolic potentials. Additionally, in the *m2m cscope* command, we included the "-m" flag to incorporate the host metabolic network in the sbml file format. The sbml files for maize (*Zea mays*), sugarcane (*Saccharum officinarum*), and sorghum (*Sorghum bicolor*) were obtained from C4GEM [36]. Furthermore, for soybean (*Glycine max*), we reconstructed non-curated GSMNs using the m2m recon command line for the genome of the *G. max* cultivar EMBRAPA BRS 537 downloaded from NCBI (Supplementary data 1). As a nutritional

constraint, we used root exudate-mimicking growth media [37,38], which were used as a "seed" (-s flag in m2m) for the targeted predicted producible metabolites (Supplementary Table S2).

The selection of the minimal community and computation of key species were performed using the m2m mincom command line. For this analysis, we considered each host GSMN in sbml format and included essential compounds for their metabolism, such as amino acids, nucleotide components, cofactors, vitamins, phytohormones, organic acids, and other compounds relevant to plant interactions (Supplementary Table S3). The identification of key species for the targeted set of compounds was carried out using MiSCoTo [39], which is implemented in the m2m suite.

The MetaCyc Metabolic Pathway Database [40] and the Kyoto Encyclopedia of Genes and Genomes were used as references to link genome annotation to metabolism.

Identification of plant growth-promoting traits (PGPTs)

To identify genes associated with plant growth-promoting traits (PGPTs) within the MAGs, we conducted an alignment of their protein sequences. This alignment was performed using a combination of BLASTP and HMMER tools available in the PGPT-Pred database from PLaBAse [41]. To ensure the accuracy of the annotations, we further validated them by conducting BLASTP searches against the NCBI non-redundant protein, RefSeq, and UniProtKB/Swiss-Prot databases. Hits with an E-value <1e−5 were considered significant for both approaches.

To assess the MAGs' potential to interact with plants, we established a criterion based on the presence or absence of 86 PGPT genes. These genes are involved in nitrogen fixation, phosphorus solubilization, as well as the production of EPSs, siderophores, and plant growth hormones. Specifically:

Nitrogen-fixing genes: *nifA, nifB, nifD, nifE, nifF, nifH, nifHD1, nifHD2, nifJ, nifK, nifM, nifN, nifQ, nifS, nifT, nifU, nifV, nifW, nifX, nifZ.*

Exopolysaccharide (EPS) production*: epsE, epsD, epsF, epsH, epsI, epsJ, epsL, epsM, epsN, epsO.*

Root colonization by nodulation: *nodA, nodB, nodC, node, nodF, nodI, nodJ, nodU, nod, nodT, nolM, noeA, noeB, noeC, noeD, noeE, nodN, nodN_like, nodO, nodP, nodS, nodS_like, nodY, nodZ, nodV, nodV_like, nodW, nodX, nodX, nodO.*

Oxidative stress|ROS scavenging: *sodN, sodC, sod3.*

Iron acquisition: *lipA, lipB, lipL, lipL2, lipM, lplA.*

Salinity stress-potassium transport: *kdpA, kdpB, kdpC, kdpD, kdpE, kdpF.*

Plant embryogenesis-spermidine: *puuA, puuB, puuC, puuD, puuE, pup*

IAA-related tryptophan metabolism*: trpA, trpB, trpC, trpCF, trpD, trpE, trpEG, trpG, trpDG, trpF, trpS, trpR*

Microbe-Microbe and Plant-Microbe interactions using Reverse Ecology

Protein-level assemblies of eight MAGs were annotated using KofamKOALA [42] (KOfam parameters: '--e-value 0.00001'). The KEGG orthologs (KO) obtained from these annotations were utilized for determining microbe-microbe interactions, specifically competition and complementarity indices. To calculate these indices, we employed the Cooperation Index package in RevEcoR [43].

Next, we created a matrix that included the substrate and product information for each species from the RevEcoR analysis. Subsequently, we applied NetCooperate to gain insights into potential ecological host-microbe interactions [44]. For pairwise interactions between hosts and microbes, we measured the biosynthetic support score (BSS) and the metabolic complementarity index (MCI). The KOs obtained from the annotated host genomes, which were sourced from the Joint Genome Institute (JGI, https://genome.jgi.doe.gov/portal/), were considered for these analyses. The compounds involved were annotated in KEGG compounds with their biological roles and the Phytochemical Compounds Database (https://www.genome.jp/kegg/ compound/) as well as the MetaCyc Metabolic Pathway Database.

Additionally, we utilized NetMet [45] to predict the metabolic performances of microbes and their combinations in user-defined environments. To achieve this, we utilized lists of species-specific enzymatic reactions (EC numbers) obtained from the JGI, along with the host compounds, as inputs for the metabolic environment analysis.

Results

Design of this study

In this study, we investigated 270 previously described MAGs obtained from Campos rupestres, a grassland ecosystem located on the geologically ancient rocky outcrops of central and eastern regions of Brazil [30]. This unique ecosystem is characterized by extremely low concentrations of essential nutrients, creating challenging conditions for plant growth [46]. However, it also

serves as a habitat for microorganisms that have adapted to these harsh conditions. The MAGs were primarily sourced from bulk soil (BS) and the rhizosphere (RX) of two dominant plant species in the campos rupestres, namely *V. epidendroide*s (BE) and *B. macrantha* (BM).

To gain insights into the metabolic capabilities of these microorganisms, we reconstructed GSMNs for the MAGs and determined the minimum number of species required to perform specific metabolic functions related to plant-microbe interactions. We considered a predefined set of target compounds and incorporated information about the associated plant hosts. By doing so, we successfully reduced the complexity of the microbiota and identified minimal communities with comparable properties (Fig. 1).

Furthermore, we conducted a comprehensive analysis of key species within these minimal communities, focusing on their PGPTs. Additionally, we calculated competition and complementarity indices for each pair of species to assess their potential interactions. This thorough workflow allowed us to identify beneficial microbes and construct SynCom that have the potential to enhance crop productivity (Fig. 1).

Soil and rhizosphere microbiome associated with dominant plant species in the Campos rupestres

Here, we utilized 270 MAGs obtained from the soil and rhizosphere microbiomes of *V. epidendroides* (BE) and *B. macrantha* (BM). Among the MAGs, those belonging to the phylum Proteobacteria were the most abundant in the samples (37.4%), followed by Actinobacteriota (15.9%) and Acidobacteriota (14.8%) (Fig 2, Fig. S1A). Interestingly, we observed a significant presence of the phylum Eremiobacterota, known for its ecological versatility and ability to thrive under various extreme environmental conditions [47], which accounted for 10.3% of the species in the dataset used in this study.

The size of the MAGs varied, ranging from 0.6 Mb in a MAG affiliated with the phylum Patescibacteria to 8 Mb in MAGs belonging to Chloroflexota, Proteobacteria, and Myxococcota (Fig 2, Fig. S1B). Generally, we did not observe a direct relationship between the distribution of phyla and the environment. Most MAGs were widely distributed throughout the Campos rupestres, which may be attributed to the heterogeneity of MAGs within and across phyla. We retrieved a total of 48 MAGs from the bulk soil of *V. epidendroides* (BE_BS), 55 MAGs from the rhizosphere of *V. epidendroides* (BE_RX), 82 MAGs from the bulk soil of *B. macrantha* (BM_BS) and 85 MAGs from the rhizosphere of *B. macrantha* (BM_RX). For detailed

information on each MAG and its taxonomic assessment, please refer to Supplementary Table S1.

Reconstruction of genome-scale metabolic networks (GSMNs) to understand metabolic complementarity between species

To identify metabolic functions and species of interest within the Campos rupestres microbiota, we employed the Pathway Tools [34] integrated within the m2m [35]. Notably, the overall statistics of GSMNs varied depending on the number of MAGs per sample, with MAGs from BE_BS displaying relatively lower numbers. In total, the reconstruction of GSMNs encompassed a range of 59,479 to 101,411 compounds, 39,700 to 72,566 reactions, and 945 to 1,742 pathways. Interestingly, MAGs from BE_RX exhibited a relatively smaller number of pathways (Fig. S2, Supplementary Table S4, Supplementary Table S5). Individual GSMNs for each MAGs are provided in Supplementary Data 1.

Next, we gained insights into the metabolic potential, referred to as scopes, of individual metabolic networks under specific nutritional conditions, specifically a seed representing a "root exudate-mimicking growth media" (see Materials and Methods for details). Overall, the sizes of individual scopes remained relatively stable across all GSMNs. The results revealed that, on average, each microbiota from the samples could access approximately 80 metabolites (as indicated by the intersection of scopes) (Fig. 3A, Supplementary Table S6). When considering the reachable metabolites for all organisms together, the combined scope comprised 521, 444, 451, and 510 metabolites in the BE_BS, BE_RX, BM_BS, and BM_RX samples, respectively (Fig. 3A, Supplementary Table S6, Supplementary Table S7).

Additionally, we conducted a metabolic potential analysis considering the metabolites reachable by the entire community, excluding the seeds. Overall, we examined the composition of the 519 newly producible metabolites accessible to the community (Fig. 3B, Supplementary Table S8). Interestingly, the number of these metabolites remained consistent regardless of the inclusion of the GSMNs of the host plant in the analysis (Fig. S3) (see Materials and Methods for details). Most of these metabolites predominantly consisted of carbohydrates, organic acids, nucleotide components, and amino acids (Fig. 3C). Among the target compounds included in the analysis of the entire community, we discovered that this community could produce 17 amino acids, five organic acids, and five aromatic compounds (Fig. 3D, Supplementary Table S9).

Defining a minimal community of microbial

We proceeded to reduce the microbiota into minimal communities that possess equivalent properties, making them suitable for further analyses. Considering our desired target compounds and the host metabolism (crop), we narrowed down the community to 68 bacterial species. Specifically, we identified 16 species associated with soybean, 18 with maize, 17 with sorghum, and 39 with sugarcane (Supplementary Table S10). Among these species, 47 MAGs belonged to Proteobacteria, 13 MAGs belonged to Actinobacteriota, 7 MAGs belonged to Eremiobacterota, and 4 MAGs belonged to Acidobacteriota. These phyla were generally present across all samples, but Eremiobacterota showed a stronger association with *V. epidendroide*s (BS/BS), while Actinobacteriota and Acidobacteriota exhibited a preference for *B. macrantha* (Fig 4A). We also observed a core set of species shared among different plant hosts, with sugarcane hosting a unique set of species (Fig 4B). At a lower taxonomic level, families such as *Beijerinckiaceae*, *Binataceae*, *Chroococcidiopsidaceae*, *Enterobacteriaceae*, *Reyanellaceae*, and *Steroidobacteraceae* were found to be widespread across all hosts. Sorghum had the highest number of unique families selected (Fig 4C). Additionally, the producible set of metabolites from this minimal community included essential amino acids, a selection of ten organic acids, aromatic compounds like indole, vitamins, and inorganic ions (Fig 4D, Supplementary Table S11).

We analyzed the MAGs to identify genes encoding proteins related to plant growth-promoting traits (PGPTs). These PGPT genes were categorized into seven classes, including Iron acquisition, EPS production, potassium solubilization, nitrogen fixation, GABA production, and IAA (indole-3-acetic acid) related tryptophan metabolism (Fig. 5). It is noteworthy that a majority of these MAGs contained genes encoding PGPT proteins, although the presence and abundance varied across different host plants. We found that genes encoding for potassium solubilization, and IAA metabolism were more abundant across the MAGs. Specifically, only a few MAGs from the species *Metakosakonia intestini* exhibited the potential for nitrogen fixation, and we identified one MAG associated with soybean that possessed a complete set of nitrogen-fixing genes (Fig. 5). Detail of PGPTs for each MAGs can be found in Supplementary Table S12 and Supplementary data 2.

Designing a SymCom based on key species in a minimal community through pairwise interactions

Within the 68 species of the minimal community, we identified eight species as essential symbionts, present in all minimal communities and enabling the producibility of the target

metabolites [35]. These essential symbionts represented four phyla (Cyanobacteria, Eremiobacterota, Proteobacteria, and Verrucomicrobiota), with only one MAG assigned at the species level, *M. intestine*. Interestingly, these essential symbionts were associated with all crop hosts (Supplementary Table S13).

We aimed to construct a microbial consortium by identifying microbe-microbe interactions using RevEcoR [43]. To assess the dynamics within the consortium, we calculate competition and complementarity indices for all pairs of species (Fig. S4). Among the species in the SymCom, we observed that members of Enterobacteriaceae, specifically *M. intestine* and an *Enterobacter* specie, exhibited the highest competitiveness index compared to other species. Consequently, we decided to exclude these two species and recalculated the competition index. Interestingly, the results revealed that none of the species exceeded a competition index greater than 0.6, indicating a lack of competition among the species (Fig. 6A).

Regarding the complementarity index, we observed that a member of the phylum Eremiobacterota tended to provide support to most members in the SymCom (Fig. 6B). Furthermore, we assessed the metabolic complementarity between species and the host plants (Fig. 6C, 6D). Initially, we calculated the Biosynthetic Support Score (BSS), which evaluates the host species' ability to meet the nutritional requirements of a parasitic or commensal species. Additionally, we computed the Metabolic Complementarity Index (MCI), which quantifies the extent to which two microbial species can support each other through biosynthetic complementarity [17].

The distribution of BSS values across the hosts ranged from approximately 0.1 to 0.8, while the MCI values ranged from 0.0 to 0.4 (Fig. 6A, Fig. 6B). As expected, there were clear differences in BSS and MCI scores between bacteria and hosts, indicating that hosts tend to better fulfill the nutritional needs of the soybean. Interestingly, a member of the Chthoniobacterales order, belonging to the Verrucomicrobiota phylum, exhibited high BSS and MCI scores in pairwise interactions with sorghum and soybean (Fig. 6C, 6D). We identified 641 BSS compounds that were supported by the SymCom and 1,037 compounds in which the SymCom was supported by the hosts (Supplementary Table S14). Most of these compounds were not assigned to known compounds, but we found that carbohydrates, esters, amino acids, and aromatic compounds were the most frequently involved in the interaction between the SymCom and the hosts (Fig. 6F). Conversely, amino acids, lipids, and coenzymes were found to be involved in the interaction between hosts and the SymCom (Fig. S5).

We have confirmed that the six species included in the SymCom encoding PGPT proteins are actively involved in various essential processes such as nitrogen fixation, phosphorus solubilization, EPS production, siderophore production, and plant growth hormone production (Supplementary Table S15) Our observations indicate that the *Beijerinckiaceae* (bei) species possess a comprehensive set of genes responsible for phosphate transport, homeostasis, and degradation (*pho*, *pts*, and *phn* clusters), as well as the production of siderophores like enterobacterin and mycobactin (*ent* and *mdt* clusters). Furthermore, all the species exhibit the potential to promote plant germination through the production of H_2S and the synthesis of IAA. They also possess the ability to solubilize potassium, with three of them capable of producing the GABA phytohormone and protecting against osmotic stress through the production of osmolytes such as glycine and betaine (Supplementary Table S15). Collectively, our findings provide strong evidence that the SymCom model can generate crucial PGPTs that significantly enhance crop productivity. In return, these enhanced plants can contribute to the maintenance and sustainability of SymCom (Fig. 6H).

Discussion

The integration of omics data acquisition and analysis, along with modeling approaches, enables computational predictions of an organism's resource utilization, biosynthetic capacities, limitations, and growth under diverse conditions [48,49]. These models rely on the reconstruction of metabolic networks from annotated genomes, which integrate all the expected metabolic reactions of an organism. This makes it possible to predict fundamental information regarding the competition of members of the microbial community, and their cooperation among microbes and their host [50,51].

In this study, we used 270 MAGs derived from the microbiomes associated with *V. epidendroides* and *B. macrantha* in campos rupestres, as described by [30]. This unique ecosystem is characterized by remarkably low concentrations of essential nutrients, creating challenging conditions for plant growth. However, these conditions also present an excellent opportunity to explore plant-microbe interactions in harsh environments. The authors noted a significant level of taxonomic novelty within this environment and highlighted the presence of microbial taxa associated with low phosphorus (P) soils, which have the potential for phosphorus turnover [46]. Our *in-silico* approach involved reducing the microbial community based on the metabolic complementarity between bacterial species and host crop plants, aiming to identify hub species that preserve the essential PGPTs for the design of a SymCom.

We showed that the selection of this minimal community was based on our target compound, encompassing essential compounds for their metabolism, such as amino acids, nucleotide components, cofactors, vitamins, phytohormones, organic acids, and other compounds relevant to plant interactions. Through this process, we successfully reduced the initial community size from 270 to 68 species, resulting in a notable change of approximately 4.5-fold. This reduction is particularly intriguing, considering the intricate microbial interactions typically observed within soil communities [52–54]. Remarkably, the microbial community exhibited significant production of amino acids, organic acids, vitamins, aromatic compounds, and various inorganic ions. These findings align with previous studies on plant microbiomes, which also preferentially utilize these nutrients [38]. These compounds have been extensively documented to play a crucial role in plant-microbe interactions, particularly through root secretions where plants can signal and attract beneficial microbes under specific conditions [55–59]. In contrast, gut microbes were found to predominantly metabolize lipids, sugar derivatives, and carboxy acids [35].

The minimal community we identified preserved important genes associated with PGPTs, including those involved in iron acquisition, EPS production, potassium solubilization, nitrogen fixation, GABA production, and IAA-related tryptophan metabolism. This emphasizes the significance of screening novel microbial taxa, particularly under harsh conditions. The findings suggest that such microbial taxa may assist plants in thriving during drought conditions. Recent studies in the Atacama Desert have highlighted the role of plant growth-promoting bacteria (PGPB) and positive gene selection in facilitating key mechanisms for plant survival [60]. In line with this study, an emphasized the presence of Acidobacteriota, Eremiobacterota, and Verrucomicrobiota as members of this group of PGPB. It is worth noting that members of these phyla are often characterized as slow-growing bacteria and most lack culture representative. A genome-centric approach has recently provided a comprehensive analysis of 756 MAGs belonging to Acidobacteriota, revealing their potential to promote plant growth through their interactions (Gonçalves et al., 2023).

Within the minimal community, we carefully selected eight hub-species or essential symbiotics that were present in all minimal communities and enabled the production of target metabolites. These hub species were chosen to compose the SymCom and represented four phyla, namely Cyanobacteria, Eremiobacterota, Proteobacteria, and Verrucomicrobiota. In our initial round of species interactions, we excluded two proteobacteria that displayed high competitiveness within the SymCom. It has been documented that genome-encoded metabolic potential tends to cluster

quantitatively and qualitatively based on phylogeny, resulting in competitive behaviors among species [38,61]. The exclusion of these species revealed that all the species within the SymCom exhibited cooperative interactions, with a member of the Eremiobacterota phylum playing a supportive role in the metabolism of most species. This finding further underscores the significance of this phylum in microbial interactions. To comprehend the interaction between SymCom and the host, we integrated five important crop plants, including soybean, maize, sorghum, and sugarcane, into our study. Interestingly, our results demonstrated that, in general, only sorghum selected unique microbial species. However, a core set of species remained consistent across different hosts, suggesting their importance independent of the specific plant type. This result was crucial in identifying six hub species that could enhance the growth of all crops. Furthermore, we found that the hosts primarily provided amino acids, lipids, and coenzymes, while the SymCom, in addition to PGPTs, supplied carbohydrates, esters, amino acids, and aromatic compounds to the hosts.

Conclusion

The soil microbiome associated with plants in stressful environments offers an exceptional opportunity to investigate how plants select beneficial microbial taxa to enhance their survival. It is well-known that soil microbe interactions are complex and depend on various biotic and abiotic factors. Recent advancements in sequencing technologies have allowed for insights into these microbes and their interactions within their natural environment. Computational modeling and prediction approaches enable the exploration of such dynamics.

In this study, we employed an *in-silico* approach using genome metabolic modeling to design a synthetic microbial community aimed at improving the yield of important crop plants. This approach relied on comprehensive knowledge of microbe-microbe-plant interactions and involved the selection of key species carrying essential plant growth-promoting traits. Similar approaches, as demonstrated here, can be combined with culturomics- and metagenomics-based techniques [62] to design synthetic microbial communities as microbial inoculants for future agricultural production.

Availability of data and materials

The genomes used in this study are publicly available in the supplementary information. Source data with genes associated with plant growth-promoting traits (PGPTs) in the MAGs and GSMNs file are available in the Zenodo repository: xx, xx. The datasets generated and analyzed during the study are provided with this paper in supplementary information.

References

1. Berg G, Grube M, Schloter M, Smalla K. Unraveling the plant microbiome: looking back and future. Frontiers in Microbiology. 2014. p. 148. https://www.frontiersin.org/article/10.3389/ fmicb.2014.00148

2. Dastogeer KMG, Tumpa FH, Sultana A, Akter MA, Chakraborty A. Plant microbiome–an account of the factors that shape community composition and diversity. Curr Plant Biol. 2020; 23:100161. https://www.sciencedirect.com/science/article/pii/S2214662820300426

3. Naylor D, Coleman-Derr D. Drought Stress and Root-Associated Bacterial Communities. Frontiers in Plant Science. 2018. p. 2223. https://www.frontiersin.org/article/10.3389/ fpls.2017.02223

4. Flemer B, Gulati S, Bergna A, Rändler M, Cernava T, Witzel K, et al. Biotic and Abiotic Stress Factors Induce Microbiome Shifts and Enrichment of Distinct Beneficial Bacteria in Tomato Roots. Phytobiomes J. 2022; 6:276–89. https://doi.org/10.1094/PBIOMES-10-21-0067-R

5. Souza R, Ambrosini A, Passaglia LMP, de Souza R, Ambrosini A, Passaglia LMP. Plant growth-promoting bacteria as inoculants in agricultural soils. Genet Mol Biol. 2015; 38:401–19. http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-47572015000400401 &lng=en& tlng=en

6. Compant S, Clément C, Sessitsch A. Plant growth-promoting bacteria in the rhizo- and endosphere of plants: Their role, colonization, mechanisms involved and prospects for utilization. Soil Biol Biochem. 2010; 42:669–78. https://www.sciencedirect.com/science/article/pii/ S0038071709004398

7. terHorst CP, Lennon JT, Lau JA. The relative importance of rapid evolution for plant-microbe interactions depends on ecological context. Proceedings of the Royal Society B: Biological Sciences. 2014; 281:20140028. https://doi.org/10.1098/rspb.2014.0028

8. Mahmud K, Missaoui A, Lee K, Ghimire B, Presley HW, Makaju S. Rhizosphere microbiome manipulation for sustainable crop production. Curr Plant Biol. 2021; 27:100210. https://www.sciencedirect.com/science/article/pii/S2214662821000153

9. Ahkami AH, Allen White R, Handakumbura PP, Jansson C. Rhizosphere engineering: Enhancing sustainable plant ecosystem productivity. Rhizosphere. 2017; 3:233–43. https://www.sciencedirect.com/science/article/pii/S2452219817300459

10. Hakim S, Naqqash T, Nawaz MS, Laraib I, Siddique MJ, Zia R, et al. Rhizosphere Engineering With Plant Growth-Promoting Microorganisms for Agriculture and Ecological Sustainability. Front Sustain Food Syst. 2021;5. https://www.frontiersin.org/articles/10.3389/ fsufs.2021.617157

11. Großkopf T, Soyer OS. Synthetic microbial communities. Curr Opin Microbiol. 2014; 18:72–7. https://www.sciencedirect.com/science/article/pii/S1369527414000198

12. Prigigallo MI, Gómez-Lama Cabanás C, Mercado-Blanco J, Bubici G. Designing a synthetic microbial community devoted to biological control: The case study of *Fusarium* wilt of banana. Front Microbiol. 2022;13. https://www.frontiersin.org/articles/10.3389/fmicb.2022.967885

13. De Vrieze M, Germanier F, Vuille N, Weisskopf L. Combining Different Potato-Associated *Pseudomonas* Strains for Improved Biocontrol of Phytophthora infestans. Front Microbiol. 2018;9. https://www.frontiersin.org/articles/10.3389/fmicb.2018.02573

14. Pradhan S, Tyagi R, Sharma S. Combating biotic stresses in plants by synthetic microbial communities: Principles, applications and challenges. J Appl Microbiol. 2022; 133:2742–59. https://doi.org/10.1111/jam.15799

15. Xu X, Zarecki R, Medina S, Ofaim S, Liu X, Chen C, et al. Modeling microbial communities from atrazine contaminated soils promotes the development of biostimulation solutions. ISME J. 2019; 13:494–508. https://doi.org/10.1038/s41396-018-0288-5

16. de Souza RSC, Armanhi JSL, Arruda P. From Microbiome to Traits: Designing Synthetic Microbial Communities for Improved Crop Resiliency. Front Plant Sci. 2020;11. https://www.frontiersin.org/articles/10.3389/fpls.2020.01179

17. Johns NI, Blazejewski T, Gomes ALC, Wang HH. Principles for designing synthetic microbial communities. Curr Opin Microbiol. 2016; 31:146–53. https://www.sciencedirect.com /science/article/pii/S136952741630025X

18. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, et al. A new view of the tree of life. Nat Microbiol. 2016; 1:16048. https://doi.org/10.1038/nmicrobiol.2016.48

19. Zhu Q, Mai U, Pfeiffer W, Janssen S, Asnicar F, Sanders JG, et al. Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and Archaea. Nat Commun. 2019; 10:5477. https://doi.org/10.1038/s41467-019-13443-4

20. Nayfach S, Roux S, Seshadri R, Udwary D, Varghese N, Schulz F, et al. A genomic catalog of Earth's microbiomes. Nat Biotechnol. 2021; 39:499–509. https://doi.org/10.1038/s41587-020-0718-6

21. Kindler GS, Wong HL, Larkum AWD, Johnson M, MacLeod FI, Burns BP. Genome-resolved metagenomics provides insights into the functional complexity of microbial mats in Blue Holes, Shark Bay. FEMS Microbiol Ecol. 2022;98: fiab158. https://doi.org/10.1093/femsec/fiab158

22. Wang Y, Zhang Y, Hu Y, Liu L, Liu S-J, Zhang T. Genome-centric metagenomics reveals the host-driven dynamics and ecological role of CPR bacteria in an activated sludge system. Microbiome. 2023; 11:56. https://doi.org/10.1186/s40168-023-01494-1

23. Gong X, del Río ÁR, Xu L, Chen Z, Langwig M V, Su L, et al. New globally distributed bacterial phyla within the FCB superphylum. Nat Commun. 2022; 13:7516. https://doi.org/ 10.1038/s41467-022-34388-1

24. Langwig M V, De Anda V, Dombrowski N, Seitz KW, Rambo IM, Greening C, et al. Large-scale protein level comparison of Deltaproteobacteria reveals cohesive metabolic groups. ISME J. 2022; 16:307–20. https://doi.org/10.1038/s41396-021-01057-y

25. de Souza LC, Procópio L. The profile of the soil microbiota in the Cerrado is influenced by land use. Appl Microbiol Biotechnol. 2021; 105:4791–803. https://doi.org/10.1007/s00253-021-11377-w

26. Haridasan M. Nutritional adaptations of native plants of the cerrado biome in acid soils. Brazilian Journal of Plant Physiology. 2008;20.

27. Procópio L, Barreto C. The soil microbiomes of the Brazilian Cerrado. J Soils Sediments. 2021; 21:2327–42. https://doi.org/10.1007/s11368-021-02936-9

28. Fadiji AE, Santoyo G, Yadav AN, Babalola OO. Efforts towards overcoming drought stress in crops: Revisiting the mechanisms employed by plant growth-promoting bacteria. Front Microbiol. 2022;13. https://www.frontiersin.org/articles/10.3389/fmicb.2022.962427

29. Jansson JK, Hofmockel KS. Soil microbiomes and climate change. Nat Rev Microbiol. 2020; 18:35–46. https://doi.org/10.1038/s41579-019-0265-7

30. Camargo AP, de Souza RSC, de Britto Costa P, Gerhardt IR, Dante RA, Teodoro GS, et al. Microbiomes of Velloziaceae from phosphorus-impoverished soils of the campos rupestres, a biodiversity hotspot. Sci Data. 2019; 6:140. https://doi.org/10.1038/s41597-019-0141-3

31. Parks D, Imelfort M, Skennerton C, Philip H, Tyson G. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25.

32. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020; 37:1530–4. https://doi.org/10.1093/molbev/msaa015

33. Letunic I, Bork P. Interactive Tree of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019; https://doi.org/10.1093/nar/gkz239

34. Karp PD, Midford PE, Billington R, Kothari A, Krummenacker M, Latendresse M, et al. Pathway Tools version 23.0 update: software for pathway/genome informatics and systems biology. Brief Bioinform. 2021; 22:109–26. https://doi.org/10.1093/bib/bbz104

35. Belcour A, Frioux C, Aite M, Bretaudeau A, Hildebrand F, Siegel A. Metage2Metabo, microbiota-scale metabolic complementarity for the identification of key species. Elife. 2020;9:e61968. https://doi.org/ 10.7554/eLife.61968

36. de Oliveira Dal'Molin CG, Quek L-E, Palfreyman RW, Brumbley SM, Nielsen LK. C4GEM, a Genome-Scale Metabolic Model to Study C4 Plant Metabolism. Plant Physiol. 2010; 154:1871–85. https://doi.org/10.1104/pp.110.166488

37. Baudoin E, Benizri E, Guckert A. Impact of artificial root exudates on the bacterial community structure in bulk soil and maize rhizosphere. Soil Biol Biochem. 2003; 35:1183–92. https://www.sciencedirect.com/science/article/pii/S0038071703001792

38. Mataigne V, Vannier N, Vandenkoornhuyse P, Hacquard S. Multi-genome metabolic modeling predicts functional inter-dependencies in the *Arabidopsis* root microbiome. Microbiome. 2022; 10:217. https://doi.org/10.1186/s40168-022-01383-z

39. Frioux C, Fremy E, Trottier C, Siegel A. Scalable and exhaustive screening of metabolic functions carried out by microbial consortia. Bioinformatics. 2018;34: i934–43. https://doi.org/ 10.1093/bioinformatics/bty588

40. Caspi R, Billington R, Keseler IM, Kothari A, Krummenacker M, Midford PE, et al. The MetaCyc database of metabolic pathways and enzymes - a 2019 update. Nucleic Acids Res. 2020;48: D445–53. https://doi.org/10.1093/nar/gkz862

41. Patz S, Gautam A, Becker M, Ruppel S, Rodríguez-Palenzuela P, Huson DH. PLaBAse: A comprehensive web resource for analyzing the plant growth-promoting potential of plant-associated bacteria. bioRxiv. 2021;2021.12.13.472471. http://biorxiv.org/content/early/ 2021/12/15/2021.12.13.472471.abstract

42. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics. 2020; 36:2251–2. https://doi.org/10.1093/bioinformatics/btz859

43. Cao Y, Wang Y, Zheng X, Li F, Bo X. RevEcoR: an R package for the reverse ecology analysis of microbiomes. BMC Bioinformatics. 2016; 17:294. https://doi.org/10.1186/s12859-016-1088-4

532 44. Levy R, Carr R, Kreimer A, Freilich S, Borenstein E. NetCooperate: a network-based tool for
533 inferring host-microbe and microbe-microbe cooperation. BMC Bioinformatics. 2015; 16:164.
534 https://www.ncbi.nlm.nih.gov/pubmed/25980407

535 45. Tal O, Selvaraj G, Medina S, Ofaim S, Freilich S. NetMet: A Network-Based Tool for Predicting
536 Metabolic Capacities of Microbial Species and their Interactions. Microorganisms. 2020.

537 46. Camargo AP, de Souza RSC, Jose J, Gerhardt IR, Dante RA, Mukherjee S, et al. Plant microbiomes
538 harbor potential to promote nutrient turnover in impoverished substrates of a Brazilian biodiversity
539 hotspot. ISME J. 2023; 17:354–70. https://doi.org/10.1038/s41396-022-01345-1

540 47. Ji M, Williams TJ, Montgomery K, Wong HL, Zaugg J, Berengut JF, et al. *Candidatus*
541 *Eremiobacterota*, a metabolically and phylogenetically diverse terrestrial phylum with acid-tolerant
542 adaptations. ISME J. 2021; 15:2692–707. https://doi.org/10.1038/s41396-021-00944-8

543 48. Muller EEL, Faust K, Widder S, Herold M, Martínez Arbas S, Wilmes P. Using metabolic networks
544 to resolve ecological properties of microbiomes. Curr Opin Syst Biol. 2018; 8:73–80.
545 https://www.sciencedirect.com/science/article/pii/S245231001730197X

546 49. Kumar M, Ji B, Zengler K, Nielsen J. Modelling approaches for studying the microbiome. Nat
547 Microbiol. 2019; 4:1253–67. https://doi.org/10.1038/s41564-019-0491-9

548 50. Du H, Pan J, Zou D, Huang Y, Liu Y, Li M. Microbial active functional modules derived from
549 network analysis and metabolic interactions decipher the complex microbiome assembly in mangrove
550 sediments. Microbiome. 2022; 10:224. https://doi.org/10.1186/s40168-022-01421-w

551 51. Kuppa Baskaran DK, Umale S, Zhou Z, Raman K, Anantharaman K. Metagenome-based metabolic
552 modelling predicts unique microbial interactions in deep-sea hydrothermal plume microbiomes. ISME
553 Communications. 2023; 3:42. https://doi.org/10.1038/s43705-023-00242-8

554 52. Young IM, Crawford JW. Interactions and Self-Organization in the Soil-Microbe Complex. Science
555 (1979). 2004; 304:1634–7. https://doi.org/10.1126/science.1097394

556 53. Philippot L, Raaijmakers JM, Lemanceau P, van der Putten WH. Going back to the roots: the
557 microbial ecology of the rhizosphere. Nat Rev Microbiol. 2013; 11:789–99. https://doi.org/
558 10.1038/nrmicro3109

559 54. Romdhane S, Spor A, Aubert J, Bru D, Breuil M-C, Hallin S, et al. Unraveling negative biotic
560 interactions determining soil microbial community assembly and functioning. ISME J. 2022; 16:296–
561 306. https://doi.org/10.1038/s41396-021-01076-9

562 55. Moe LA. Amino acids in the rhizosphere: From plants to microbes. Am J Bot. 2013; 100:1692–705.
563 https://doi.org/10.3732/ajb.1300033

564 56. Korenblum E, Massalha H, Aharoni A. Plant–microbe interactions in the rhizosphere via a circular
565 metabolic economy. Plant Cell. 2022; 34:3168–82. https://doi.org/10.1093/ plcell/koac163

566 57. Macias-Benitez S, Garcia-Martinez AM, Caballero Jimenez P, Gonzalez JM, Tejada Moral M,
567 Parrado Rubio J. Rhizospheric Organic Acids as Biostimulants: Monitoring Feedbacks on Soil
568 Microorganisms and Biochemical Properties. Front Plant Sci. 2020;11. https:/
569 www.frontiersin.org/articles/10.3389/fpls.2020.00633

570 58. Ali S, Tyagi A, Park S, Mir RA, Mushtaq M, Bhat B, et al. Deciphering the plant microbiome to
571 improve drought tolerance: Mechanisms and perspectives. Environ Exp Bot. 2022; 201:104933.
572 https://www.sciencedirect.com/science/article/pii/S0098847222001551

59. Moormann J, Heinemann B, Hildebrandt TM. News about amino acid metabolism in plant- microbe interactions. Trends Biochem Sci. 2022; 47:839–50. https://doi.org/10.1016/j.tibs.2022. 07.001

60. Eshel G, Araus V, Undurraga S, Soto DC, Moraga C, Montecinos A, et al. Plant ecological genomics at the limits of life in the Atacama Desert. Proceedings of the National Academy of Sciences. 2021;118: e2101177118. http://www.pnas.org/content/118/46/e2101177118.abstract

61. Borenstein E, Kupiec M, Feldman MW, Ruppin E. Large-scale reconstruction and phylogenetic analysis of metabolic environments. Proceedings of the National Academy of Sciences. 2008; 105:14482 LP – 14487. http://www.pnas.org/content/105/38/14482.abstract

62. Lian W-H, Mohamad OAA, Dong L, Zhang L-Y, Wang D, Liu L, et al. Culturomics- and metagenomics-based insights into the microbial community and function of rhizosphere soils in Sinai desert farming systems. Environ Microbiome. 2023; 18:4. https://doi.org/10.1186/s40793-023-00463-3

Funding

This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq (Process APQ-02381-21), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior/Programa de Excelência Acadêmica-Finance Code 001 (CAPES ProEx grant 23038.019105/2016-86), CAPES-PrInt (process 88887.696147/2022-00) and Fundação de Amparo à Pesquisa do Estado de Minas Gerais—FAPEMIG (Process 402644/2021-2) for the financial support.

Acknowledgements

The authors expressed their gratitude to the technical support team of the Cluster at Universidade Federal de Viçosa.

Contributions

Conceptualization: OG, and MF. Data curation: OG. Funding acquisition: MF. Investigation: OG, CC. Methodology: OG, CC. Supervision: CC, MF. Visualization: OG. Writing – original draft: OG. Writing – review & editing: OG, MF, CC. All the authors read and approved the final manuscript.

Ethics approval and consent to participate.

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Figures

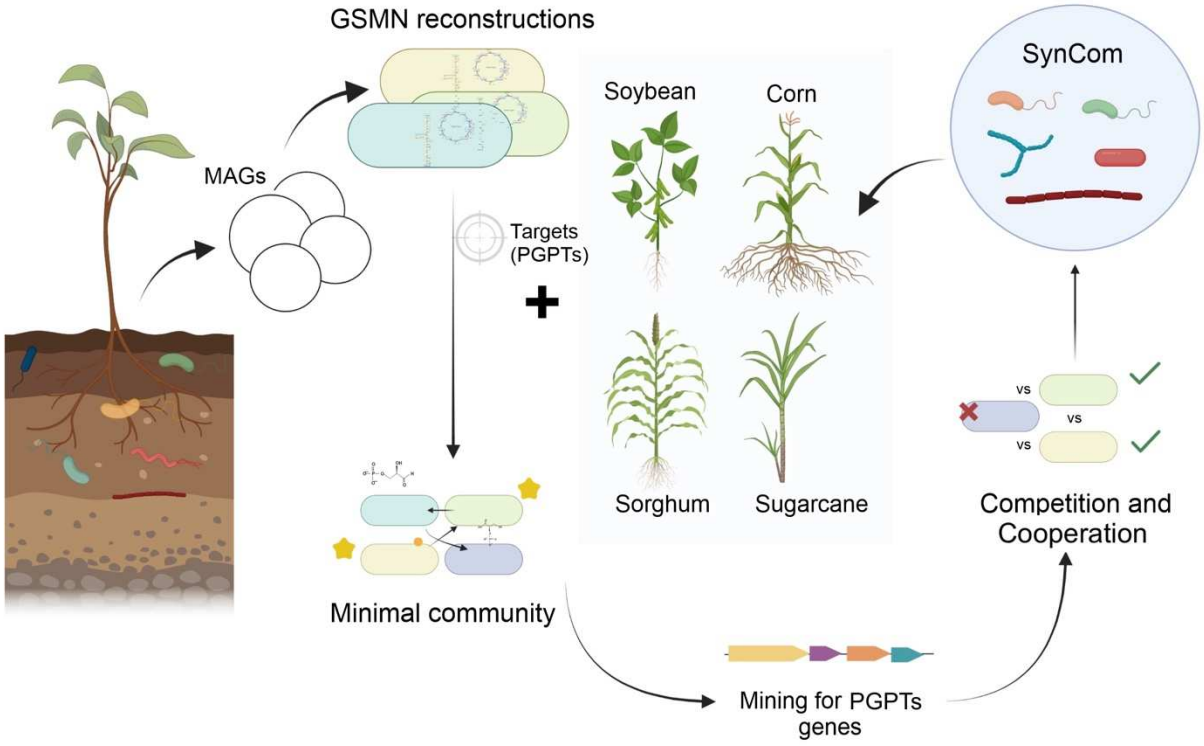


Figure 1. illustrates the workflow employed in this study to design the SymCom. The study starts with the use of 270 MAGs obtained from the microbiomes associated with *V. epidendroides* and *B. macrantha* in Campos rupestres, as detailed in Camargo et al. (2020). Subsequently, genome-scale metabolic networks of individual MAGs were reconstructed using the M2M tool suite. To determine the target compound, which encompasses crucial compounds for their metabolism (such as amino acids, nucleotide components, cofactors, vitamins, phytohormones, organic acids, and other compounds relevant to plant interactions), as well as by incorporating GSMNs of significant crop plants, we curated a minimal community (MinCom) from the original microbiome community linked to *V. epidendroides* and *B. macrantha*. From this MinCom, we identified genes associated with plant growth-promoting traits (PGPTs) and employed a reverse ecology approach to select species that collectively constitute our SymCom. The figure was designed using BioRender.

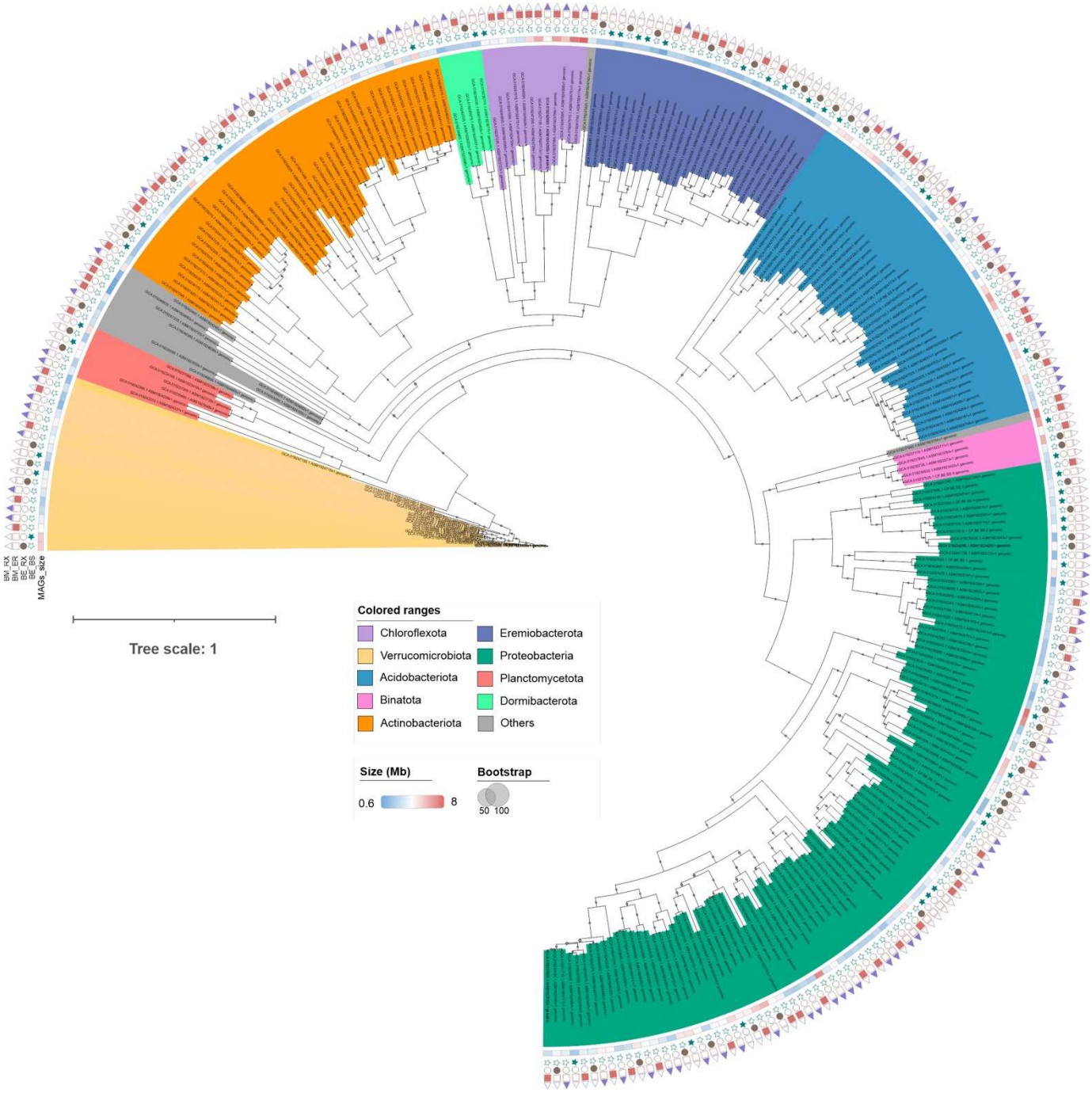


Figure 2. Phylogeny of 270 metagenome-assembled genomes (MAGs) from Campos rupestres. A maximum likelihood phylogenetic tree based on 43 single-copy; protein-coding marker genes identified using CheckM. Phyla are marked in different background colors. The first arc of the tree represents the genome size of each MAG, followed by sequential arcs indicating the source of the obtained MAGs, distinguished by filled geometric forms. A phylogenetic tree was built under the model of rate heterogeneity G+F+I+G4, and a maximum of 1,000 bootstrap replicates. Bootstraps are shown in black circles. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree.

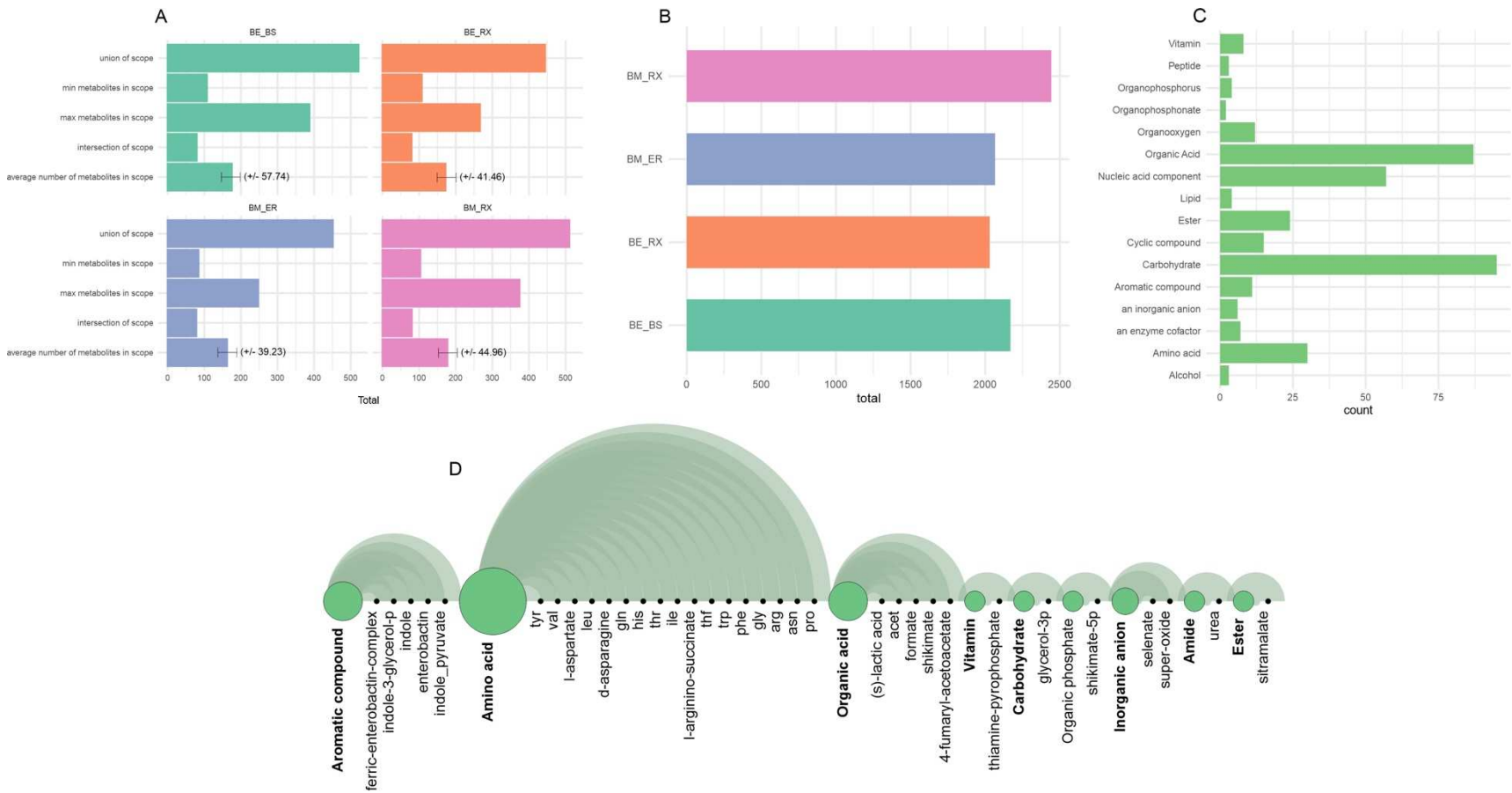


Figure 3. Metabolic producibility and cooperation potential of the microbiome from Campos rupestres. A) Metrics displaying individual metabolic potentials (scopes), showcasing the minimum, maximum, and arithmetic mean number of compounds across all scopes for each sample collection. Sample IDs are indicated above each graph bar. B) The total number of producible compounds accessible to the community in each sample. C) Classification of the accessible compounds into categories within the community. D) Detailed information regarding the target compounds accessible to the entire community. Circles represent the proportion of each class, while arcs depict the set of compounds within each class. Abbreviations: BE_BS (bulk soil of *V. epidendroides*), BE_RX (Rhizosphere of *V. epidendroides*), BM_BS (bulk soil of *B. macrantha*), and BM_RX (rhizosphere of *B. macrantha*)

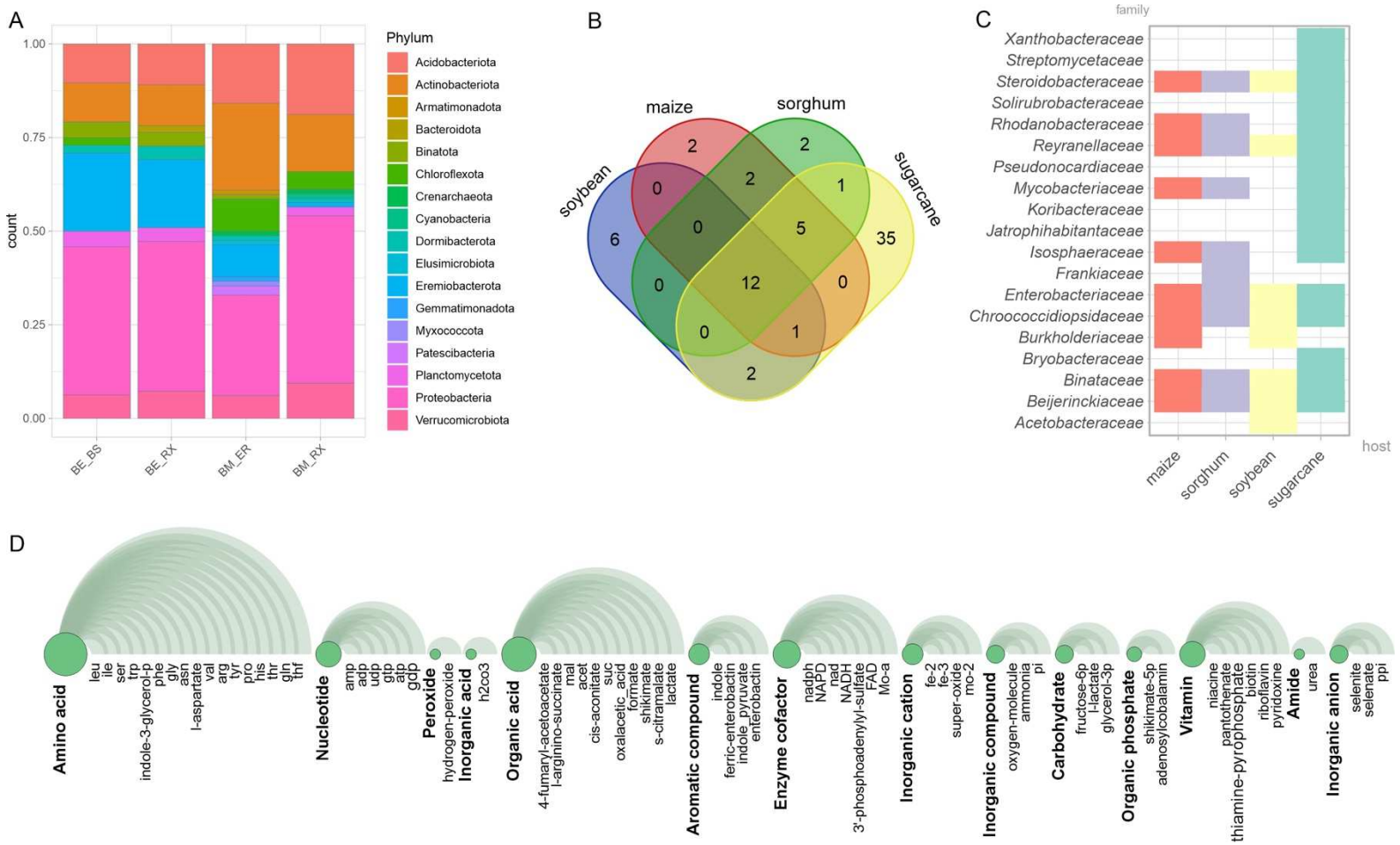


Figure 4.
Metabolic producibility of the minimal community of the microbiome from Campos rupestres. A) Taxonomic abundance of the minimal communities for each sample collection. Abbreviations: BE_BS (bulk soil of *V. epidendroides*), BE_RX (Rhizosphere of *V. epidendroides*), BM_BS (bulk soil of *B. macrantha*), and BM_RX (rhizosphere of *B. macrantha*). B) The core set of species shared among different plant hosts of the reduced community. C) Distribution of taxonomic family across different plant hosts. D) Community reduction analysis of the target categories in microbiome from Campos rupestres

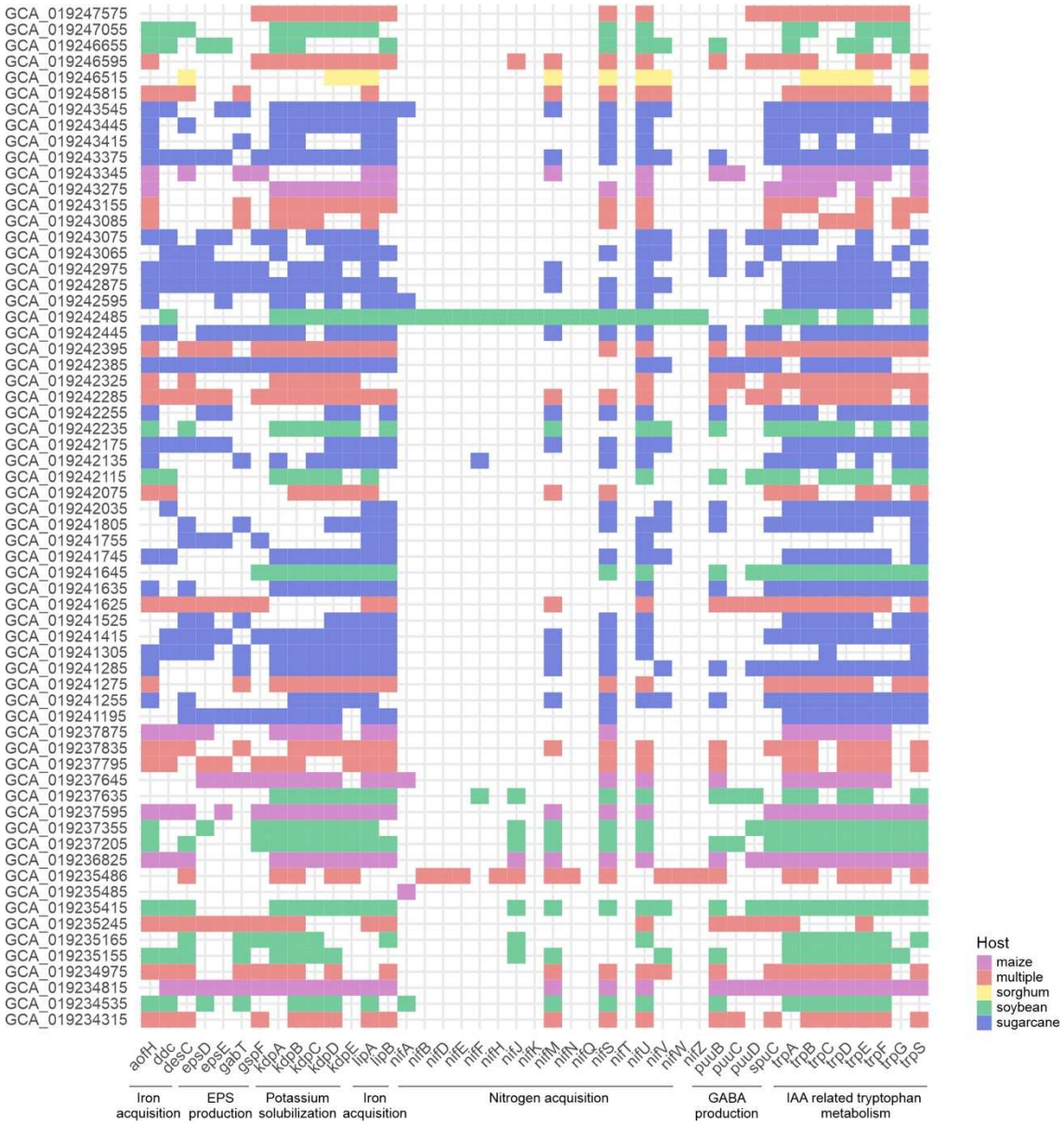


Figure 5. The distribution of key genes associated with plant growth-promoting traits among members of the minimal community in the microbiome from Campos rupestres. The Y-axes represent the GenBank access numbers of individual members within the reduced community, while the X-axes group the PGPT genes into categories based on their shared functions. The graph is color-coded according to the host plant, and the absence of color in squares indicates the absence of the gene. The presence of multiple, red-colored squares indicates genes found across more than two hosts. Details of all PGPT genes are in Supplementary Table S12 and Supplementary Data 2.

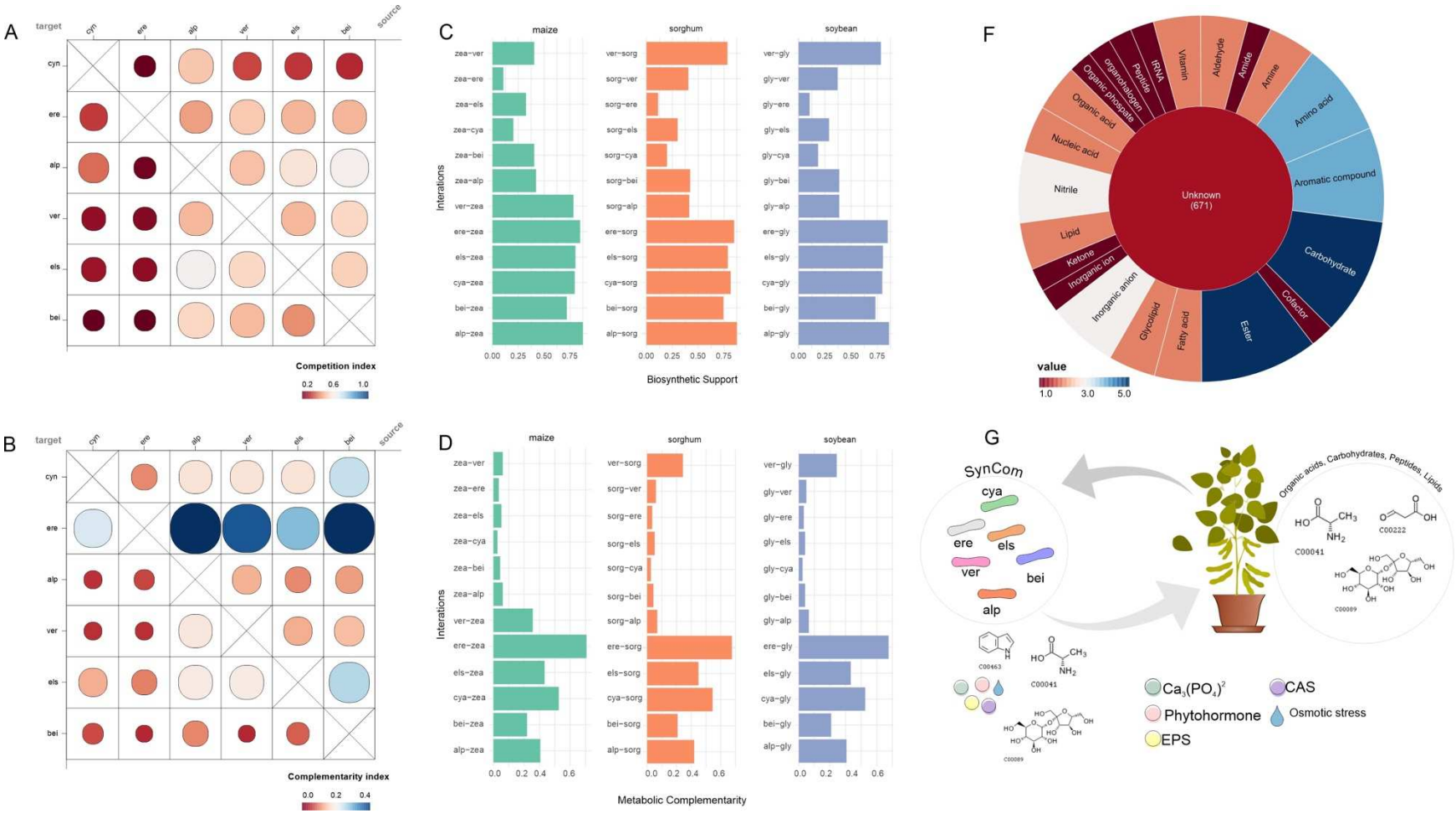


Figure 6. A SymCom designed based on comprehensive knowledge of microbe-microbe-plant interactions. A) Matrix plot of the competition index among six species of SymCom. The competition index ranges from 0 (indicating the lowest competition) to 1 (representing the highest competition index). B) Matrix plot of the complementary index among six species of SymCom. The size and color of each circle correspond to the competition and complementary index. C) The distribution of the biosynthetic support score for the hosts and microbial species. D) The distribution of the metabolic complementary score for the hosts and microbial species. The Y-axes represent the pairwise interactions and indicate that the first microbe/plant is metabolically supported or complemented by the second microbe/plant. F) Biosynthetic support with biological pathways found

in the compounds of the bacteria-host interaction. The color of each circle corresponds to its abundance, except for the inner circle, which shows the number of each compound for that category. G) The SymCom was designed with six microbial species showing potential PGPT features that can enhance plant growth. In turn, the host plant may supply potential compounds to the bacteria to support this interaction. Abbreviations: alp (Alphaproteobacteria), cya (Cyanobacteria), bei (*Beijerinckiaceae*), els (Elsterales), ere (Eremiobacterota), ver (Verrucomicrobiota).

Chapter 4

Insights into Plant Interactions and the Biogeochemical Role of the Globally Widespread Acidobacteriota Phylum

Gonçalves, O.S., Fernandes, A.S., Creevey C.J., Santana, M.F. Insights into Plant Interactions and the Biogeochemical Role of the Globally Widespread Acidobacteriota Phylum. The manuscript will be submitted to the ISME journal.

Insights into Plant Interactions and the Biogeochemical Role of the Globally Widespread Acidobacteriota Phylum

Osiel S. Gonçalves[1], Christopher J. Creevey[2], Mateus F. Santana[1*]

[1]Grupo de Genômica Eco-evolutiva Microbiana, Laboratório de Genética Molecular de Microrganismos, Departamento de Microbiologia, Instituto de Biotecnologia Aplicada à Agropecuária, Universidade Federal de Viçosa, Minas Gerais, Brazil.

[2]School of Biological Sciences, Institute for Global Food Security, Queen's University Belfast, Belfast, BT9 5DL, UK.

[*]Correspondence authors: mateus.santana@ufv.br

Abstract

The prevalence and abundance of Acidobacteriota raise concerns about their ecological function and metabolic activity in the environment. Studies have reported the potential of some members of Acidobacteriota to interact with plants and play a significant role in biogeochemical cycles. However, their role in this context has not been extensively studied. Here, we performed a comprehensive genomic analysis of 758 metagenome-assembled genomes (MAGs) and 121 RefSeq genomes of Acidobacteriota. Our analysis revealed a high frequency of plant growth-promoting traits (PGPTs) genes in the *Acidobacteriaceae*, *Bryobacteraceae*, *Koribacteraceae*, and *Pyrinomonadaceae* families. These PGPTs include genes involved in nitrogen fixation, phosphorus solubilization, exopolysaccharide production, siderophore production, and plant growth hormone production. Expression of such genes was found to be transcriptionally active in different environments. In addition, we expanded the knowledge of the functional potential for biological nitrogen fixation in Acidobacteriota. We identified numerous carbohydrate-active enzymes and peptidases involved in plant polymer degradation and revealed the distributional role of the phylum in carbon, nitrogen, sulfur, and trace metal cycling. This study underscores the distinct potential ecological roles for each of these taxonomic groups, providing valuable insights for future research.

Keywords: Ecological roles, metagenomics, plant growth-promoting bacteria, soil bacteria.

Main

Acidobacteriota is a highly prevalent bacterial phylum found in various environments such as soils, freshwater, and marine ecosystems worldwide. It was initially discovered in 1991 with the identification of *Acidobacterium capsulatum* from an acidic mineral habitat [1]. Since then, the phylum's widespread distribution was revealed by its distinctive 16S rRNA gene sequence [2–4]. Studies have shown that Acidobacteriota is a phylogenetically diverse group, consisting of 26 major sequence clades or subdivisions (SDs) [5]. It is also a geographically widespread and numerically significant component of the soil microbiota [6].

The prevalence and abundance of these microorganisms raise concerns about their ecological function and metabolic activity, particularly in soil environments [7–9]. In soil ecosystems, members of this phylum are suggesting contributing to nutrient cycling and organic matter decomposition [10–12]. They are known for their ability to tolerate a wide range of soil pH levels and have been found to be particularly abundant in acidic soils (ref). Some Acidobacteriota members have been classified as k-strategists, which means they can thrive in settings with limited nutritional availability, slower growth rates, and high tolerance to toxic compounds [13, 14].

Acidobacteriota strains from subdivision 1 (SD1) were identified to act as plant growth-promoting bacteria [14]. These strains produced plant growth-promoting traits (PGPTs), which increased the biomass of roots and shoots in *Arabidopsis thaliana* [14]. This was the first report of Acidobacteriota interacting with the plant. Moreover, given their dominance and metabolic activity in the soil, Acidobacteriota is believed to play a significant role in the biogeochemical cycles of rhizosphere soil [8]. The diversity of Acidobacteriota in various soils and environments indicates their potential significance in plant-microbe interactions. However, their role in this context has not been extensively studied.

Here, we performed a comprehensive genomic analysis of 758 metagenome-assembled genomes (MAGs) from catalog of Earth’s microbiomes (GEM) [15] and 121 RefSeq genomes of Acidobacteriota to better understand the plant interactions and the biogeochemical role of this phylum. Using this genomic database, we identified four ecologically distinct taxonomic groups that have the potential to play a direct role in plant growth. These groups possess genes encoding plant growth-promoting traits (PGPTs), such as nitrogen fixation, phosphorus solubilization, exopolysaccharide (EPS) production, siderophore and plant growth hormones

production. Furthermore, we uncovered distinct potential ecological roles for each of these taxonomic groups, which can provide valuable insights for future research.

Results

A comprehensive analysis of 758 MAGs reveals the potential for plant growth promotion of Acidobacteriota in diverse environments

To investigate the potential interaction between the Acidobacteriota phylum and plants, we collected a total of 758 metagenome-assembled genomes (MAGs) from the Global Earth Microbiome (GEM) catalog [15] and 121 RefSeq genomes from NCBI (Supplementary Table 1, Supplementary Table 2), which provided us with a comprehensive dataset to explore the diversity and abundance of Acidobacteriota within different environments. The dataset included MAGs from all continents and oceans on Earth, with a notable emphasis on samples from North America and Europe (Fig. 1A). These MAGs were obtained from diverse environments, including soils and other terrestrial habitats (243), ocean and other aquatic environments (173), plant host-associated (52), and engineered environments (52) (Fig. 1B, Supplementary Table 3). In addition, the sizes of Acidobacteriota MAGs ranged from 1 Mb to 11 Mb, with the largest one discovered in the UBA5066 and Bryobacterales order (Supplementary Table 4, Supplementary Fig 1).

These MAGs represented 13 different taxonomic classes, with Acidobacteriaceae (n= 372) being the most abundant class in the dataset (Supplementary Fig. 1). It is noteworthy that most of the isolates cultivated to date from this class are affiliated with formal subdivision 1, which suggests that this subdivision is the most well-studied group within the Acidobacteriota phylum [16]. Furthermore, we observed that the predominant members of the *Acidobacteriaceae* family were bacteria of the orders Acidobacteriales (Supplementary Table 1).

We constructed a phylogenetic tree using 43 single-copy marker genes to understand the evolutionary relationships between Acidobacteriota genomes (Fig. 1B). Our dataset included 758 MAGs obtained from various environments, as well as 121 Acidobacteriota reference genomes publicly available. The phylogenetic tree analysis revealed the formation of 17 distinct branches, thereby the presence of high yet uncharacterized Acidobacteriota species-level diversity (Fig. 1B). These branches represented different taxonomic groups within the dataset, with most genomes (n=211) were found to belong to the Acidobacteriales order, indicating their abundance and widespread distribution in various environments.

Subsequently, we conducted a genome mining analysis on Acidobacteriota MAGs to identify genes associated with PGPTs, to explore the potential of this phylum in plant growth promotion. We categorized these genes into six classes associated with PGPTs, namely bio-fertilization, stress-control/biocontrol, colonization, competitive exclusion, phytohormone production, and plant immune response stimulation (Supplementary Table 5). Our analysis revealed that the colonization and competitive exclusion PGPTs classes had a broad presence in the Acidobacteriota phylum (Fig. 1B), and their genes were linked to indirect effects on plant growth, such as the utilization of amino acids, carbohydrates, and lipids derived from plants, as well as motility and chemotaxis. In general, we noticed that the abundance of these categories varied significantly, particularly within certain taxa such as members of the *Acidobacteriaceae* (Supplementary Fig. 2). Notably, the Acidobacteriales and Holophagales order showed levels of PGPTs associated with bio-fertilization and phytohormone production. These classes are known to play a direct role in plant growth and contain genes involved in iron acquisition, nitrogen assimilation and regulation, mineral solubilization (K, P), auxin and cytokinin synthesis, vitamin production, plant germination, H_2S production, and metabolism (Supplemental data 1).

After observing that numerous genes within these categories may not be directly contribute to plant growth, we specifically chose 86 PGPTs genes that are involved in nitrogen fixation, phosphorus solubilization, as well as the production of EPSs, siderophore and plant growth hormones (Supplementary Table 6). A clear division into two clusters based on family and gene frequency was observed (Fig. 2). The first cluster comprised the families *Acidobacteriaceae*, *Bryobacteraceae*, *Koribacteraceae*, and *Pyrinomonadaceae*, which exhibited a high frequency of PGPT genes. Conversely, *Aminicenantaceae*, *Chloracidobacteriaceae*, *Holophagaceae*, *Thermoanaerobaculaceae*, and *Vicinamibacteraceae* displayed lower frequencies of these genes (Supplementary Fig. 3). The genes associated with tryptophan metabolism (*trpABCDEGFS*), potassium solubilization (*kdpABCDEF*), and EPS production (*epsDFHILMNO*) were predominantly identified in the second cluster. Furthermore, we observed a high occurrence of nitrogen-fixing genes in *Acidobacteriaceae* and *Holophagaceae* (Supplementary Table 6, Fig. 2). These findings suggest that the second cluster is potentially involved in a range of functions related to nutrient cycling and soil health.

The potential of Acidobacteriota to act in plant polymer degradation

Carbohydrate Active Enzymes (CAZymes) and peptidases play a crucial role in plant growth and development. These enzymes are responsible for the synthesis, modification, and

breakdown of carbohydrates in plants [17]. Our analysis mapped of 39,722 CAZymes distributed across 758 MAGs of Acidobacteriota (Supplementary Table 7). Among these enzymes, a predominance of GH (Glycosyl Hydrolases) and the presence of some PL (Polysaccharide Lyases) in small number were identified. In particular, the GH enzyme family stood out, representing the seven most abundant CAZymes in these genomes. These enzymes were identified as GH3, GH5, GH13, GH23, GH28, GH29, and GH109, with varying in numbers, ranging from 832 to 4299 Hits, and abundance within the phylum, being *Acidobacteriaceae* and *Bryobacteriaceae* exhibited a higher abundance of these enzymes in their genomes (Supplementary Table 7, Supplementary Fig. 4A, Fig.3A). Particularly noteworthy was the significant presence of GH109 in both families, suggesting their involvement in activities related to α-N-acetylgalactosaminidase. Additionally, *Acidobacteriaceae* showed high quantities of GH3 and GH13 enzymes, indicating their capacity for cellulose, glucan, starch, and peptidoglycan degradation. In addition to these activities, the seven most abundant enzymes also individually contribute to the degradation of mannan, chitooligosaccharides, pectin, and fucose.

A total of 32,566 proteolytic enzymes were mapped, grouped according to their catalytic type, including five families of aspartic peptidases (A), 18 families of cysteine peptidases (C), 62 families of metallo peptidases (M), 2 families of asparagine peptide lyases (N), 1 family of mixed peptidase (P), 37 families of serine peptidases (S), 6 families of threonine peptidases (T), and 3 families of peptidases with unknown catalytic type. Additionally, 761 inhibitors belonging to 5 distinct families were mapped in these OTUs (Supplementary Table 8).

Among the peptidases, the eight most abundant families were selected from the mapped genomes. These families include: C26, corresponding to the family of gamma-glutamyl hydrolases; M103, containing the peptidase TldD; M13, containing metalloendopeptidases with restricted activity to substrates smaller than proteins; M20A, containing carboxypeptidases; M38, corresponding to the family of beta-aspartyl dipeptidases; S09X, containing a diverse set of serine-dependent peptidases; S12, containing serine-type D-Ala-D-Ala carboxypeptidases; S15, containing Xaa-Pro dipeptidyl peptidase; S33, containing exopeptidases that act on the N-terminus of peptides; and T03, which exhibits activities of aminopeptidase and aminotransferase (Fig. 3B). The *Acidobacteriaceae* family hosted most proteolytic enzymes, followed, in smaller quantities, by the *Koribacteriaceae*, *Pyrinomonadaceae*, and *Bryobacteriaceae* families, respectively (Supplementary Fig. 4B, Fig. 3B).

Acidobacteriaceae showed a higher number of enzymes related to plant polymer degradation, such as GH2, GH42, GH13, GH144, GH149, GH6, GH9, GH55, GH77, GH97, GH1, GH23, GH15, GH31, GH10, GH115, GH30, GH67, GH8, GH74, GH120, and GH39 (Supplementary Table 7), indicating the possible involvement of *Acidobacteriaceae* in the degradation of cellulose, pectin, cellobiose, galactose, mannose, glucuronic acid, glucan, and arabinofuranose cleavage (Fig. 3C). Furthermore, the families *Bryobacteraceae*, *Koribacteraceae*, and *Pyrinomonadaceae* also exhibited a significant quantity of these enzymes. Altogether, this highlights the significant potential of this bacterial family to act in plant polymer degradation.

Role of Acidobacteriota in Biogeochemical Cycles

Our findings highlight the significant role of organisms from the phylum Acidobacteriota in four biogeochemical cycles: carbon, nitrogen, sulfur, and elemental metal cycles (Supplementary Table S9, Fig. 4). We identified 618 MAGs from this phylum that contribute to these reactions, and most of them carry *acs* genes, which are important for acetate oxidation (Fig. 4A). Moreover, we observed a similar pattern of classes participating in the sulfur cycle pathways, which exhibited substantial Acidobacteriota involvement (Supplementary Fig. 5). Within this cycle, we identified eight classes of the Acidobacteriota, comprising a total of 436 MAGs engaged in sulfur oxidation reactions (Fig. 4B). Additionally, in pathways involving the reduction of sulfur, 278 MAGs were found, which shared the same classes present in the oxidation pathways, differing only by the inclusion of subgroup 26. Notably, the MAGs primarily encoded genes associated with sulfide oxidation (*fccB*, *sqr*), sulfite reduction (*dsrABD*, *asrABC*), sulfur oxidation (*sdo*, *sor*), sulfur reduction (*sreABC*, *sor*), thiosulfate oxidation (*soxBCY*), sulfate reduction (*aprA*, *sat*), and thiosulfate disproportionation (*aprA*).

Furthermore, a significant contribution of Acidobacteriota to nitrogen balance in the environment (Supplementary Fig. 5). While individuals from all eleven taxonomic classes of the phylum Acidobacteriota participated in eight stages of the nitrogen cycle, they were notably absent in the ammonia oxidation pathway (Fig. 4C). Conversely, in pathways crucial for nitrogen absorption and assimilation by plants and microbiota, such as nitrification (*norBC*) and nitrite ammonification (*nrfADH*, *nirBD*), a higher number of taxa were predicted to be involved. Specifically, 308 MAGs were associated with nitrification, and 186 MAGs were associated with nitrite ammonification (Fig. 4C). Moreover, we identified 17 MAGs belonging to the classes *Acidobacteriae* and *Holophagae* that possess the capability to engage in biological nitrogen fixation. Marker genes involved in biological nitrogen fixation, namely *nifDK* and *nifH*, were successfully identified in these MAGs.

Lastly, with regard to elemental metals, all 746 MAGs were found to participate in redox reactions involving iron (Fig. 4D). However, the elements selenium and arsenic displayed minimal participation from individuals within the taxonomic classes. Specifically, only one MAG was identified in the arsenite oxidation pathways (Fig. 4D). In summary, our research underscores the crucial role played by Acidobacteriota in important biogeochemical processes. Their involvement spans the carbon, nitrogen, sulfur, and elemental metal cycles, demonstrating their significance in maintaining environmental balance and ecosystem functioning.

Expansion of Nitrogen-fixing activity in Acidobacteriota

Although nitrogen-fixing genes have been acknowledged for their significance [18], their distribution and role within the phylum remain largely unexplored. Consequently, additional research is necessary to uncover the prevalence of these genes in various bacterial families and their functional implications in different ecological settings. A search was conducted to identify evolutionary conserved genes associated with nitrogen fixation in the genomes under study. A total of seventeen MAGs were discovered to contain genes linked to nitrogen fixation. Upon analysis, it was determined that the genetic capacity for N_2 fixation was linked to the following families: *Holophagaceae* (6), *Acidobacteriaceae* (5), *Bryobacteraceae* (5), and *Koribacteraceae* (1). Notably, *Holophagaceae* displayed the highest prevalence, while *Koribacteraceae* exhibited the lowest frequency. At genus level, we identified *Holophaga* (4), *Geothrix* (2), *Terracidiphilus* (1), and *Granulicella* (1). Furthermore, it was observed that 8 MAGs did not belong to any classified genera within the phylum (Supplementary Table S9). The majority of genes were associated with MAGs predominantly distributed in terrestrial environments (15), with a smaller presence in aquatic environments (2) and only one MAG identified in engineered environments (1). Within the terrestrial environment, some operational taxonomic units (OTUs) were found in peat soils (2), while the majority were detected in permafrost (10) (Supplementary Fig. 6A).

We mapped eight nif genes directly involved in the N2 fixation, including *nifA* (5), *nifB* (14), *nifD* (17), *nifH* (6), *nifJ* (16), *nifK* (16), *nifS* (12), and *nifW* (4). The *nifD* gene was found in all analyzed genomes, either in its complete form or as fragments. Notably, OTU-17217, affiliated with the *Terracidiphilus* genus, exhibited a higher abundance of nitrogen assimilation genes, as well as other genes associated with the organization of nitrogen fixation operons (Supplementary Fig. 6B) [19]. However, contig fragmentation resulted in several reference sequences containing more than one fragment per contig. Overall, our analysis provided a

comprehensive understanding of how nitrogen-fixing genes are distributed, their functional implications, and their organization within various ecological settings in the Acidobacteriota phylum.

Evidence of transcriptional activity of PGPTs in metatranscriptomic data

Lastly, we employed metatranscriptomic data from *Acidobacteriaceae* bacterium URHE0068 and *Chloracidobacterium thermophilum* to investigate their transcriptional activity related to PGPTs (Fig. 5). These data were collected from various environments, including grassland soil microbial communities (Fig. 5A), *Avena fatua* rhizosphere and bulk microbial communities (Fig. 5B), and anoxygenic and chlorotrophic microbial communities from Yellowstone National Park (Fig. 5B), being the first two RNA-seq for *Acidobacteriaceae* bacterium (Supplementary Table S10).

Our analysis revealed that both species exhibited metabolic activity in these environments, as evidenced by the high expression levels of general metabolism (Fig. 5). Furthermore, PGPTs associated with phosphorus solubilization, siderophore production, EPS, and genes related to plant growth hormones exhibited above-average relative transcriptional activity, particularly in terrestrial ecosystems associated with plants. Interestingly, several peptidase genes showed the highest levels of transcriptional activity in grassland soil and the anoxygenic community (Fig. 5B, C). Additionally, we observed differential expression of the aminopeptidase family M9 gene between soil associated with root (rhizosphere) and soil not associated (bulk soil) of *A. fatua*. Similarly, phosphorus transport genes exhibited above-average relative transcriptional activity in the rhizosphere of the host plant. A high relative transcriptional activity of genes involved in iron metabolism (such as bacterioferritin, ferritin-like, and iron complex outer membrane receptor) in the anoxygenic microbial communities of *C. thermophilum* (Fig. 5C). Overall, our findings shed light on the transcriptional activity of Acidobacteriota and their involvement in PGPTs. These results highlight their potential contributions to nutrient cycling and plant-microbe interactions in various ecosystems.

Discussions

Acidobacteriota is a prominent bacterial taxon widely observed in soils worldwide, often constituting a substantial portion of the total bacterial community, reaching up to 52% in some cases [6, 7, 20]. These bacteria are commonly associated with the rhizosphere, the region surrounding plant roots, where intricate interactions between plants and microorganisms occur, mutually influencing each other. Despite their high abundance in this environment, our

understanding of how Acidobacteriota members specifically interact with plants and contribute to biogeochemical processes remains limited. Here, we took advantage of metagenome data spanning various environmental contexts within the Acidobacteriota to elucidate the roles played by this phylum in plant-microbe interactions and their impact on the biogeochemical dynamics.

We explored 758 MAGs from GEM catalog and 121 RefSeq genomes of Acidobacteriota from wide range of habitats worldwide, which allowed us to explore the ecological roles and adaptations of this phylum in different environments. Our comprehensive analysis of the metagenome data allowed us to identify a wide distribution of PGPTs genes within the Acidobacteriota phylum. Notably, the colonization and competitive exclusion classes of PGPTs were found to be present in numerous Acidobacteriota members. The colonization PGPT class encompasses genes that are involved in establishing robust interactions between Acidobacteriota and plants. These genes contribute to processes such as adhesion, biofilm formation, and the ability to colonize plant tissues [21–23]. Similarly, the competitive exclusion PGPT class plays a vital role in plant-microbe interactions. Genes within this class enable Acidobacteriota to outcompete other microorganisms [24]. Previous studies have provided evidence that members of the Acidobacteriota possess multiple conjugative and integrative elements within their genomes and upregulate the expression of specific genes to persistence in the soil environments [25, 26].

Our analysis identified four major Acidobacteriota families, namely *Acidobacteriaceae*, *Bryobacteraceae*, *Koribacteraceae*, and *Pyrinomonadaceae*, which exhibited a considerable number of 86 keys PGPTs genes. These families have the potential to directly interact with plants, suggesting their ability to promote plant growth and development. The presence of these PGPT genes within these Acidobacteriota families highlights their importance in establishing beneficial relationships with plants and their potential contributions to enhancing plant growth. *Acidobacteriaceae*, a family within the SD1 subdivision of the Acidobacteriota phylum, stands out as the most abundant and extensively cultured group within the phylum [16, 27]. It is considered the dominant division of Acidobacteriota. The initial discovery of plant growth-promoting abilities within the Acidobacteriota phylum was associated with *Acidobacteriaceae*. Subsequently, another study reported the presence of plant growth-promoting members in two additional families: *Bryobacteraceae* (SD3) and *vicinamibacteraceae* (SD6). Although no typical bacterial plant growth-promoting traits were identified in vitro for these strains, they

were found to enhance the growth of duckweed and exhibited the ability to colonize plant surfaces [28].

This study built upon previous evidence of nitrogen fixation by Acidobacteriota [18]. We discovered 17 metagenome-assembled genomes (MAGs) exclusively belonging to the *Acidobacteriaceae* and *Holophagae* families, representing nine novel genera involved in this important process. Interestingly, most of these genera have not been taxonomically classified yet. We did not observe any direct correlation, beyond the family level, between these genera and their involvement in nitrogen fixation. Therefore, we hypothesized that this phenomenon could potentially be attributed to horizontal gene transfer events. However, due to the inherent limitations of MAGs, we were unable to definitively determine the origin of these nitrogen-fixing capabilities. Further investigations are needed to fully understand the genetic basis and evolutionary dynamics of nitrogen fixation in these Acidobacteriota genera.

We profiled the carbohydrate degradation and peptidase activities of Acidobacteriota. Our analysis revealed that the enzymatic repertoire of Acidobacteriota, being some members encode up to 200 CAZymes and more than 100 peptidases, specifically in the four major families associated with PGPTs, was enriched with genes involved in the degradation of plant polymers. This suggests a potential connection between carbohydrate degradation and PGPTs in Acidobacteriota. However, it is important to consider that this correlation could be influenced by the dataset itself, as the number of MAGs available for these specific Acidobacteriota families may are abundant compared to others.

Acidobacteriota may play a significant role in biogeochemical cycling, although many aspects of their specific contributions are still not well understood [27]. Here, we applied an in-depth insight into the diversity of the phylum and their role in carbon, nitrogen, sulphur, and trace metal cycling. Our analysis revealed a significant presence of Acidobacteriota in both the sulfur and nitrogen cycles, which are vital processes in ecosystems, incorporating essential nutrients into the soil through sulfur oxidation and nitrogen fixation [29, 30], making them available to plants. Finally, our study provided evidence of the metabolic activity of Acidobacteriota members in diverse environments, demonstrating their contribution to the expression of crucial genes involved in PGPTs. These findings shed light on the roles played by Acidobacteriota in plant-microbe interactions and their impact on biogeochemical processes. The transcriptional activity of highly expressed core genes further emphasized the metabolic relevance of Acidobacteriota in different environments.

We aimed to shed light on the role of specific families within Acidobacteriota in plant-microbe interactions and their potential impact on biogeochemical processes that contribute to soil functioning (Fig. 6). We focused on six representatives of Acidobacteriota known to have a more direct relationship with plant interactions. These strains possess the ability to form flagella, enabling them to encounter root exudates and mucilage components [31]. Chemotaxis driven by flagella likely plays a crucial role in root colonization. Additionally, we found a consensus among these members regarding the presence of the high-affinity phosphate transport system PstSCAB, located on the outer side of the cell membrane (Fig. 6). This indicates their ability to transport inorganic phosphate [32]. Furthermore, we demonstrated that these bacterial strains may degrade complex plant polymers such as chitin, cellulose, and hemicellulose, as reported in previous studies [33–35].

We also observed that the solubilization of insoluble potassium (K) is a common trait among these representatives, with the ability to convert it into a soluble form suitable for plant growth [36]. Additionally, we found that some members of Acidobacteriota are capable of surviving in environments contaminated with heavy metals, suggesting their potential as bioremediation agents. Furthermore, the presence of CRISPR loci and other antiviral mechanisms in *Acidobacteriaceae* and *Pyrinomonadaceae* members may provide a selective advantage in the interaction between bacteria and phages in the soil environment [37].

In the biogeochemical context, we observed that many members of Acidobacteriota are involved in nitrogen transformation processes such as ammonification, nitrification, and biological nitrogen fixation, as well as sulfur-related processes including sulfate reduction, sulfur oxidation, and sulfide oxidation (Fig. 6). Although it is challenging to establish a direct effect of these bacterial plant growth-promoting traits *in vitro* [28], either due to the lack of understanding of their mechanisms or the difficulty of isolating and cultivating in laboratory conditions, the transformations performed by Acidobacteriota in oxidizing essential elements for plant growth may improve overall plant growth. Altogether, this study highlights the importance of this intriguing phylum and opens new avenues for future research aimed at exploring specific Acidobacteriota taxa and their implications in plant interactions for applications in agriculture and environmental sustainability.

Material and Methods

Dataset compilation

We retrieved 758 MAGs from the publicly available GEM catalog (https://portal.nersc.gov/GEM/) [38]that were taxonomically affiliated with the phylum Acidobacteriota. To gain insights into these MAGs, we employed QUAST v5.0.2102 to calculate fundamental features such as size, GC content, N50 value, and other relevant metrics. In addition to the MAGs, we also acquired 129 genome RefSeq from the National Center for Biotechnology Information (NCBI). These RefSeq were selected for the purpose of comparison and further analysis. By including these additional genomes, we aimed to broaden the scope of our investigation and potentially uncover valuable similarities or differences between the MAGs and the RefSeq.

We employed a two-step approach to obtain biogeographic and environmental metadata. First, we utilized the genome ID information from the GEM catalog and searched a search on the Metagenome Bin Search tool available on Integrated Microbial Genomes and Microbiomes (JGI IMG, https://img.jgi.doe.gov). This allowed us to gather relevant biogeographic and environmental information associated with the MAGs. Additionally, for the RefSeq genomes, we accessed the metadata available on Biosample on NCBI, which provided us with details of the geographical origin and environmental context of these genomes.

Phylogenetic analyses

CheckM v1.0.13 [39]with the tree_qa command was employed to extract 43 single-copy, protein-coding marker genes. These marker genes were utilized to assess phylogenetic markers in a dataset comprising 758 assembled metagenomic bins and 121 genomes. The concatenated protein alignments of 43 universal marker genes obtained from CheckM were subsequently used to reconstruct maximum-likelihood phylogenetic tree using IQ-TREE v1.6.11 [40]. The reconstruction process incorporated specific parameters ('-m TEST -bb 1000') to ensure accurate tree generation. The phylogenetic tree was uploaded to iTOL [41], where it underwent visual annotation, such as color coding and heatmaps were applied.

Annotations and metabolic predictions

MAG gene prediction was determined using Prokka v1.14.5 [42] with specific parameters ('—metagenome') then annotated by KOfam and custom HMM profiles within METABOLIC v.4.0 [43] and eggNOG-emapper v.2.1.2 [44] with default settings. Additionally, METABOLIC-G v.4.0 was used to profile metabolic and biogeochemical traits, and functional networks in the Acidobacteriota MAGs and genomes. dbCAN2 [45] within METABOLIC v.4.0 was used to annotate proteins carbohydrate-active enzymes (CAZymes) using default thresholds.

Peptidases were searched against MEROPS 'pepunit' database [46] also implemented in METABOLIC. The annotations of genes of interest were compared among the outputs of different annotation tools. Amino acid sequences of selected MAGs were used to detect known prokaryotic antiviral systems using DefenseFinder (https://defense-finder.mdmparis-lab.com) [48].

Identification of plant growth-promoting traits (PGPTs)

We performed alignment of MAGs' protein sequences to identify genes related to plant growth-promoting traits (PGPTs) within the genomes. This alignment was carried out using a combination of BLASTP and HMMER tools in PGPT-Pred database from PLaBAse [49]. In addition, to ensure accuracy, the annotations were further validated through BLASTP searches against the NCBI nonredundant protein, RefSeq, and UniprotKB/Swiss-Prot databases. Hits with an E-value <1e−5 were considered as significant for both approaches. To determine the MAGs potential to interact with plant, we established a criterion based on the presence/absence of 86 PGPTs genes involved in nitrogen fixation, phosphorus solubilization, as well as the production of EPSs, siderophore and plant growth hormones. Specifically:

- Nitrogen-fixing genes: *nifA, nifB, nifD, nifE, nifF, nifH, nifHD1, nifHD2, nifJ, nifK, nifM, nifN, nifQ, nifS, nifT, nifU, nifV, nifW, nifX, nifZ.*
- *Exopolysaccharide production: epsE, epsD, epsF, epsH, epsI, epsJ, epsL, epsM, epsN, epsO.*
- Root colonization by nodulation: *nodA, nodB, nodC, node, nodF, nodI, nodJ, nodU, nod, nodT, nolM, noeA, noeB, noeC, noeD, noeE, nodN, nodN_like, nodO, nodP, nodS, nodS_like, nodY, nodZ, nodV, nodV_like, nodW, nodX, nodX, nodO.*
- Oxidative stress|ROS scavenging: *sodN, sodC, sod3.*
- Iron acquisition: *lipA, lipB, lipL, lipL2, lipM, lplA.*
- Salinity stress-potassium transport: *kdpA, kdpB, kdpC, kdpD, kdpE, kdpF.*
- Plant embryogenesis-spermidine: *puuA, puuB, puuC, puuD, puuE, pup*
- IAA related tryptophan metabolism*: trpA, trpB, trpC, trpCF, trpD, trpE, trpEG, trpG, trpDG, trpF, trpS, trpR*

See Supplementary Table 6 for more details. The frequency of PGPT genes was transformed into ln(x) values, and a heatmap was generated using Clustvis v.1.0 (https://biit.cs.ut.ee/clustvis/) [50]. To *nif*-genes analysis of 17 MAGs with role in N2 fixation

were selected. Nif-cluster were extracted and syntenic analysis was performed using clinker and clustermap.js [51]).

Metatranscriptomic analysis

Publicly available metatranscriptomic data from JGI IMG was used to investigate the expression of PGPT genes, as well as metabolic and biogeochemical traits. We specifically searched for Acidobacteriota RNASeq Expression Studies, and *Acidobacteriaceae* bacterium URHE0068 and *Chloracidobacterium thermophilum* B were considered. We selected the studies "*Avena fatua* rhizosphere and bulk microbial communities" (submission ID 49153, 48966, 48929, 49411, 49409, and 49669) and "Grassland soil microbial communities" (submission ID 49669, 49668, 49671, 49675, 49677, 49686) for *Acidobacteriaceae* bacterium URHE0068. In addition, we chose the study "Anoxygenic and chlorotrophic microbial mat microbial communities from Yellowstone National Park" (submission ID 64582, 64581, and 64584) for *C. thermophilum*.

Subsequently, we mapped PGPT genes, metabolic traits, and biogeochemical traits in these selected studies. To quantify gene expression, read counts per transcript were normalized by the total number of samplings reads and the length of each transcript. FPKM values were then calculated using the formula FPKM = (Number of mapped reads for the gene / Length of the gene in KB) / (Total number of mapped reads in the sample / 1 million). This normalization and calculation procedure allowed for the comparison of gene expression levels across different samples and genes of interest.

Data availability

The genomes used in this study are publicly available https://portal.nersc.gov/GEM/. Source data with genes associated with plant growth-promoting traits (PGPTs) in Acidobacteriota MAGs are available in the Zenodo repository: https://doi.org/10.5281/zenodo.7957159. The datasets generated and analyzed during the study are provided with this paper in supplementary information.

References

1. Kishimoto N, Kosako Y, Tano T. Acidobacterium capsulatum gen. nov., sp. nov.: An acidophilic chemoorganotrophic bacterium containing menaquinone from acidic mineral environment. *Curr Microbiol* 1991; **22**: 1–7.
2. Felske A, De Vos WM, Akkermans ADL. Spatial distribution of 16S rRNA levels from uncultured acidobacteria in soil. *Lett Appl Microbiol* 2000; **31**: 118–122.

3. Ludwig W, Bauer SH, Bauer M, Held I, Kirchhof G, Schulze R, et al. Detection and in situ identification of representatives of a widely distributed new bacterial phylum. *FEMS Microbiol Lett* 1997; **153**: 181–190.
4. Rheims H, Rainey FA, Stackebrandt E. A molecular approach to search for diversity among bacteria in the environment. *J Ind Microbiol* 1996; **17**: 159–169.
5. Dedysh SN, Yilmaz P. Refining the taxonomic structure of the phylum Acidobacteria. *Int J Syst Evol Microbiol* 2018; **68**: 3796–3806.
6. Kuske CR, Barns SM, Busch JD. Diverse uncultivated bacterial groups from soils of the arid southwestern United States that are present in many geographic regions. *Appl Environ Microbiol* 1997; **63**: 3614–3621.
7. Barns SM, Takala SL, Kuske CR. Wide Distribution and Diversity of Members of the Bacterial Kingdom Acidobacterium in the Environment. *Appl Environ Microbiol* 1999; **65**: 1731 LP – 1737.
8. Lee S-H, Ka J-O, Cho J-C. Members of the phylum Acidobacteria are dominant and metabolically active in rhizosphere soil. *FEMS Microbiol Lett* 2008; **285**: 263–269.
9. Quaiser A, Ochsenreiter T, Lanz C, Schuster SC, Treusch AH, Eck J, et al. Acidobacteria form a coherent but highly diverse group within the bacterial domain: evidence from environmental genomics. *Mol Microbiol* 2003; **50**: 563–575.
10. Chen L-F, He Z-B, Zhao W-Z, Kong J-Q, Gao Y. Empirical evidence for microbial regulation of soil respiration in alpine forests. *Ecol Indic* 2021; **126**: 107710.
11. Woodcroft BJ, Singleton CM, Boyd JA, Evans PN, Emerson JB, Zayed AAF, et al. Genome-centric view of carbon processing in thawing permafrost. *Nature* 2018; **560**: 49–54.
12. Crits-Christoph A, Diamond S, Butterfield CN, Thomas BC, Banfield JF. Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. *Nature* 2018; **558**: 440–444.
13. Andrews JH, Harris RF. r- and K-Selection and Microbial Ecology BT - Advances in Microbial Ecology. In: Marshall KC (ed).1986. Springer US, Boston, MA, pp 99–147.
14. Kielak AM, Cipriano MAP, Kuramae EE. Acidobacteria strains from subdivision 1 act as plant growth-promoting bacteria. *Arch Microbiol* 2016; **198**: 987–993.
15. Nayfach S, Roux S, Seshadri R, Udwary D, Varghese N, Schulz F, et al. A genomic catalog of Earth's microbiomes. *Nat Biotechnol* 2021; **39**: 499–509.
16. Kielak AM, Barreto CC, Kowalchuk GA, van Veen JA, Kuramae EE. The Ecology of Acidobacteria: Moving beyond Genes and Genomes. *Front Microbiol* 2016; **7**: 744.
17. Drula E, Garron M-L, Dogan S, Lombard V, Henrissat B, Terrapon N. The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res* 2022; **50**: D571–D577.
18. Linta R, Xinning Z. Genome-Resolved Metagenomics Informs the Functional Ecology of Uncultured Acidobacteria in Redox Oscillated Sphagnum Peat. *mSystems* 2022; **7**: e00055-22.
19. Garcia AK, McShea H, Kolaczkowski B, Kaçar B. Reconstructing the evolutionary history of nitrogenases: Evidence for ancestral molybdenum-cofactor utilization. *Geobiology* 2020; **18**: 394–411.
20. Delgado-Baquerizo M, Oliverio AM, Brewer TE, Benavent-González A, Eldridge DJ, Bardgett RD, et al. A global atlas of the dominant bacteria found in soil. *Science (1979)* 2018; **359**: 320 LP – 325.
21. Compant S, Clément C, Sessitsch A. Plant growth-promoting bacteria in the rhizo- and endosphere of plants: Their role, colonization, mechanisms involved and prospects for utilization. *Soil Biol Biochem* 2010; **42**: 669–678.
22. Knights HE, Jorrin B, Haskett TL, Poole PS. Deciphering bacterial mechanisms of root colonization. *Environ Microbiol Rep* 2021; **13**: 428–444.

23. Ishizawa H, Kuroda M, Inoue D, Ike M. Genome-wide identification of bacterial colonization and fitness determinants on the floating macrophyte, duckweed. *Commun Biol* 2022; **5**: 68.
24. Ghoul M, Mitri S. The Ecology and Evolution of Microbial Competition. *Trends Microbiol* 2016; **24**: 833–845.
25. Greening C, Carere CR, Rushton-Green R, Harold LK, Hards K, Taylor MC, et al. Persistence of the dominant soil phylum Acidobacteria by trace gas scavenging. *Proceedings of the National Academy of Sciences* 2015; **112**: 10497–10502.
26. Gonçalves OS, Santana MF. The coexistence of monopartite integrative and conjugative elements in the genomes of Acidobacteria. *Gene* 2021; **777**.
27. Kalam S, Basu A, Ahmad I, Sayyed RZ, El-Enshasy HA, Dailin DJ, et al. Recent Understanding of Soil Acidobacteria and Their Ecological Significance: A Critical Review. *Front Microbiol* 2020; **11**.
28. Yoneda Y, Yamamoto K, Makino A, Tanaka Y, Meng X-Y, Hashimoto J, et al. Novel Plant-Associated Acidobacteria Promotes Growth of Common Floating Aquatic Plants, Duckweeds. *Microorganisms* . 2021. , **9**
29. Vavourakis CD, Mehrshad M, Balkema C, van Hall R, Andrei A-Ş, Ghai R, et al. Metagenomes and metatranscriptomes shed new light on the microbial-mediated sulfur cycle in a Siberian soda lake. *BMC Biol* 2019; **17**: 69.
30. Ågren GI, Wetterstedt JÅM, Billberger MFK. Nutrient limitation on terrestrial plant growth – modeling the interaction between nitrogen and phosphorus. *New Phytologist* 2012; **194**: 953–960.
31. Turnbull GA, Morgan JAW, Whipps JM, Saunders JR. The role of bacterial motility in the survival and spread of Pseudomonas fluorescens in soil and in the attachment and colonisation of wheat roots. *FEMS Microbiol Ecol* 2001; **36**: 21–31.
32. Martin J, Liras P. Molecular Mechanisms of Phosphate Sensing, Transport and Signalling in Streptomyces and Related Actinobacteria. *Int J Mol Sci* 2021; **22**: 1129.
33. Ward NL, Challacombe JF, Janssen PH, Henrissat B, Coutinho PM, Wu M, et al. Three genomes from the phylum Acidobacteria provide insight into the lifestyles of these microorganisms in soils. *Appl Environ Microbiol* 2009; **75**: 2046–2056.
34. Rawat SR, Männistö MK, Bromberg Y, Häggblom MM. Comparative genomic and physiological analysis provides insights into the role of Acidobacteria in organic carbon utilization in Arctic tundra soils. *FEMS Microbiol Ecol* 2012; **82**: 341–355.
35. Belova SE, Ravin N V, Pankratov TA, Rakitin AL, Ivanova AA, Beletsky A V, et al. Hydrolytic Capabilities as a Key to Environmental Success: Chitinolytic and Cellulolytic Acidobacteria From Acidic Sub-arctic Soils and Boreal Peatlands. *Front Microbiol* 2018; **9**: 2775.
36. Etesami H, Emami S, Alikhani HA. Potassium solubilizing bacteria (KSB):: Mechanisms, promotion of plant growth, and future prospects A review. *J Soil Sci Plant Nutr* 2017; **17**: 897–911.
37. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, et al. Systematic discovery of antiphage defense systems in the microbial pangenome. *Science (1979)* 2018; **359**: eaar4120.
38. Nayfach S, Roux S, Seshadri R, Udwary D, Varghese N, Schulz F, et al. A genomic catalog of Earth's microbiomes. *Nat Biotechnol* 2021; **39**: 499–509.
39. Parks D, Imelfort M, Skennerton C, Philip H, Tyson G. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015; **25**.
40. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* 2020; **37**: 1530–1534.

41. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* 2019.
42. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014; **30**: 2068–2069.
43. Zhou Z, Tran PQ, Breister AM, Liu Y, Kieft K, Cowley ES, et al. METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. *Microbiome* 2022; **10**: 33.
44. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol Biol Evol* 2021; **38**: 5825–5829.
45. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* 2018; **46**: W95–W101.
46. Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res* 2018; **46**: D624–D632.
47. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res* 2020; **48**: D517–D525.
48. Tesson F, Hervé A, Mordret E, Touchon M, d'Humières C, Cury J, et al. Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nat Commun* 2022; **13**: 2561.
49. Patz S, Gautam A, Becker M, Ruppel S, Rodríguez-Palenzuela P, Huson DH. PLaBAse: A comprehensive web resource for analyzing the plant growth-promoting potential of plant-associated bacteria. *bioRxiv* 2021; 2021.12.13.472471.
50. Metsalu T, Vilo J. ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Res* 2015; **43**: W566–W570.
51. Gilchrist CLM, Chooi Y-H. clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics* 2021; **37**: 2473–2475.

Funding

This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq (Process APQ-02381-21), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior/Programa de Excelência Acadêmica-Finance Code 001 (CAPES ProEx grant 23038.019105/2016-86), CAPES-PrInt (process 88887.696147/2022-00) and Fundação de Amparo à Pesquisa do Estado de Minas Gerais—FAPEMIG (Process 402644/2021-2) for the financial support.

Acknowledgements

The authors expressed their gratitude to the technical support team of the Cluster at Universidade Federal de Viçosa. We expressed our appreciation to Felipe Xavier for providing valuable artistic insights into Figure 6 of this paper.

Contributions

577 Conceptualization: OG, and MF. Data curation: OG, AF, ST, LA, TX. Funding acquisition:
578 MF. Investigation:. Methodology:. Project administration: Resources:. Supervision:.
579 Visualization: Writing – original draft:. Writing – review & editing:

580 **Conflict of interest**

581 The authors declare that they have no conflict of interest.

Figures

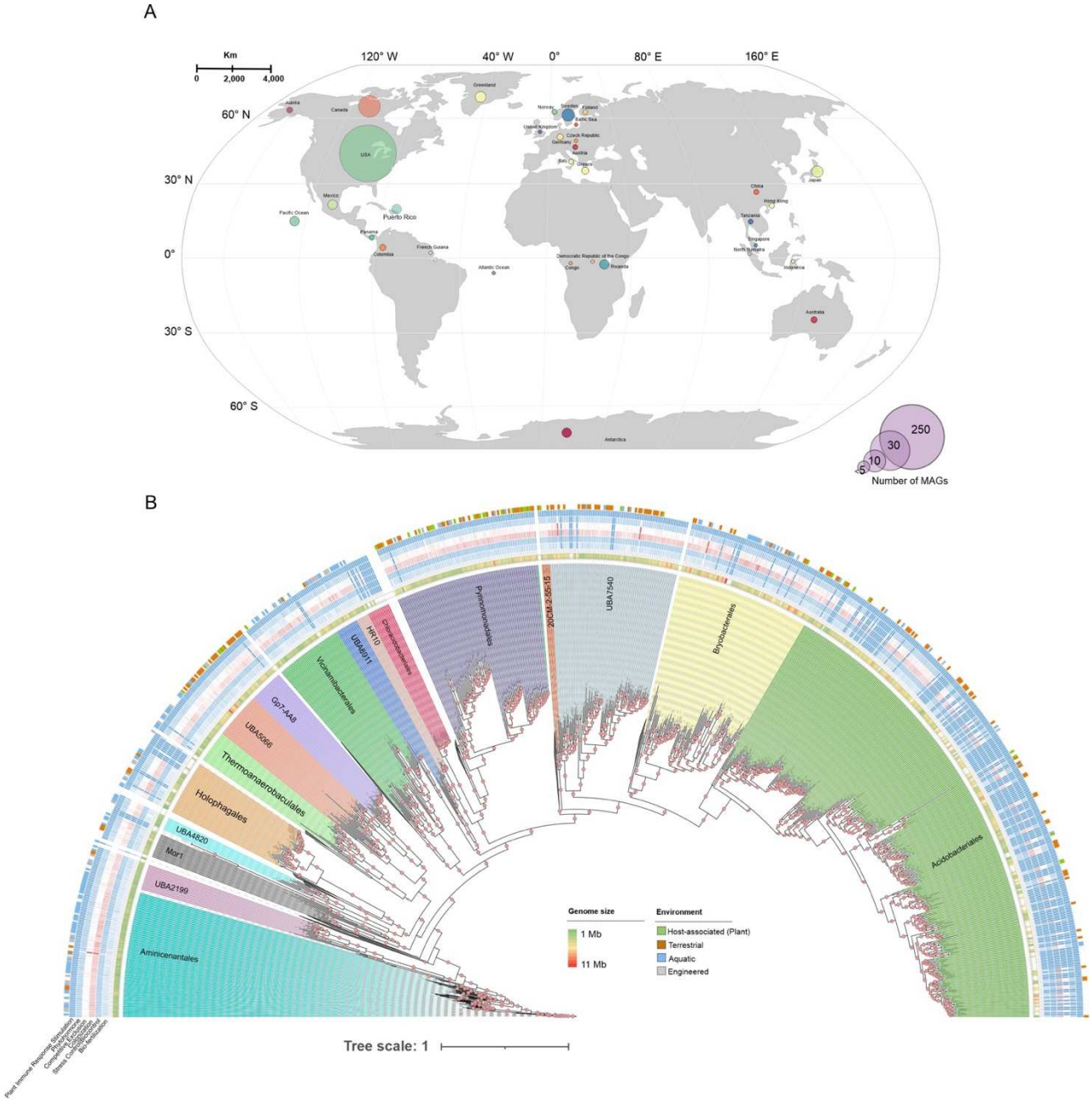


Figure 1. Phylogeny and an overview of their PGPTs potential and global distribution of 885 genomes of Acidobacteriota phylum. (**A**) The global distribution of the Acidobacteriota described in this study in a map generated using 'ggplot2' package in R. The abundance is highlighted in distinct circle size. (**B**) A maximum likelihood phylogenetic tree of 879 genomes including the 758 metagenome assembled genomes (MAGs) from GEM catalog and 121 RefSeq genomes. The phylogeny is based on 43 single-copy, protein-coding marker genes identified using CheckM. Acidobacteriota families are marked in different background colors. The first arc of the tree displays the environmental source of each genome, followed by sequential arcs representing the relative abundance of genes associated with plant growth-promoting traits (PGPTs), and the last arc represents the genome size of each MAGs. Phylogenetic tree was built under the model of rate heterogeneity LG+F+I+64, and a maximum of 1,000 bootstrap replicates. Bootstraps are shown in pink circles. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree.

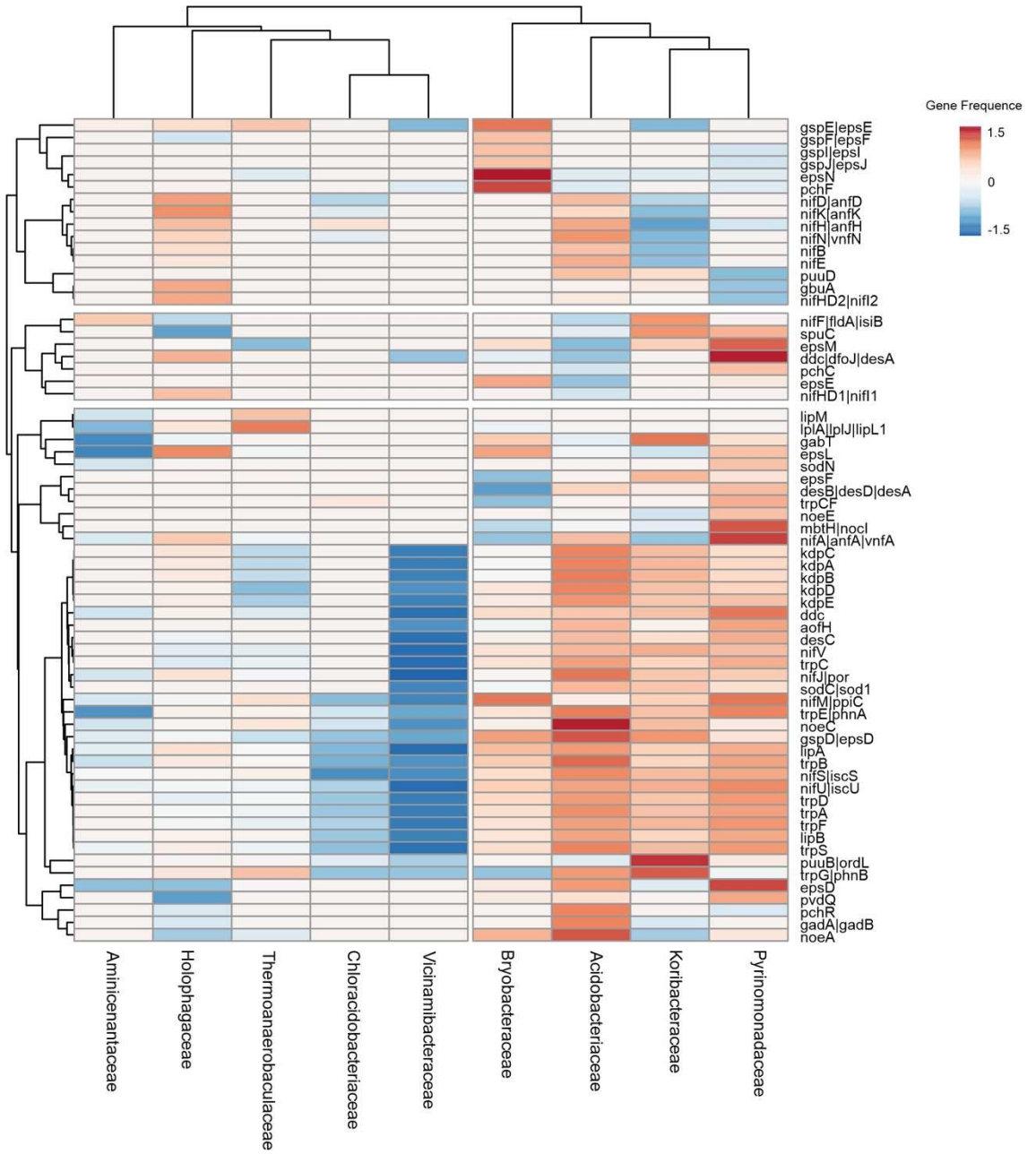


Figure 2. Heatmap depicting the distribution of 86 genes associated with plant growth-promoting traits (PGPTs) that directly contribute to plant growth. The abundance values have been logarithmically transformed (ln(x)-transformed). Row centering and unit variance scaling have been applied to enhance visualization. Missing values were estimated using imputation techniques. The rows and columns have been clustered using correlation distance and average linkage methods.

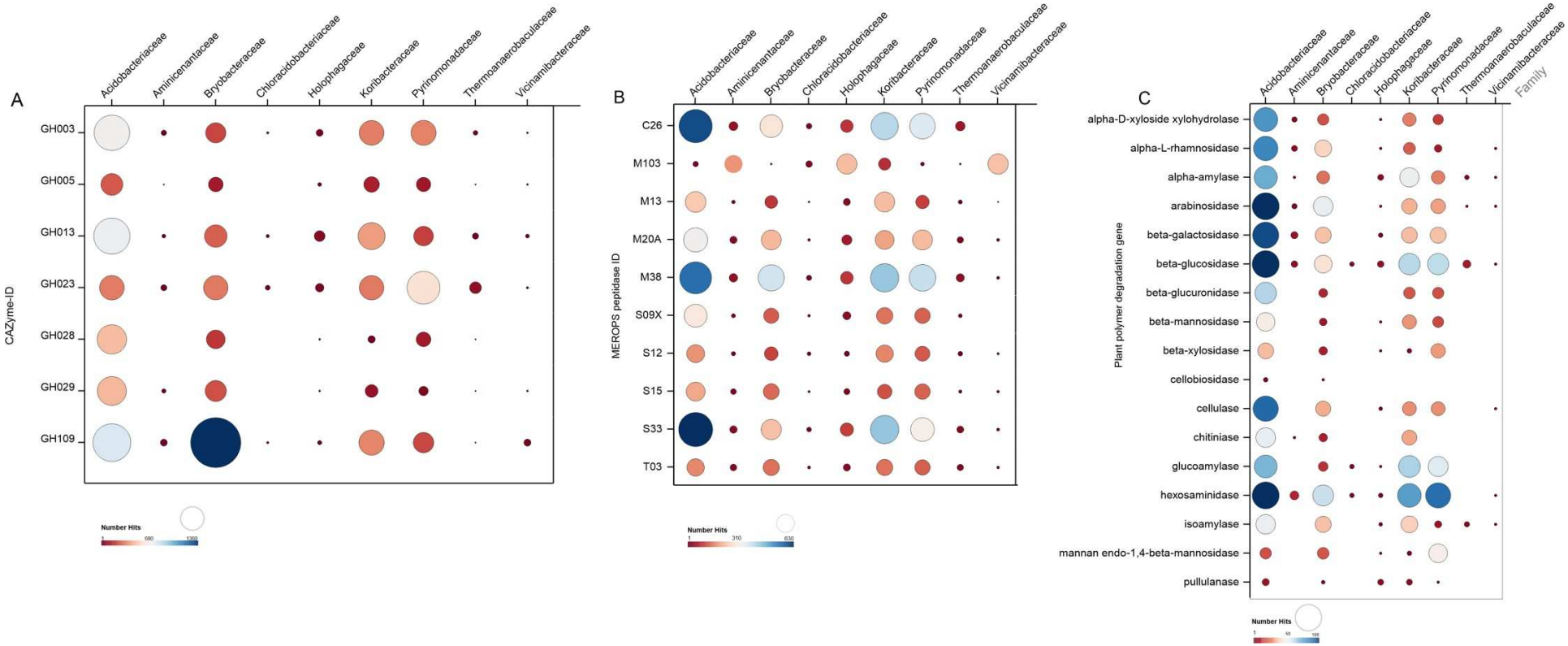


Figure 3. The abundances of genes involved in carbohydrate degradation (A), peptidases (B), and plant polymer degradation (C) across the MAGs. Each circle's size represents the number of hits for the corresponding gene. The abbreviation "GH" refers to glycoside hydrolases. The x-axis represents the genes, while the y-axis displays the taxonomic family of Acidobacteriota. For a comprehensive description of the carbohydrate-active enzymes (CAZymes) and peptidases present in the genomes, please refer to Supplementary Table 7 and Table 8, respectively.

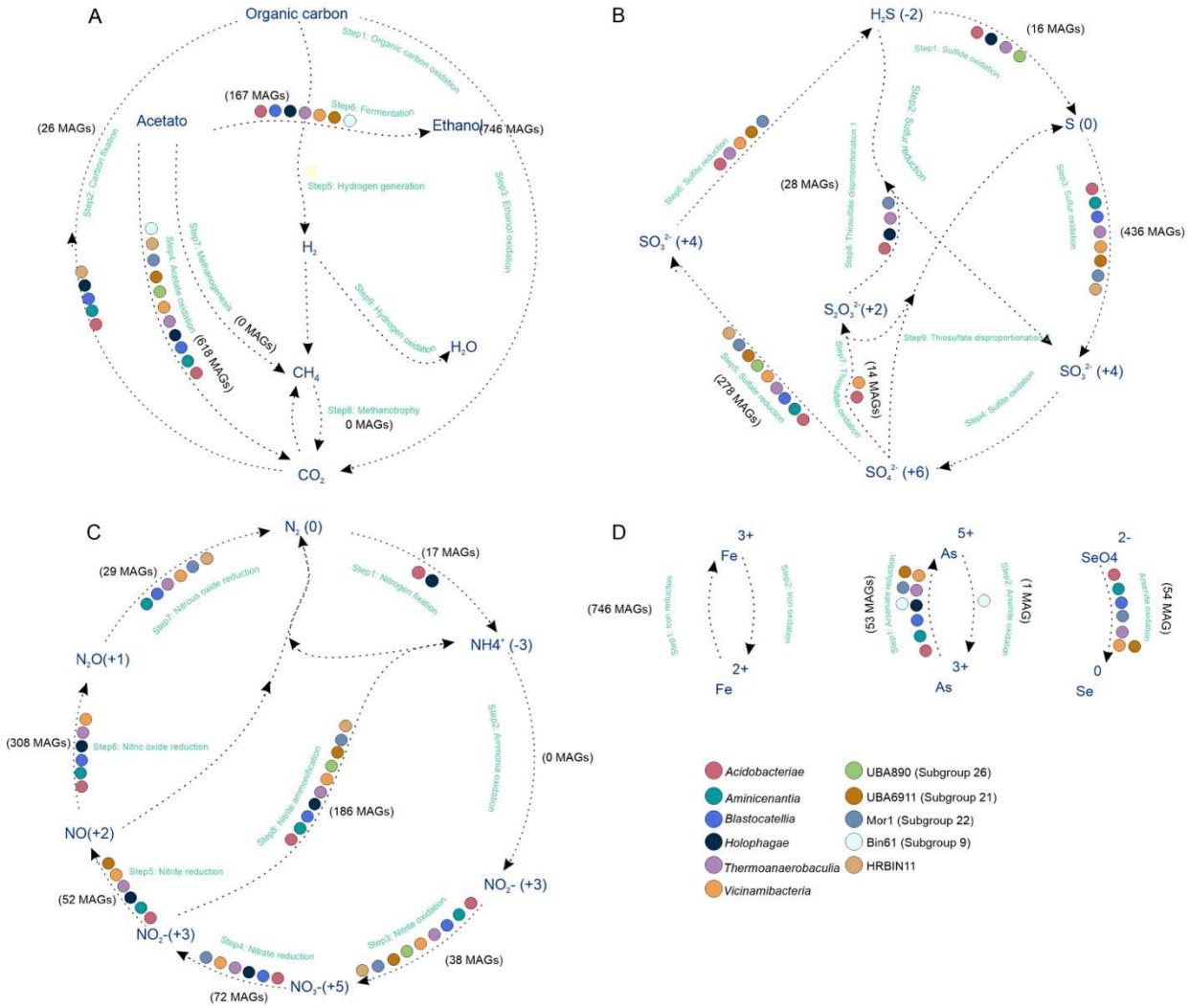


Figure 4. The genomic-based predictions regarding the potential biogeochemical role of Acidobacteriota in various elemental cycles. The taxonomic classes are color-coded to represent their involvement in different steps within the Carbon (**A**), Nitrogen (**B**), Sulfur (**C**), and trace metal (**D**) cycles. The number of Metagenome-Assembled Genomes (MAGs) participating in each element transformation is indicated within parentheses. It is important to note that only pathways encoded in at least one MAG are displayed, and when all MAGs are involved in a particular pathway, taxonomic classes are not color-coded. Each arrow in the figure signifies a single transformation or step within a specific cycle.

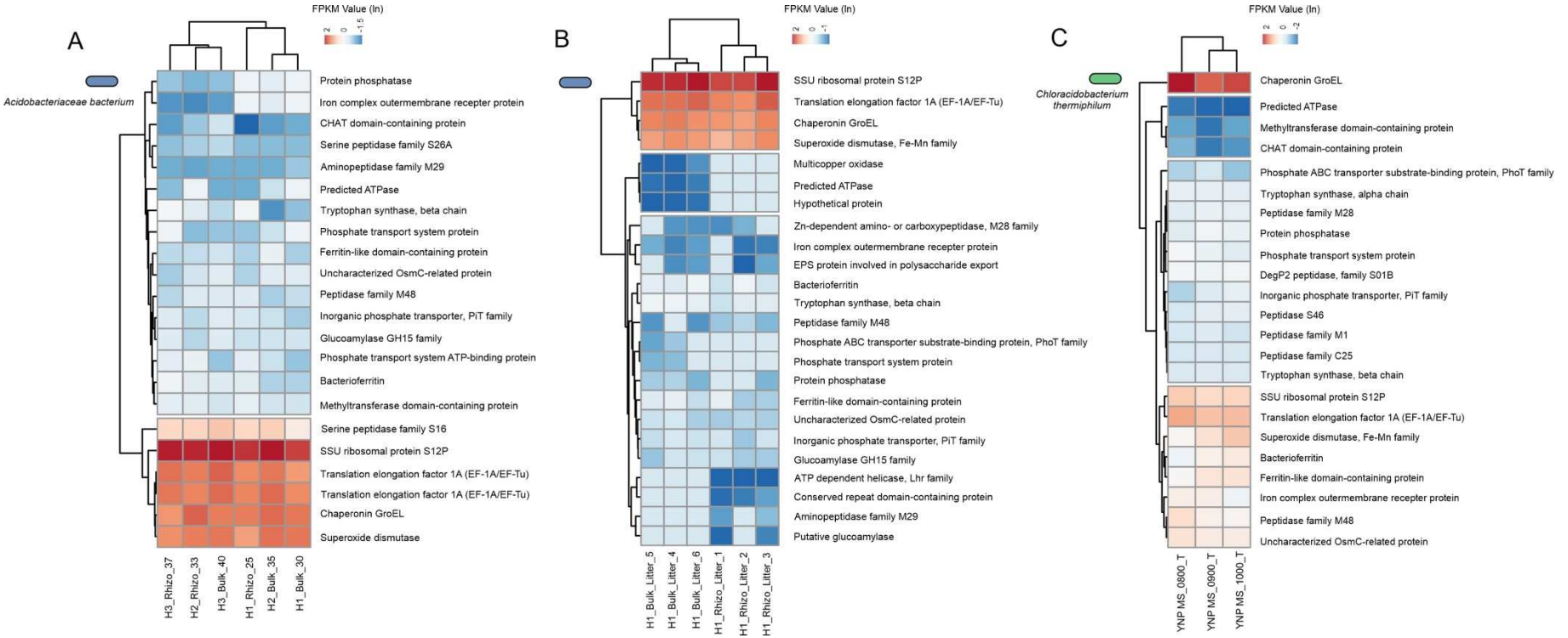


Figure 5. Metatranscriptomic activity of genes associated with PGPTs in two strains of Acidobacteriota across different environmental settings. The expression levels, measured in Fragments Per Kilobase per Million mapped fragments (FPKM), are shown for *Acidobacteriaceae* bacterium URHE0068 in Grassland soil microbial communities (A) and *Avena fatua* rhizosphere and bulk soil microbial communities (B). FPKM values for *Chloracidobacterium thermophilum* B in anoxygenic and chlorotrophic microbial communities (C). The expression levels are reported in ln-transformed FPKM. The rows in the figure represent individual samples, without any correlation, and have been clustered using Euclidean distance and average linkage.

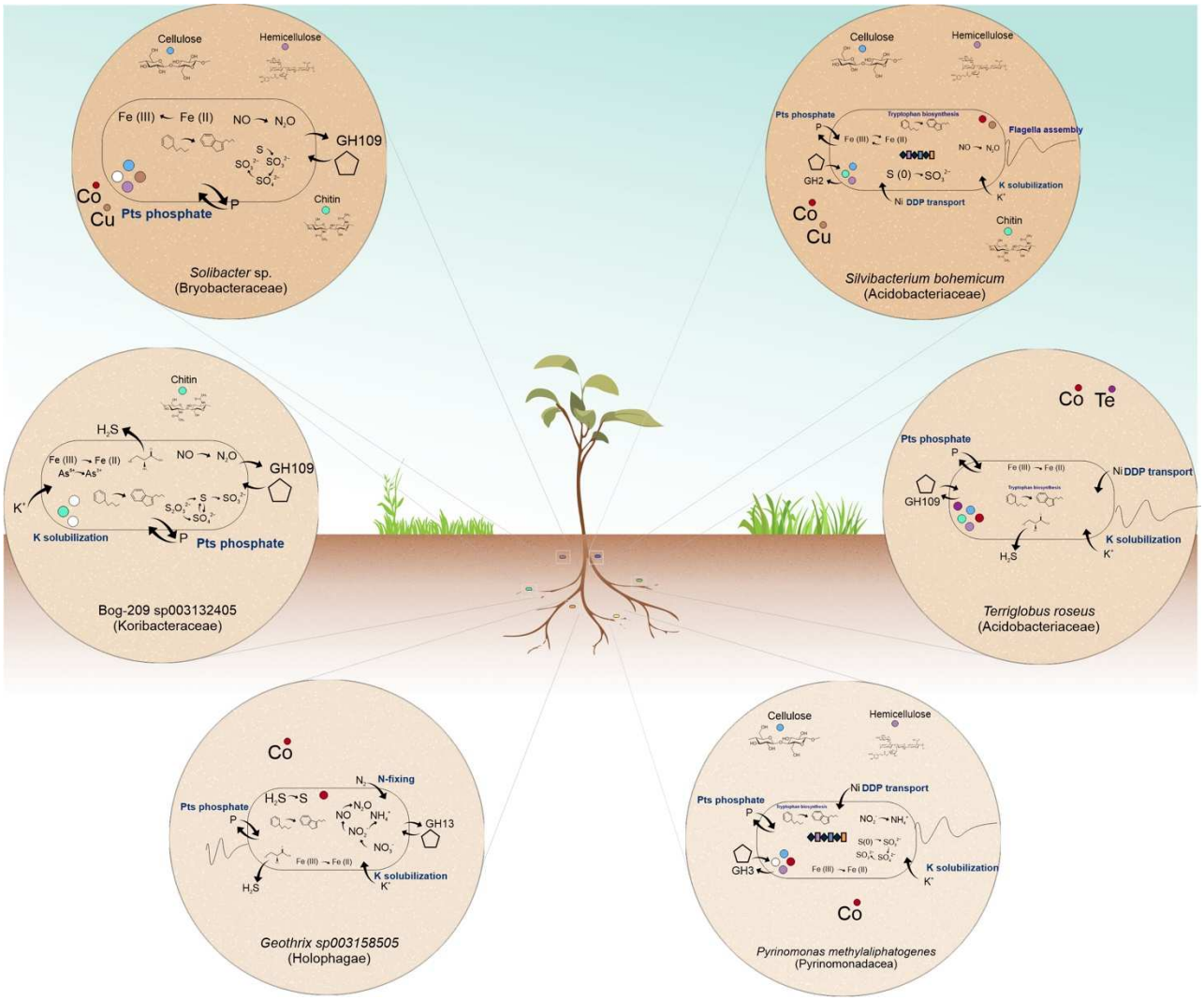


Figure 6. **Schematic illustration highlighting some representatives of Acidobacteriota families involved in plant-microbe interactions and their potential influence on the biogeochemical processes contributing to soil functioning**. Distinct bacterial groups are distinguished by various circle augmentations, while biogeochemical reactions are depicted within each cell. The ability to break down complex carbon compounds and withstand heavy metals is visually represented through circle colors. Curved lines illustrate transformations occurring in the periplasm or involving membrane-bound processes, while outer pentagon shapes represent extracellular CAZYmes. The lower part of the figure showed general microbiome services and plant-specific traits.

General conclusions

The isolation of these remarkable slow-growing bacteria has uncovered a group of novel species with significant potential for enhancing plant growth, particularly in drought conditions. These findings not only provide new insights into their taxonomic and ecological roles but also offer fresh perspectives on their interactions with plants.

The integration of reverse ecology and genomics approaches has proven to be a valuable tool, supplementing traditional methods for selecting microbial species to compose bioinoculants. Additionally, gaining an overview of the natural microbial community offers a valuable means of manipulating and selecting beneficial microorganisms to construct synthetic communities that can improve crop productivity and enhance stress resistance.

These results significantly contribute to advancing our understanding of slow-growing bacteria, their ecological importance, and their potential applications in agriculture. They provide crucial insights into ecosystem dynamics and plant-microbe interactions, laying the foundation for future breakthroughs in this field.

Appendix 1

GONÇALVES, Osiel Silva; SANTANA, Mateus Ferreira. The coexistence of monopartite integrative and conjugative elements in the genomes of Acidobacteria. Gene, v. 777, p. 145476, 2021. https://doi.org/10.1016/j.gene.2021.145476

The coexistence of monopartite integrative and conjugative elements in the genomes of Acidobacteria

Osiel Silva Gonçalves[1] and Mateus Ferreira Santana[1*]

[1]Departamento de Microbiologia, Instituto de Biotecnologia Aplicada à Agropecuária (BIOAGRO), Universidade Federal de Viçosa, Viçosa, MG, 36570-000; Brazil.
[*]Correspondence: mateus.santana@ufv.br (M.F.S.); Tel.: +55 (31) 3612-2452

Highlights

- Ten novel ICEs were identified in the chromosome of Acidobacteria.
- ICEs were found co-existing as three separated monopartite ICEs into the single chromosome.
- ICEs carry a repertoire of genes with potential environmental roles.

Abstract

Soil bacteria can rapidly adapt to environmental perturbations through horizontal gene transfer. Acidobacteria is one of the most persistent dominant phyla in the soil. However, the role of these organisms in terrestrial ecosystems remains elusive. Here we identified and describe the integrative and conjugative elements (ICEs) in the published complete genomes of Acidobacteria. In total, ten novel ICEs were identified, in which nine were found integrated as three separated monopartite ICEs in the single chromosome sequences of three Acidobacteria. These ICEs carry a repertoire of genes with potential environmental roles, including heavy metal resistance, iron uptake, secondary metabolism, and antibiotic resistance. To our knowledge, these are the first evidence of three monopartite ICEs identified in the single chromosome, and this might be due to the absence of recognizable entry exclusion systems. We hypothesis that the coexistence of multiples ICEs in the chromosome of Acidobacteria might reflect a major advantage for the survival, resistance, and persistence of phylum in the environment.

Keywords: Bacterial evolution; Mobile DNA; Soil bacteria.

1. Introduction

Mobile genetic elements (MGEs) are segments of DNA that encode proteins mediating inter- and intragenomic mobility. The prokaryotic MGE pool constitutes the "mobilome" (Frost et al., 2005), which encompasses transposons, insertion sequences, integrons, introns, plasmids, bacteriophages, and integrative and conjugative elements (ICEs) (Siguier et al., 2014). Special attention has been paid to ICEs, as they can play a major role in microbial evolution, allowing microbes to acquire new genes and phenotypes (Johnson and Grossman, 2015). ICEs are the most abundant conjugative elements in prokaryotic genomes (Guglielmini et al., 2011). Most of these elements have been extensively studied through phenotypic analysis, including antibiotic resistance and virulence factors, in bacterial human pathogens (Beker et al., 2018; Farzand et al., 2019). However, ICEs are poorly documented in soil bacteria.

The phylum Acidobacteria is one of the most dominant bacterial taxa found in soils around the world, representing, in some cases, up to 52% of the total bacterial community (Barns et al., 1999; Delgado-Baquerizo et al., 2018). Its members are physiologically diverse and are found in a variety of environments, including hot springs, oceans, and polluted environments. However, despite the ubiquity and abundance of Acidobacteria in soils, little information is available regarding the potential function of this phylum in the environment (Kielak et al., 2016). In this study, we identified and describe ICEs within the complete genomes of Acidobacteria. Ten novel ICEs were identified, in which nine were found integrated as three separated monopartite ICEs in the single chromosome sequences of three Acidobacteria.

2. Materials and methods

2.1. Data

Fourteen RefSeq Acidobacteria complete genomes were retrieved from the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/genome) (Table S1), in August 2019. The common tree taxonomy of Acidobacteria was created in the NCBI Taxonomy Tool (https://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/ wwwcmt.cgi).

2.2. Detection of conjugative systems and delimitation of ICEs

ICEs were identified through a search for conjugative systems with the CONJscan module of MacSyFinder (https://galaxy.pasteur.fr) (Cury et al., 2017); by a standard BLASTn search against the ICEberg database version 2.0 (https://db-mml.sjtu.edu.cn/ICEberg/) (Liu et al., 2018) and ICEfinder (https://db-mml.sjtu.edu.cn/ICEfinder/ICEfinder.html). The genomes

were inspected for MGE-encoding relaxases, type-IV coupling proteins (T4CP), and the type-IV secretion system (T4SS) gene cluster using the oriTfinder tool (http://bioinfo-mml.sjtu.edu.cn/oriTfinder) (Li et al., 2018). Exclusion proteins were screened in the predicted ICEs using BLASTp (McGinnis and Madden, 2004) against the following references proteins (Accession No.): TraT_F (BAA97971), TraS_F (NP_061479), TraS_R100 (NP_052977) (Garcillán-Barcia and de la Cruz, 2008). Hits were regarded as significant when their e-value was $\leq 10^{-5}$ and their alignment covered at least 30% of the protein profile. ICEs sequences in FASTA format were annotated using GeneMarkS (Besemer, 2001). The amino acid output was used to predict the relaxase MOB family in the MOBscan server (https://castillo.dicom.unican.es/mobscan/) (Garcillán-Barcia et al., 2020). Core genes flanking ICE loci were analyzed to determine their upper bounds. This was achieved by aligning ICE upstream regions against closely related genomes where no ICE sequences had been found. We also use the Repeat Finder plugins on Geneious® to identify the attachment sites (attL and attR) delimiting ICEs. The elements were named in accordance with the nomenclature system proposed by Burrus et al., (2002).

2.3. Cargo-associated genes carried by ICEs in Acidobacteria

ICE sequences, in FASTA format, were submitted for secondary metabolite biosynthetic gene cluster identification using antiSMASH 5.0 (https://antismash.secondary metabolites.org) (Blin et al., 2019) with default settings. In addition, ICEs were searched for the presence of antimicrobial resistance-associated genes by BLASTN against the Comprehensive Antibiotic Resistance Database (CARD, http://arpcard. mcmaster.ca) (Jia et al., 2016).

2.4. Element Comparisons

TrbL, integrase, and relaxase amino acid sequences were aligned using CLUSTALW (Larkin et al., 2007) to build pairwise-alignment-based matrix calculations in the Sequence Demarcation Tool (SDT) program v.1 (Muhire et al., 2014). Whole ICE sequence alignment was made with MAFFT (Katoh and Standley, 2013) in the EMBL-EBI (https://www.ebi.ac.uk/Tools/msa/mafft/).

3. Results and discussion

3.1.Co-existence of three monopartite ICEs into the single chromosome of the Acidobacteria

Among the fourteen Acidobacteria genomes analyzed (Table S1), we found ICEs in four Acidobacteria belong to the family of Acidobacteriaceae (Fig. 1A). Genome inspection for

MGEs-encoding relaxases, T4CP, and T4SS cluster using oriTfinder revealed several open reading frames (ORFs) within three regions in the genomes of *Acidobacterium capsulatum* ATCC 51196, *Acidisarcina polymorpha* SBC82, and *Terriglobus saanensis* SP1PR4, indicating the high potential of these regions for self-transferability (Fig. 1B). These regions were manually examined and we found three monopartite ICEs in these Acidobacteria chromosomes, named ICE*Acap1,* ICE*Acap2*, and ICE*Acap3* (in the *A. capsulatum* ATCC 51196), ICE*Apol1*, ICE*Apol2*, and ICE*Apol3* (in the *A. polymorpha* SBC82), as well as ICE*Tsaa1*, ICE*Tsaa2*, and ICE*Tsaa3* (in the *T. saanensis* SP1PR4). In addition to these ICEs, we also mapped a solo ICE in the chromosome of *Granulicella tundricola* MP5ACTX9 (Table 1).

The ICEs in the genome of *A. capsulatum* ATCC 51196 has a similar size pattern. While for *A. polymorpha* SBC82 and *T. saanensis* SP1PR4, the ICEs sizes were widely varied (Table 1 and Fig. 1C). The whole sequence alignment of these ICEs was revealed to share a low degree of nucleotide identity (Fig. 1D). Therefore, we looked more closely at these ICEs, the amino acid sequences of conjugal transfer protein TrbL, integrase, and MOB_F relaxase were aligned and compared. The alignment scores for these gene markers were highly different, being TrbL sequences the ones with the highest similarity (Fig. S1).

ICEs are mobile intercellular elements that can be vertically spread during DNA replication and cell division, or horizontally transferred to another cell by conjugation (Johnson and Grossman, 2015). These MGEs are mediators of horizontal gene flow in bacteria. In general, ICEs are structure organized as contiguous modules integrated within a bacterial genome, which together guarantee the conjugation and maintenance of the element. Recently, a unique group of ICEs was found in *Mesorhizobium* spp.; this group of ICEs was characterized to form three entirely separate but linked chromosomal elements, namely tripartite integrative and conjugative elements (ICE^3s) (Haskett et al., 2016). Haskett et al., (2017) has shown that these elements undergo a series of chromosomal recombinations mediated by integrase proteins to be assembled into a single circular ICE, followed by transfer, integration, and disassembly into three parts in the chromosome of the recipient cell.

To determine whether there is the existence of ICE^3s in the Acidobacteria chromosome, we looked for the attachment site (AttL and AttR) in each identified ICEs (Table S2). In sum, all the regions have no intertwined site for recombination systems, as were previously demonstrated for the ICE^3s (Haskett et al., 2016). Therefore, we conclude the existence of three

separate monopartite ICE integrated into the single chromosome sequences of the Acidobacteria.

3.2.General description of ICEs integrated into A. capsulatum ATCC 51196 genome

ICE*Acap1* is a 66 kb long ICE found in the chromosome of *A. capsulatum* ATCC 51196, a member of the Acidobacteria subdivision 1, first isolated from acidic drainage (Kishimoto et al., 1991). ICE*Acap1* encodes a site-specific tyrosine recombinase, xerC; it is integrated next to the 3′ end of the ssrA tmRNA gene. This integrase is necessary for both the integration and excision of the ICE; it binds to specific DNA consensus sequences, separated from XerD binding sites, to form a heterotetrameric XerC-XerD complex that promotes recombination among DNA substrates (Hallet et al., 1999). ICE*Acap1* also codes for proteins associated with ICE excision and transfer, including a conjugative MOB_F relaxase and elements of the conjugative T4SS component (VirB, VirB4, and TrbI). The ICE carries gene clusters associated with mercury resistance (*merP*, *merC*, and the transcriptional regulator *merR* gene), the cobalt-zinc-cadmium resistance protein cation efflux system (CusA1, CusB1, and outer membrane efflux proteins), and the heavy metal efflux pump, including the copper-transporting ATPase and the putative copper resistance protein B (Fig. 2a and Table 1).

ICE*Acap2* encodes a site-specific tyrosine recombinase that is integrated into next to the 5′ end of the $tRNA^{Met}$ gene. This ICE also encodes proteins associated with excision and transfer, including a conjugative MOB_F relaxase, proteins of the conjugative T4SS component (VirB, VirB4, VirB5, TrbI, and TrbG/VirB9), ParB, and an additional recombinase (RecA). ICE*Acap2* encodes a plasmid stabilization system protein (RelE/ParE family), an mRNA interferase toxin RelE component of a type II toxin-antitoxin (TA) system recognized as a small module composed of a toxin protein, and its cognate antitoxin. The major function of the TA system is related to DNA stabilization through a phenomenon denoted as addiction (Leplae et al., 2011). As a cargo gene, the ICE*Acap2* codes for the putative beta-lactamase protein (Fig. 2a and Table 1), which confers resistance to different beta-lactam antibiotics (penicillin, cephalosporin, and carbapenem) (Wilke et al., 2005).

The ICE*Acap3* lacks an integrase gene. It is located upstream of the RNA polymerase sigma-70 factor gene and downstream of the *dapA2* gene, and it encodes proteins associated with its excision and transfer, including a conjugative MOB_F relaxase and elements of the conjugative T4SS (VirB, VirB4, VirB5, and TrbI), and carries the Lasso peptides (Fig. 2a and Table 1), a

class of ribosomally synthesized and post-translationally modified peptides, as cargo gene (Hegemann et al., 2015).

3.3. General description of ICEs integrated into A. polymorpha SBC82 genome

ICE*Apol1* is a 25 kb long ICE found in the chromosome of *A. polymorpha* SBC82, a member of the Acidobacteria subdivision 1, recovered from lichen-covered acidic soils of a forested tundra (Belova et al., 2018). ICE*Apol1* has a tyrosine-phage integrase directs the integration of this region next to the 3′ end of the $tRNA^{Phe}$ gene. It encodes a conjugative T4SS component (VirB3, VirB4, VirB9, VirB11, TrbI, and TraD) and a TraI protein (DNA helicase). No notable feature associated with a potential environmental role for the ICE*Apol1* was noticed (Fig. 2b and Table 1).

The ICE*Apol2* encodes a tyrosine-phage integrase; it is integrated adjacent to the *tsaA* gene. As core ICE genes, this region encodes conjugative MOB_F relaxase, conjugative T4SS proteins (VirB3, VirB4, VirB5, VirB6, VirB9, VirB11, TrbI, and TraD), and a DNA helicase. Furthermore, this region also codes for the type II toxin-antitoxin system HipA family toxin (Fig. 2b and Table 1). ICE*Apol2* carries an SOS response-associated peptidase (SRAP), a DNA-associated autoproteolytic switch that recruits diverse repair enzymes onto DNA damage (Aravind et al., 2013).

The ICE*Apol3* encodes a resolvase integrase; it is integrated next to the 3′ end of the $tRNA^{Phe}$ gene. Regarding core genes, ICE*Apol3* encodes T4SS component (VirB10), conjugative transfer pilus assembly proteins (TraC and TraD), ParB, and ParC. Also, it carries several cargo genes related to iron uptake (corresponding to the ABC-type Fe3+-siderophore transport system and the periplasmic iron-binding protein FecB) with potential environmental roles (Fig. 2b and Table 1).

3.4. General description of ICEs integrated into T. saanensis SP1PR4 genome

The length of ICE*Tsaa1* is 50 kb; it was located on the *T. saanensis* SP1PR4 chromosome, a member of the Acidobacteria subdivision 1 from tundra soils (Rawat et al., 2012). A site-specific tyrosine recombinase that integrates the region downstream of the 5′ end of the $tRNA^{Asp}$ is encoded by ICE*Tsaa1*. This element contains genes that encode ICE-associated proteins, such as conjugative MOB_F relaxase, T4SS proteins (VirB3, VirB4, VirB5, VirB6, VirB9, and TrbI), ParB-like partition protein, DNA helicase I, DNA primase, and single-strand protein-binding protein. In addition, it encodes a plasmid stabilization system protein (RelE/ParE family). The most notable feature carried by the ICE*Tsaa1* is that the acriflavin resistance

protein (Fig. 2c and Table 1), a multidrug efflux system that is believed to protect the bacterium against hydrophobic inhibitors (Ma et al., 1993).

The ICE*Tsaa2* encodes a site-specific tyrosine recombinase; it is integrated near the 5′ end of the $tRNA^{Val}$ gene. It also encodes a conjugative MOB_F relaxase, T4SS component proteins (VirB3, VirB4, VirB5, VirB6, VirB9, and TrbI), a ParB-like partition protein, DNA helicase I, a DNA primase, an HNH endonuclease, and a single-strand binding protein. The ICE carries iron uptake coding genes (a siderophore-interacting protein, 2OG-Fe(II) oxygenase, and putative iron uptake protein), a transporter of drug resistance, and an efflux pump of cobalt-zinc-cadmium resistance (Fig. 2c and Table 1).

The ICE*Tsaa3* encoding a tyrosine-phage integrase is integrated next to the 3′ end of the $tRNA^{Phe}$ gene. Similar to the ICE*Tsaa2*, ICE*Tsaa3* encodes the same component regarding the ICE core genes. However, there is no significant feature associated with a potential function in the environment (Fig. 2c and Table 1).

3.5. General description of ICE integrated into G. tundricola MP5ACTX9

ICEG*tun1* is the only solo ICE identified in *G. tundricola* MP5ACTX9, another member of the Acidobacteria subdivision 1 from tundra soils (Rawat et al., 2013). ICE*Gtun1* is 38-kb long and encodes a site-specific tyrosine recombinase that integrates the region near the $tRNA^{Ala}$ gene. This ICE codes for a conjugative MOB_F relaxase, T4SS components (VirB3, VirB4, VirB5, VirB6, VirB9, TraG), conjugative transfer proteins (TrbJ and TrbI), a ParB-like partition protein, a single-strand binding protein, and the plasmid mobilization relaxosome protein MobC. Also, ICE*Gtun1* carries several cargo genes, such as those for an ABC transporter protein and cobalt-zinc-cadmium resistance efflux pump proteins (Fig. 2d and Table 1).

Altogether, these ten novel ICEs carry a pool of poorly characterized genes with unknown functions that might enhance bacterial fitness or play a critical role in the maintenance of these elements (data not shown).

Bacteria in the soil are subjected to various environmental perturbations. A rapid adaptation to persist in the environment might be achieved through horizontal gene transfer (Koskella and Vos, 2015). Acidobacteria is one of the most dominant bacterial taxa found in soils worldwide. However, how the members of this phylum, considered to be slow-growing bacteria, are so

abundant, and what role they play in the functioning of terrestrial ecosystems remains elusive (Kielak et al., 2016).

The presence of MGEs affecting the structure and plasticity of Acidobacteria genomes has been observed in other studies (Challacombe and Kuske, 2012; Eichorst et al., 2018). Here, we identified and characterized ten ICEs in the Acidobacteria carrying a repertoire of genes with potential environmental roles, including heavy metal resistance, iron uptake, secondary metabolism, and antibiotic resistance. To our knowledge, these are the first evidence of three monopartite ICEs identified in the single chromosome. Entry exclusion is a mechanism of conjugative MGEs by which prevents the cells to host additional conjugation elements, and this system relies mostly on the entry exclusion gene, such as T4SS-related *tra* genes (*traT* and *traS*) (Garcillán-Barcia and de la Cruz, 2008). We found that all monopartite ICEs in the Acidobacteria chromosome lack recognizable entry exclusion systems and this might explain why Acidobacteria are recipients of additional ICEs.

There have been previous records of the co-existence of various ICEs in the same host. Isolates of the emergent pathogen *Shewanella* spp. have been shown to occur hosting the ICE SXT and ICE*Sh955* for many generations (Parmeciano et al., 2019). The coexistence of multiples ICEs in the chromosome of Acidobacteria might reflect a major advantage for the persistence of phylum in the environment since most of the ICEs carry important genes for bacteria fitness. Altogether, this result sheds light on the ecology of Acidobacteria, which may help to understand the survival, resistance, and persistence of this unknown phylum in the soil environment.

CRediT authorship contribution statement

Osiel Gonçalves: Software, Data curation, Visualization, Investigation, Writing- Original draft preparation. **Mateus Santana:** Conceptualization, Supervision, Writing- Reviewing, and Editing.

Declaration of Competing Interest

The authors declare no conflicts of interest.

Funding

This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; Brasília, Brazil).

References

Aravind, L., Anand, S., Iyer, L.M., 2013. Novel autoproteolytic and DNA-damage sensing components in the bacterial SOS response and oxidized methylcytosine-induced eukaryotic DNA demethylation systems. Biol. Direct 8, 20. https://doi.org/10.1186/1745-6150-8-20

Barns, S.M., Takala, S.L., Kuske, C.R., 1999. Wide Distribution and Diversity of Members of the Bacterial Kingdom Acidobacterium in the Environment. Appl. Environ. Microbiol. 65, 1731 LP – 1737. https://doi.org/10.1128/AEM.65.4.1731-1737.1999

Beker, M., Rose, S., Lykkebo, C.A., Douthwaite, S., 2018. Integrative and Conjugative Elements (ICEs) in Pasteurellaceae Species and Their Detection by Multiplex PCR. Front. Microbiol. 26, 1329. https://doi.org/10.3389/fmicb.2018.01329

Belova, S.E., Ravin, N. V, Pankratov, T.A., Rakitin, A.L., Ivanova, A.A., Beletsky, A. V, Mardanov, A. V, Sinninghe Damsté, J.S., Dedysh, S.N., 2018. Hydrolytic capabilities as a key to environmental success: chitinolytic and cellulolytic Acidobacteria from acidic sub-arctic soils and boreal peatlands. Front. Microbiol. 9, 2775. https://doi.org/10.3389/fmicb.2018.02775

Besemer, J., 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res. 29, 2607–2618. https://doi.org/10.1093/nar/29.12.2607

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

Burrus, V., Pavlovic, G., Decaris, B., Guédon, G., 2002. Conjugative transposons: the tip of the iceberg. Mol. Microbiol. 46, 601–610. https://doi.org/10.1046/j.1365-2958.2002.03191.x

Challacombe, J., Kuske, C., 2012. Mobile genetic elements in the bacterial phylum Acidobacteria. Mob. Genet. Elements 2, 179–183. https://doi.org/10.4161/ mge.21943

Cury, J., Touchon, M., Rocha, E.P.C., 2017. Integrative and conjugative elements and their hosts: composition, distribution and organization. Nucleic Acids Res. 45, 8943–8956. https://doi.org/10.1093/nar/gkx607

Delgado-Baquerizo, M., Oliverio, A.M., Brewer, T.E., Benavent-González, A., Eldridge, D.J., Bardgett, R.D., Maestre, F.T., Singh, B.K., Fierer, N., 2018. A global atlas of the dominant bacteria found in soil. Science 359, 320 LP – 325. https://doi.org/ 10.1126/science.aap9516

Eichorst, S.A., Trojan, D., Roux, S., Herbold, C., Rattei, T., Woebken, D., 2018. Genomic insights into the Acidobacteria reveal strategies for their success in terrestrial environments. Environ. Microbiol. 20, 1041–1063. https://doi.org/10.1111/1462-2920.14043

Farzand, R., Rajakumar, K., Zamudio, R., Oggioni, M.R., Barer, M.R., O'Hare, H.M., 2019. ICEKp2: description of an integrative and conjugative element in *Klebsiella pneumoniae*, co-occurring and interacting with ICE*Kp1*. Sci. Rep. 9, 13892. https://doi.org/10.1038/s41598-019-50456-x

Frost, L.S., Leplae, R., Summers, A.O., Toussaint, A., 2005. Mobile genetic elements: the agents of open source evolution. Nat. Rev. Microbiol. 3, 722.

Garcillán-Barcia, M.P., de la Cruz, F., 2008. Why is entry exclusion an essential feature of conjugative plasmids? Plasmid 60, 1–18. https://doi.org/10.1016/ j.plasmid.2008.03.002

Garcillán-Barcia, M.P., Redondo-Salvo, S., Vielva, L., de la Cruz, F., 2020. MOBscan: Automated Annotation of MOB Relaxases. Methods Mol. Biol. 2075, 295–308. https://doi.org/10.1007/978-1-4939-9877-7_21

Guglielmini, J., Quintais, L., Garcillán-Barcia, M.P., de la Cruz, F., Rocha, E.P.C., 2011. The repertoire of ICE in Prokaryotes underscores the unity, diversity, and ubiquity of conjugation. PLOS Genet. 7, e1002222. https://doi.org/ 10.1371/journal.pgen.1002222

Hallet, B., Arciszewska, L.K., Sherratt, D.J., 1999. Reciprocal control of catalysis by the tyrosine recombinases XerC and XerD: an enzymatic switch in site-specific recombination. Mol. Cell 4, 949–959. https://doi.org/10.1016/s1097-2765(00)80224-5

Haskett, T.L., Ramsay, J.P., Bekuma, A.A., Sullivan, J.T., O'Hara, G.W., Terpolilli, J.J., 2017. Evolutionary persistence of tripartite integrative and conjugative elements. Plasmid 92, 30–36. https://doi.org/10.1016/j.plasmid.2017.06.001

Haskett, T.L., Terpolilli, J.J., Bekuma, A., O'Hara, G.W., Sullivan, J.T., Wang, P., Ronson, C.W., Ramsay, J.P., 2016. Assembly and transfer of tripartite integrative and conjugative genetic elements. Proc. Natl. Acad. Sci. U. S. A. 113, 12268–12273. https://doi.org/10.1073/pnas.1613358113

Hegemann, J.D., Zimmermann, M., Xie, X., Marahiel, M.A., 2015. Lasso Peptides: An intriguing class of bacterial natural products. Acc. Chem. Res. 48, 1909–1919. https://doi.org/10.1021/acs.accounts.5b00156

Jia, B., Raphenya, A.R., Alcock, B., Waglechner, N., Guo, P., Tsang, K.K., Lago, B.A., Dave, B.M., Pereira, S., Sharma, A.N., Doshi, S., Courtot, M., Lo, R., Williams, L.E., Frye, J.G., Elsayegh, T., Sardar, D., Westman, E.L., Pawlowski, A.C., Johnson, T.A., Brinkman,

F.S.L., Wright, G.D., McArthur, A.G., 2016. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res. 45, D566–D573. https://doi.org/10.1093/nar/gkw1004

Johnson, C.M., Grossman, A.D., 2015. Integrative and Conjugative Elements (ICEs): what they do and how they work. Annu. Rev. Genet. 49, 577–601. https://doi.org/10.1146/annurev-genet-112414-055018

Katoh, K., Standley, D.M., 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol. Biol. Evol. 30, 772–780. https://doi.org/10.1093/molbev/mst010

Kielak, A.M., Barreto, C.C., Kowalchuk, G.A., van Veen, J.A., Kuramae, E.E., 2016. The Ecology of Acidobacteria: moving beyond genes and genomes. Front. Microbiol. 7, 744. https://doi.org/10.3389/fmicb.2016.00744

Kishimoto, N., Kosako, Y., Tano, T., 1991. *Acidobacterium capsulatum* gen. nov., sp. nov.: An acidophilic chemoorganotrophic bacterium containing menaquinone from acidic mineral environment. Curr. Microbiol. 22, 1–7. https:// doi.org/10.1007/BF02106205

Koskella, B., Vos, M., 2015. Adaptation in Natural Microbial Populations. Annu. Rev. Ecol. Evol. Syst. 46, 503–522. https://doi.org/10.1146/annurev-ecolsys-112414-054458

Larkin, M.A., Wilm, A., Higgins, D.G., Valentin, F., Blackshields, G., McWilliam, H., Wallace, I.M., Thompson, J.D., Brown, N.P., McGettigan, P.A., Chenna, R., Lopez, R., Gibson, T.J., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948. https://doi.org/10.1093/bioinformatics/btm404

Leplae, R., Geeraerts, D., Hallez, R., Guglielmini, J., Drèze, P., Van Melderen, L., 2011. Diversity of bacterial type II toxin–antitoxin systems: a comprehensive search and functional analysis of novel families. Nucleic Acids Res. 39, 5513–5525. https://doi.org/10.1093/nar/gkr131

Li, X., Xie, Y., Liu, M., Tai, C., Sun, J., Deng, Z., Ou, H.-Y., 2018. oriTfinder: a web-based tool for the identification of origin of transfers in DNA sequences of bacterial mobile genetic elements. Nucleic Acids Res. 46, W229–W234. https://doi.org/10.1093/nar/gky352

Liu, M., Li, X., Xie, Y., Bi, D., Sun, J., Li, J., Tai, C., Deng, Z., Ou, H.-Y., 2018. ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. Nucleic Acids Res. 47, D660–D665. https://doi.org/10.1093/nar/gky1123

Ma, D., Cook, D.N., Alberti, M., Pon, N.G., Nikaido, H., Hearst, J.E., 1993. Molecular cloning and characterization of acrA and acrE genes of *Escherichia coli*. J. Bacteriol. 175, 6299–

6313. https://doi.org/10.1128/jb.175.19.6299-6313.1993

McGinnis, S., Madden, T.L., 2004. BLAST: at the core of a powerful and diverse set of sequence analysis tools BT - Nucleic Acids Res.

Muhire, B., Varsani, A., Martin, D., 2014. SDT: A Virus Classification Tool Based on Pairwise Sequence Alignment and Identity Calculation. PLoS One 9, e108277. https://doi.org/10.1371/journal.pone.0108277

Rawat, S.R., Männistö, M.K., Starovoytov, V., Goodwin, L., Nolan, M., Hauser, L., Land, M., Davenport, K.W., Woyke, T., Häggblom, M.M., 2012. Complete genome sequence of *Terriglobus saanensis* type strain SP1PR4(T), an Acidobacteria from tundra soil. Stand. Genomic Sci. 7, 59–69. https://doi.org/10.4056/sigs.3036810

Rawat, S.R., Männistö, M.K., Starovoytov, V., Goodwin, L., Nolan, M., Hauser, L.J., Land, M., Davenport, K.W., Woyke, T., Häggblom, M.M., 2013. Complete genome sequence of *Granulicella mallensis* type strain MP5ACTX8(T), an acidobacterium from tundra soil. Stand. Genomic Sci. 9, 71–82. https://doi.org/10.4056 /sigs.4328031

Siguier, P., Gourbeyre, E., Chandler, M., 2014. Bacterial insertion sequences: their genomic impact and diversity. FEMS Microbiol. Rev. 38, 865–891.

Wilke, M.S., Lovering, A.L., Strynadka, N.C.J., 2005. Beta-lactam antibiotic resistance: a current structural perspective. Curr. Opin. Microbiol. 8, 525–533. https://doi.org/10.1016/j.mib.2005.08.016

Figures

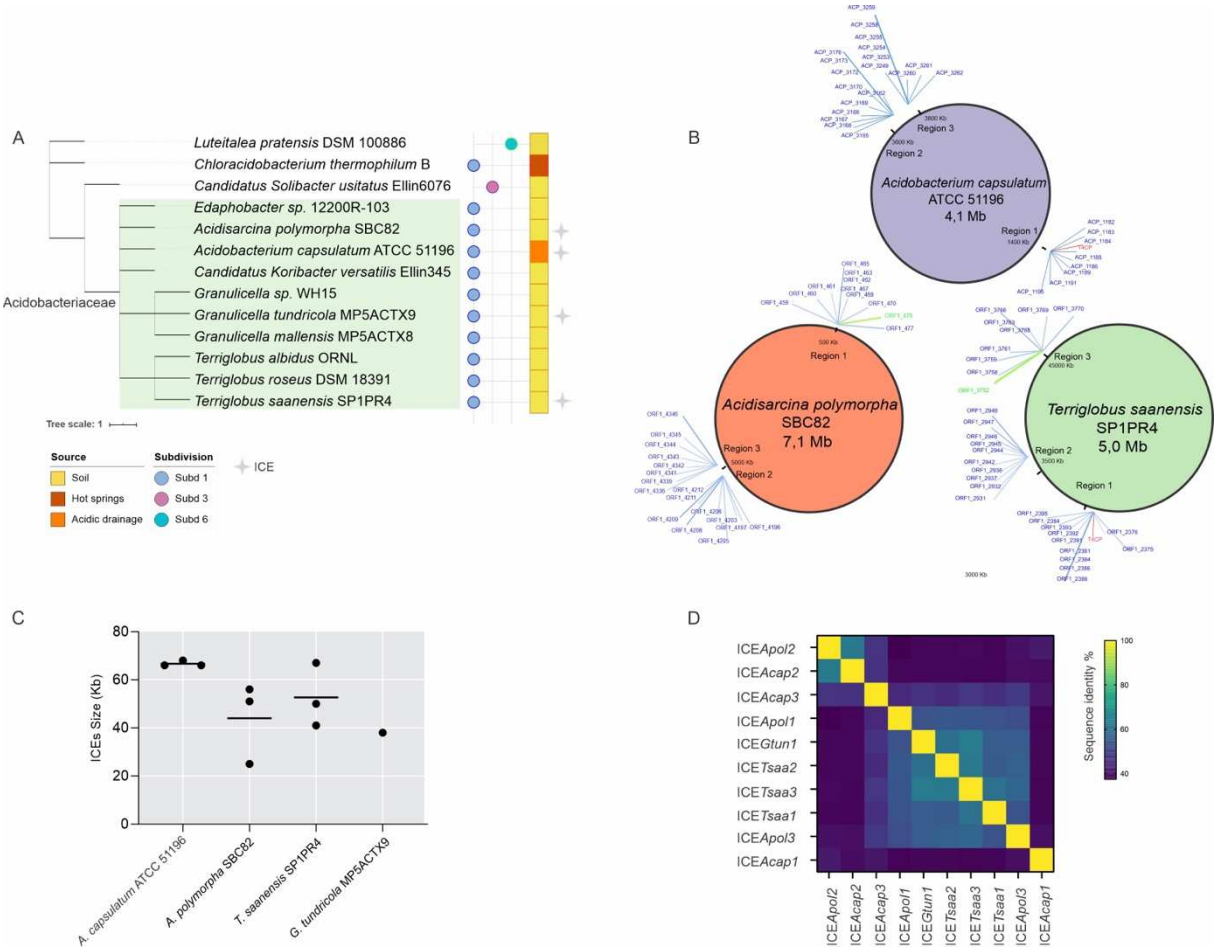


Fig. 1. Overview of monopartite ICEs in chromosomes of Acidobacteria. (A) Phylogenetic tree of Acidobacteria unrooted using the genomes used in this work (Table S1) showing the subdivision for the phylum, the source of each genome, and the presence of ICE. (B) Coordinates of the ORFs related to relaxases (green), type-IV coupling proteins (red), and type-IV secretion system (blue) gene clusters in the genomes of *Acidobacterium capsulatum* ATCC 51196, *Acidisarcina polymorpha* SBC82, and *Terriglobus saanensis* SP1PR4. (C) The sizes of the ICEs that were found in the genome of Acidobacteria. (D) Identity Matrix of nucleotide sequences of whole ICE alignment.

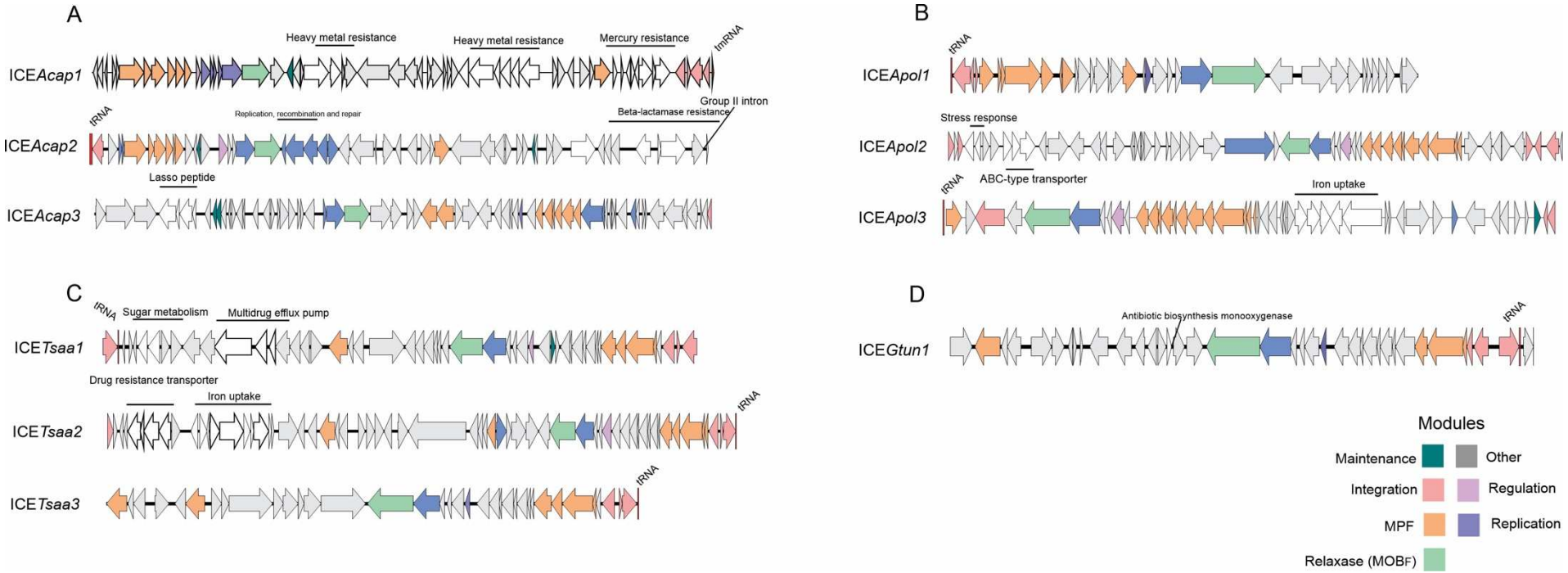


Fig. 2. Genetic content in the different regions of the ICEs. Genes are represented by arrows with different colors according to their functions. Modules are color-coded, and their functional designations are labeled in the figure. MPF, Mating Pair Formation.

Table 1. Characteristics and coordinates of putative monopartite integrative and conjugative element (ICEs) in the genomes of Acidobacteria

Host	Name	Coordinates	Size (kb)	Integration	Integrase	Relaxase	Cargo genes code for
Acidobacterium capsulatum ATCC 51196	ICE*Acap1*	1386780-1453488	66 kb	ssrA tmRNA	XerC	MOB_F	Mercury, Copper resistance, heavy metal efflux
	ICE*Acap2*	3658945-3725767	66 kb	tRNA-Met	XerC	MOB_F	beta-lactamase resistance
	ICE*Acap3*	3727209-3795596	68 kb	*dapA2*	-	MOB_F	Lasso peptides
Acidisarcina polymorpha SBC82	ICE*Apol1*	505917-531056	25 kb	tRNA-Phe	Tyr-phage	MOB_F	Unknown function
	ICE*Apol2*	4848388-4904726	56 kb	*tsaA*	Ser-Rec	MOB_F	Stress response
	ICE*Apol3*	5037772-5086665	51 kb	tRNA-Phe	Rec	MOB_F	ABC-type Fe3+-siderophore transport
Terriglobus saanensis SP1PR4	ICE*Tsaa1*	2870403-2921190	50 kb	tRNA-Asp	XerC	MOB_F	Acriflavin resistance protein, multi-drug efflux system
	ICE*Tsaa2*	3600787-3667585	67 kb	tRNA-Val	XerC	MOB_F	ABC-type Fe3+-siderophore transport, Iron dicitrate transport
	ICE*Tsaa3*	4507722-4549030	41 kb	tRNA-Phe	Tyr-phage	MOB_F	Unknown function
Granulicella tundricola MP5ACTX9	ICE*Gtun1*	3633215-3672162	38 kb	tRNA-Ala	XerC	MOB_F	Antibiotic resistance efflux pump

Appendix 2

GONÇALVES, Osiel Silva; BARREIROS, Ralph Bonandi; TUPY, Sumaya Martins; SANTANA Mateus Ferreira. A reverse-ecology framework to uncover the potential metabolic interplay among '*Candidatus Liberibacter' species*, citrus hosts and psyllid vector. **Gene**, v. 837, p. 146679, 2022. https://doi.org/10.1016/j.gene.2022.146679

A reverse-ecology framework to uncover the potential metabolic interplay among '*Candidatus Liberibacter*' species, *Citrus* hosts and psyllid vector

Osiel Silva Gonçalves[1], Ralph Bonandi Barreiros[2], Sumaya Martins Tupy[1], Mateus Ferreira Santana[1*]

[1]Grupo de Genômica Evolutiva Microbiana, Laboratório de Genética Molecular de Microrganismos, Departamento de Microbiologia, Instituto de Biotecnologia Aplicada à Agropecuária, Universidade Federal de Viçosa, Minas Gerais, Brazil.
[2]Departmento de Fitotecnia, Laboratório de Biotecnologia de Plantas Hortícolas, Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Brazil.

* Corresponding author: mateus.santana@ufv.br.

Abstract

'*Candidatus Liberibacter'* species have developed a dependency on essential nutrients and metabolites from the host cell, as a result of substantial genome reduction. Still, it is difficult to state which nutrients they acquire and whether or not they are metabolically reliant. We used a reverse-ecology model to investigate the potential metabolic interactions of '*Ca Liberibacter'* species, *Citrus*, and the psyllid *Diaphorina citri* in the huanglongbing disease pyramid. Our findings show that hosts (citrus and psyllid) tend to support the nutritional needs of '*Ca. Liberibacter'* species, implying that the pathogen's metabolism has become tightly linked to hosts, which may reflect in the parasite lifestyle of this important genus.

Keywords*:* Greening disease, metabolism, microbial genomics, plant-pathogen interaction

Highlights

- More than ten pathways were found to be involved in the interaction between '*Candidatus Liberibacter'* species, citrus, and psyllids.
- *'Ca. Liberibacter'* spp. and closely related strains differed in the compounds provided by hosts.

- *Citrus* spp. and psyllid tend to support the nutritional requirements of '*Ca. Liberibacter'* spp.

Introduction

The citrus tree is a member of the *Rutaceae* family and is one of the major fruit trees in the world, including important crops such as oranges (*C. sinensis)*, lemons (*C. limon*), grapefruits (*C. paradisi*), pomelos (*C. maxima*), and limes (*C. aurantiifolia*). In addition, several biotic and abiotic issues have challenged the production and quality of citrus fruits across the world, with huanglongbing (HLB), or greening disease, standing out for its ability to have a significant impact on *Citrus* trees. HLB is caused by a group of Gram-negative bacteria '*Candidatus Liberibacter'* spp., which include the '*Candidatus Liberibacter asiaticus'* (CLas), '*Candidatus Liberibacter americanus'* (CLam), '*Candidatus Liberibacter africanus'* (CLaf) (Jagoueix et al., 1994). Together, these species are phloem-colonizing, psyllid-transmitted fastidious bacteria, classified in the *Rhizobiaceae* (CLass: Alphaproteobacteria) (Fagen et al., 2014). '*Ca. Liberibacter'* spp. is transmitted in the hemolymph and salivary glands of psyllids *Diaphorina citri*, and since psyllids feed on sap, this allows bacteria to enter the plant's phloem (Nadarasah and Stavrinides, 2011). They cause significant metabolic and regulatory changes in the plant, causing damage to the transport system, affecting the plant's defensive mechanisms, and altering the chemical and sensory properties of the fruit. Blotchy mottle leaves, stunted growth, reduced fruit size, premature fruit drop, corky veins, and root decline are also observed (Baldwin et al., 2010; Bové and Barros, 2006; Dala Paula et al., 2018; Wang et al., 2017).

'Ca. Liberibacter' spp. are obligate host-associated bacteria with specialized living environments in the host plant (i.e., phloem) or vector (specific psyllid cells, tissues, and organs). As a result, their genomes undergo extensive gene and DNA loss, resulting in a small genome (approximately 1.2 Mb) with few genes and correspondingly restricted metabolic capacities that rely on the regular supply of certain energy substrates from hosts (Moran, 2002; Thapa et al., 2020). Although several studies have addressed the interaction of '*Ca. Liberibacter'* spp., citrus, and psyllids (Mafra et al., 2013; Ramsey et al., 2020; de Francesco et al., 2022), it is still unclear which nutrients they acquire and whether there is metabolic dependency. Therefore, we employed a reverse ecology analysis to obtain insights into the potential metabolic interplay of '*Ca. Liberibacter'* spp., *C. sinensis* and the psyllid *D. citri*. This approach is based on several computational tools to translate high-throughput genetic data into

large-scale ecological data, which potentially turns ecology into a high-throughput field (Levy and Borenstein, 2012).

Reverse ecology approaches have been used in different fields, including microbial-plant interactions to uncover metabolic environments (Karpinets et al., 2014; Ofaim et al., 2017); in human health, to quantify microbes' biosynthetic capabilities across the human oral microbiome (Bernstein et al., 2019) and to identify immune-beneficial infant gut bacteria by mining their metabolism for prebiotic feeds (Michelini et al., 2018). Here, we investigated the potential metabolic interplay among '*Ca. Liberibacter'* spp., *C. sinensis,* and *D. citri* using reverse ecology analysis.

Material and methods

Public metabolic data were retrieved from the KEGG (Kyoto Encyclopedia of Genes and Genomes) database using the function *getOrgMetabolicData* in the Package RevEcoR (Cao et al., 2016). The following '*Ca. Liberibacter'* spp. was employed in this study: CLas, CLam, CLaf, and '*Ca Liberibacter solanacearum'* CLso-ZC1 (CLso-ZC1 - *lso*). In addition, we included *Liberibacter crescens* BT-1 (*lcc*), a culture representative surrogate for plant pathogen '*Ca. Liberibacter'* (Jain et al. 2019), and *Sinorhizobium meliloti* 1021 (*sme*), a closely related model bacterium to *Liberibacter*. We also retrieved metabolic data from the hosts *C. sinensis* (*cit*) and *D. citri* (*dci*). Each KEGG Organism code is displayed in the front of the species name in parenthesis.

Next, a matrix was created including the substrate and product for each species, and then NetCooperate was applied for inferring host-microbe cooperation (cooperative and competitive potential) (Levy et al., 2015). The biosynthetic support score (BSS), and metabolic complementarity index (MCI) were measured for species interactions. The compounds were then annotated in KEGG compounds with biological roles and the Phytochemical compounds Database (https://www.genome.jp/kegg/compound/). In addition, NetMet (https://freilich-lab-tools.com/) (Tal et al., 2020) was used for predicting the metabolic performances of microbes and their corresponding combinations in user-defined environments. For that, we took the lists of species-specific enzymatic reactions (EC numbers) in the Joint Genome Institute (JGI, https://genome.jgi. doe.gov/portal/) for the '*Ca. Liberibacter'* spp. and used the *C. sinensis* and *D. citri* compounds from RevEcoR output as environmental input (nutritional resources).

Results and Discussion

We first calculate the Biosynthetic Support Score (BSS), which evaluates a host species' capacity to provide the nutritional requirements of a parasitic or commensal species (Levy et al., 2015). According to network analysis (Fig. 1A), the majority of '*Ca. Liberibacter'* spp. metabolic pathways were supported by metabolic pathways of *C. sinensis* and *D. citri*, as shown by the support seeds (red dots), indicating that the bacteria obtain compounds exogenously from the hosts. These seed sets were proven to be compatible with the lifestyles of many bacteria and typically correspond with various basic aspects of the species' surroundings, as well as biological observations of key adaptations (Borenstein et al., 2008). Additionally, the distribution of BSS values of '*Ca. Liberibacter'* spp. against *C. sinensis* and *D. citri* ranged from approximately 0.1 to 0.8 (Fig. 1B), and there was a significant difference in BSS scores between bacteria and hosts (Fig. 1B). Our findings suggest that CLam and CLas are more dependent on *D. citri* than CLaf and CLso-ZC1, but the same evidence was not observed for *C. sinensis* (Fig. 1B). In addition, we revealed that the '*Ca. Liberibacter'* spp. and closely related species *L. crescens* and *S. meliloti* have a similar score, despite the fact that *S. meliloti* has low score support from the host (Fig. 1B).

Furthermore, we seek to provide which pathways and compounds are involved in these interactions in both directions. In the '*Ca. Liberibacter'* spp. – *C. sinensis* interaction, we mapped a total of 1702 compounds grouped into 17 pathways, while in the '*Ca. Liberibacter'* spp. – *D. citri* interaction, we mapped 1141 compounds grouped into 14 pathways (Fig. S1, Supplementary Table S1). These pathways were involved mostly in metabolic pathways, biosynthesis of secondary metabolites, biosynthesis of cofactors, and biosynthesis of amino acids. We observe that the hosts may supply some compounds involved in specific pathways for '*Ca. Liberibacter'* spp., *S. meliloti,* and *L. crescens*, such as compounds in D-amino acid metabolism in the *C. sinensis*-CLam interaction; and compounds in fatty acid and pyrimidine metabolism in the *C. sinensis*-CLas/ *D. citri*-CLas interaction (Fig. S2, Fig. S3).

In general, most of these compounds were involved in metabolic pathways, including ATP, UDP-glucose, L-Alanine, 5-Phospho-alpha-D-ribose 1-diphosphate, and Folinic acid, following biosynthesis of secondary metabolites. Details of the compounds that were found in the hosts-pathogen interaction can be found in Supplementary Table S2. Other key pathways for bacterial survival have been found, including amino acid biosynthesis, carbon metabolism, and fatty acid metabolism. It has been proposed that very-long-chain fatty acids play an important role in the axenic growth of pathogenic *Liberibacter* spp., but this gene is missing in all '*Ca. Liberibacter'* spp. (Leonard et al., 2012). Remarkably, we only mapped this compound

in the interaction among *S. meliloti*, a closely related model bacterium, *C. sinensis* and *D. citri* (Supplementary Table S2). Besides, CLas lacks the ability to synthesize phosphatidylcholine, which is associated with the fluidity, permeability, and potential of bacterial membranes, but encodes a predicted ABC transporter system for choline, indicating that it is capable of utilizing extracellular choline (Li et al., 2012). We found evidence that *D. citri* may support this compound for CLas (Supplementary Table S2).

The compounds involved in the BSS were then contracted against each '*Ca. Liberibacter'* spp. We found that *L. crescens* and CLaf had the most exclusive compounds mapped, 33 and 31, respectively (Fig. 1C). We mapped *L. crescens*, compounds with biological roles in the classes of organic acids (2-oxoisovalerate), carbohydrates (d-mannose), peptides (l-ornithine), and vitamins and cofactors (phylloquinone, menaquinone, and nicotinamide). In addition, for CLaf, we found organic acids (malate), nucleic acids (dAMP), vitamins and cofactors (pyridoxamine, UDP-glucose, biotin) (Supplementary Table S3).

Finally, we used NetMet (Tal et al., 2020), a tool for predicting the metabolic performances of microorganisms and their environments, which we defined as a compilation of nutrients provided by the hosts citrus and psyllid (Fig. 2). Our analysis shows that the hosts may provide practically all needed cellular components, however, some categories are host-dependent and/or produced individually by the strain, such as a few essential amino acids, cofactors and nucleotides (Fig. 2).

Throughout the longitudinal study, our *in-silico* approach was also congruent with experimental data of metabolite extracts obtained from leaf samples taken from CLas and control grafted lemon plants (Ramsey et al., 2020). The authors profiled 25 metabolites using 1H-NMR spectroscopy, including sugars, amino acids, and other primary and secondary metabolites. Interestingly, adenosine, along with many amino acids (proline, arginine, and the branched-chain amino acids (BCAA), was among the metabolites with the significant concentration changes between fruit taken from $CLas^{+}$ trees and those collected from CLas trees (Slisz et al. 2012). Similar metabolites were also reported by other studies using different *Citrus* varieties (Chin et al. 2014; Liu et al. 2020).

Fan et al., (2012) observed different transcriptional changes in host responses to CLas in highly susceptible *C. sinensis* and tolerant rough lemon (*C. jambhiri* Lush.) throughout the time course, including lipid metabolism and hormone metabolisms, using microarray analysis.

Furthermore, the authors demonstrated that pathways such as glucose metabolism, cell wall metabolism, and stress response were completely changed in disease development. Overall, these results indicate that several metabolites are available during the infection of '*Ca. Liberibacter'* spp. in *Citrus* trees may be used by the pathogen to thrive in the plant environment.

Here, we demonstrated that the main common compounds abundantly available in the phloem sap and hemolymph environments, such as organic acids, and amino acids, vitamins, saccharides, and fatty acids (Duan et al., 2009; Killiny, 2017; Killiny et al., 2017; Merfa et al., 2019), including the metabolism of purines and pyrimidines, which CLas lacks essential enzymes (Hartung et al., 2011), may be provided by the hosts for '*Ca. Liberibacter'* spp. In addition, several studies have shown that '*Ca. Liberibacter'* spp. obtain essential nutrients from microbial communities other than citrus and psyllid hosts ((Zuñiga et al., 2020; Hu et al., 2021).

Taken together, our findings show that hosts (citrus and psyllid) tend to support the nutritional requirements of '*Ca. Liberibacter'* spp. This suggests that the pathogen's metabolism has become tightly linked to hosts, which may reflect in the parasite lifestyle and the complexity to isolate in axenic culture '*Ca. Liberibacter'* spp. in an artificial medium. Understanding the interaction between the host and pathogen could aid in the development of new disease-control strategies, as well as the design of sustainable media culture that supports the growth of the bacteria.

References

Baldwin, E., Plotto, A., Manthey, J., McCollum, G., Bai, J., Irey, M., Cameron, R., Luzio, G., 2010. Effect of *Liberibacter Infection* (Huanglongbing Disease) of *Citrus* on Orange Fruit Physiology and Fruit/Fruit Juice Quality: Chemical and Physical Analyses. Journal of Agricultural and Food Chemistry 58, 1247–1262. https://doi.org/10.1021/jf9031958

Bernstein, D.B., Dewhirst, F.E., Segrè, D., 2019. Metabolic network percolation quantifies biosynthetic capabilities across the human oral microbiome. Elife 8, 1–33. https://doi.org/10.7554/eLife.39733.001

Borenstein, E., Kupiec, M., Feldman, M.W., Ruppin, E., 2008. Large-scale reconstruction and phylogenetic analysis of metabolic environments. Proceedings of the National Academy of Sciences 105, 14482 LP – 14487. https://doi.org/10.1073/pnas.0806162105

Bové, J., Barros, A., 2006. Huanglongbing: A destructive, newly emerging, century-old disease of citrus. J Plant Pathol 88. https://doi.org/10.4454/jpp.v88i1.828

Cao, Y., Wang, Y., Zheng, X., Li, F., Bo, X., 2016. RevEcoR: an R package for the reverse ecology analysis of microbiomes. BMC Bioinformatics 17, 294. https://doi.org/10.1186/s12859-016-1088-4

Dala Paula, B.M., Raithore, S., Manthey, J.A., Baldwin, E.A., Bai, J., Zhao, W., Glória, M.B.A., Plotto, A., 2018. Active taste compounds in juice from oranges symptomatic for Huanglongbing (HLB) citrus greening disease. LWT 91, 518–525. /https://doi.org/10.1016/j.lwt.2018.01.083

de Francesco, A., Lovelace, A.H., Shaw, D., Qiu, M., Wang, Y., Gurung, F., Ancona, V., Wang, C., Levy, A., Jiang, T., Ma, W., 2022. Transcriptome Profiling of '*Candidatus Liberibacter asiaticus*' in *Citrus* and Psyllids. Phytopathology® 112, 116–130. https://doi.org/10.1094/PHYTO-08-21-0327-FI

Duan, Y., Zhou, L., Hall, D.G., Li, W., Doddapaneni, H., Lin, H., Liu, L., Vahling, C.M., Gabriel, D.W., Williams, K.P., Dickerman, A., Sun, Y., Gottwald, T., 2009. Complete genome sequence of citrus huanglongbing bacterium, '*Candidatus Liberibacter asiaticus*' obtained through metagenomics. Molecular Plant-Microbe Interactions® 22, 1011–1020. https://doi.org/10.1094/MPMI-22-8-1011

Fagen, J.R., Leonard, M.T., Coyle, J.F., McCullough, C.M., Davis-Richardson, A.G., Davis, M.J., Triplett, E.W., 2014. *Liberibacter crescens* gen. nov., sp. nov., the first cultured member of the genus Liberibacter. International Journal of Systematic and Evolutionary Microbiology 64, 2461–2466. https://doi.org/https://doi.org/1 0.1099/ijs.0.063255-0

Fan, J., Chen, C., Yu, Q., Khalaf, A., Achor, D.S., Brlansky, R.H., Moore, G.A., Li, Z.-G., Gmitter, F.G., 2012. Comparative transcriptional and anatomical analyses of tolerant rough lemon and susceptible sweet orange in response to '*Candidatus Liberibacter asiaticus'* infection. Molecular Plant-Microbe Interactions® 25, 1396–1407. https://doi.org/10.1094/MPMI-06-12-0150-R

Hartung, J.S., Shao, J., Kuykendall, L.D., 2011. Comparison of the '*Ca. Liberibacter asiaticus*' genome adapted for an intracellular lifestyle with other members of the Rhizobiales. PLOS ONE 6, e23289-.

Hu, B., Rao, M.J., Deng, X., Pandey, S.S., Hendrich, C., Ding, F., Wang, N., Xu, Q., 2021. Molecular signatures between citrus and *Candidatus Liberibacter asiaticus*. PLOS Pathogens 17, e1010071-.

Jagoueix, S., Bove, J.-M., Garnier, M., 1994. The phloem-limited bacterium of greening disease of citrus is a member of the α subdivision of the Proteobacteria. International Journal of Systematic and Evolutionary Microbiology 44, 379–386. https://doi.org/https://doi.org/10.1099/00207713-44-3-379

Karpinets, T. v, Park, B.H., Syed, M.H., Klotz, M.G., Uberbacher, E.C., 2014. Metabolic environments and genomic features associated with pathogenic and mutualistic interactions between bacteria and plants. Molecular Plant-Microbe Interactions® 27, 664–677. https://doi.org/10.1094/MPMI-12-13-0368-R

Killiny, N., 2017. Metabolite signature of the phloem sap of fourteen citrus varieties with different degrees of tolerance to *Candidatus Liberibacter asiaticus*. Physiological and Molecular Plant Pathology 97, 20–29. https://doi.org/10.1016/j.pmpp. 2016.11.004

Killiny, N., Hijaz, F., El-Shesheny, I., Alfaress, S., Jones, S.E., Rogers, M.E., 2017. Metabolomic analyses of the haemolymph of the Asian citrus psyllid *Diaphorina citri*, the vector of huanglongbing. Physiological Entomology 42, 134–145. https://doi.org/https://doi.org/10.1111/phen.12183

Leonard, M.T., Fagen, J.R., Davis-Richardson, A.G., Davis, M.J., Triplett, E.W., 2012. Complete genome sequence of Liberibacter crescens BT-1. Standards in Genomic Sciences 7, 271–283. https://doi.org/10.4056/sigs.3326772

Levy, R., Borenstein, E., 2012. Reverse Ecology: from systems to environments and back. Adv Exp Med Biol 751, 329–345. https://doi.org/10.1007/978-1-4614-3567-9_15

Levy, R., Carr, R., Kreimer, A., Freilich, S., Borenstein, E., 2015. NetCooperate: a network-based tool for inferring host-microbe and microbe-microbe cooperation. BMC Bioinformatics 16, 164. https://doi.org/10.1186/s12859-015-0588-y

Li, W., Cong, Q., Pei, J., Kinch, L.N., Grishin, N. v, 2012. The ABC transporters in *Candidatus Liberibacter asiaticus*. Proteins: Structure, Function, and Bioinformatics 80, 2614–2628. https://doi.org/https://doi.org/10.1002/prot.24147

Mafra, V., Martins, P.K., Francisco, C.S., Ribeiro-Alves, M., Freitas-Astúa, J., Machado, M.A., 2013. *Candidatus Liberibacter americanus* induces significant reprogramming of the transcriptome of the susceptible citrus genotype. BMC Genomics 14, 247. https://doi.org/10.1186/1471-2164-14-247

Merfa, M. v, Pérez-López, E., Naranjo, E., Jain, M., Gabriel, D.W., de La Fuente, L., 2019. Progress and Obstacles in Culturing '*Candidatus Liberibacter asiaticus*', the Bacterium Associated with Huanglongbing. Phytopathology® 109, 1092–1101. https://doi.org/10.1094/PHYTO-02-19-0051-RVW

Michelini, S., Balakrishnan, B., Parolo, S., Matone, A., Mullaney, J.A., Young, W., Gasser, O., Wall, C., Priami, C., Lombardo, R., Kussmann, M., 2018. A reverse metabolic approach to weaning: *In silico* identification of immune-beneficial infant gut bacteria, mining their metabolism for prebiotic feeds and sourcing these feeds in the natural product space. Microbiome 6, 1–18. https://doi.org/10.1186/s40168-018-0545-x

Moran, N.A., 2002. Microbial Minimalism: Genome Reduction in Bacterial Pathogens. Cell 108, 583–586. https://doi.org/10.1016/S0092-8674(02)00665-7

Nadarasah, G., Stavrinides, J., 2011. Insects as alternative hosts for phytopathogenic bacteria. FEMS Microbiology Reviews 35, 555–575. https://doi.org/10.1111/j.1574-6976.2011.00264.x

Ofaim, S., Ofek-Lalzar, M., Sela, N., Jinag, J., Kashi, Y., Minz, D., Freilich, S., 2017. Analysis of microbial functions in the rhizosphere using a metabolic-network based framework for metagenomics interpretation. Frontiers in Microbiology.

Ramsey, J.S., Chin, E.L., Chavez, J.D., Saha, S., Mischuk, D., Mahoney, J., Mohr, J., Robison, F.M., Mitrovic, E., Xu, Y., Strickler, S.R., Fernandez, N., Zhong, X., Polek, M., Godfrey, K.E., Giovannoni, J.J., Mueller, L.A., Slupsky, C.M., Bruce, J.E., Heck, M., 2020. Longitudinal transcriptomic, proteomic, and metabolomic analysis of citrus limon response to graft inoculation by *Candidatus Liberibacter asiaticus*. J Proteome Res 19, 2247–2263. https://doi.org/10.1021/acs.jproteome. 9b00802

Tal, O., Selvaraj, G., Medina, S., Ofaim, S., Freilich, S., 2020. NetMet: A Network-Based Tool for Predicting Metabolic Capacities of Microbial Species and their Interactions. Microorganisms. https://doi.org/10.3390/microorganisms8060840

Thapa, S.P., de Francesco, A., Trinh, J., Gurung, F.B., Pang, Z., Vidalakis, G., Wang, N., Ancona, V., Ma, W., Coaker, G., 2020. Genome-wide analyses of Liberibacter species provides insights into evolution, phylogenetic relationships, and virulence factors. Mol Plant Pathol 21, 716–731. https://doi.org/10.1111/mpp.12925

Wang, N., Pierson, E.A., Setubal, J.C., Xu, J., Levy, J.G., Zhang, Y., Li, J., Rangel, L.T., Martins, J., 2017. The *Candidatus Liberibacter*–Host Interface: Insights into pathogenesis mechanisms and disease control. Annual Review of Phytopathology 55, 451–482. https://doi.org/10.1146/annurev-phyto-080516-035513

Zuñiga, C., Peacock, B., Liang, B., McCollum, G., Irigoyen, S.C., Tec-Campos, D., Marotz, C., Weng, N.-C., Zepeda, A., Vidalakis, G., Mandadi, K.K., Borneman, J., Zengler, K., 2020. Linking metabolic phenotypes to pathogenic traits among "*Candidatus Liberibacter asiaticus*" and its hosts. npj Systems Biology and Applications 6, 24. https://doi.org/10.1038/s41540-020-00142-w

Funding

This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq (Process Nos. 143132/2019-9 and 147654/2021-1), Coordenacão de Aperfeiçoamento de Pessoal de Nível Superior/Programa de Excelência Acadêmica-Finance Code 001.

CRediT authorship contribution statement

Osiel Silva Gonçalves: Conceptualization, Software, Data curation, Visualization, Investigation, Writing - original draft. **Ralph Bonandi Barreiros**: Conceptualization, Investigation, Writing - original draft. **Sumaya Martins Tupy**: Software, Data curation. **Mateus Ferreira Santana**: Conceptualization, Supervision, Writing - review & editing.

Declaration of Competing Interest

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

Figures

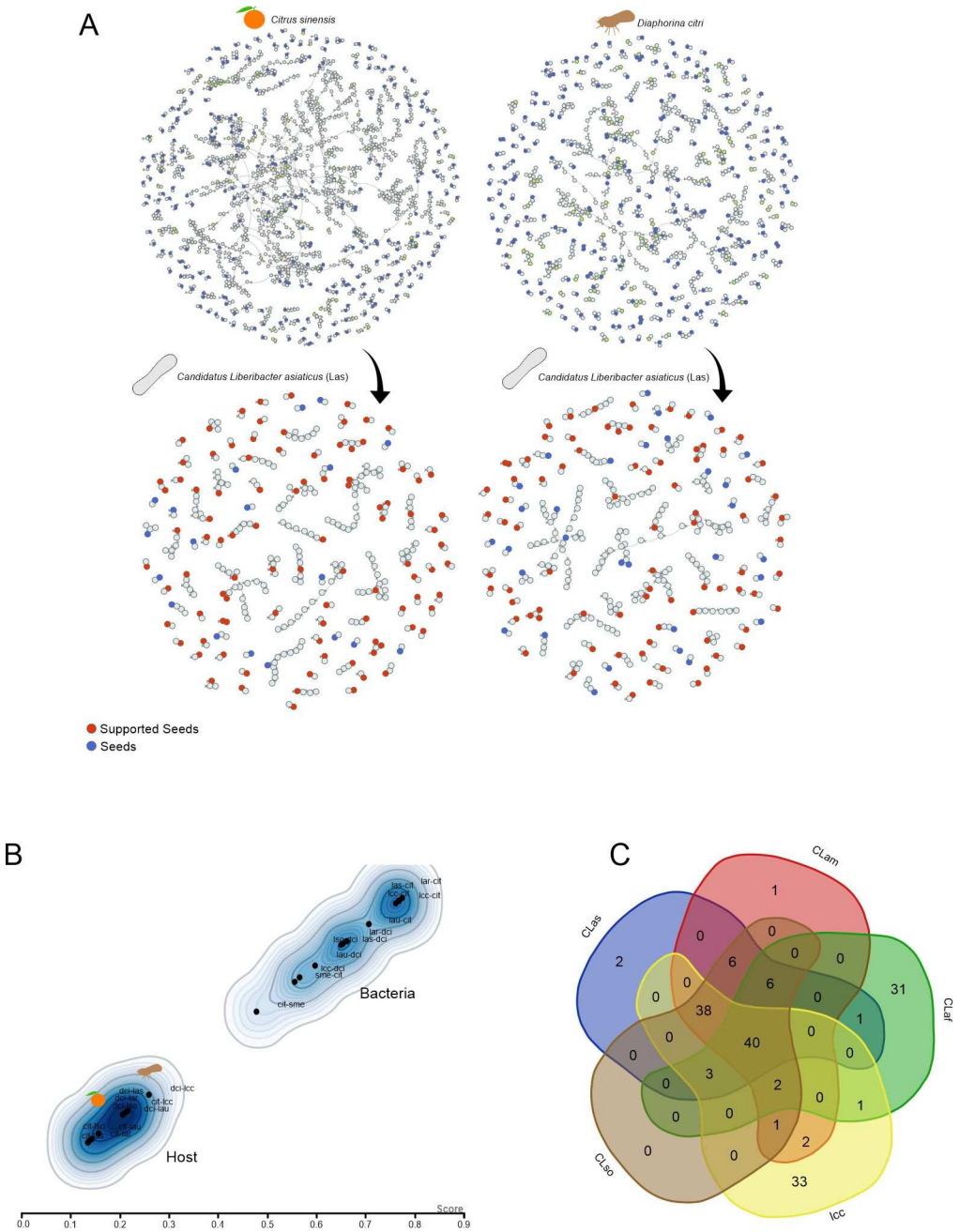


Fig 1. The potential interplay among '*Ca Liberibacter'* species, *C. sinensis*, and the psyllid *D. citri* depicted through reverse ecology. **A**, The metabolic networks of '*Ca' L. asiaticus* (CLas), *Citrus*, and *D. citri.* Nodes represent compounds and edges connecting substrates to products. Each network highlights which metabolites were identified as seeds (blue), which of these seeds were identified as being potentially supported by the second species (red), and the corresponding supporting metabolites in the second network (green). Only the giant components are considered here. **B**, the distribution of the biosynthetic support score. **C**, Diagram Ven showing shared compounds among the '*Ca' Liberibacter* species. Abbreviations: *Candidatus Liberibacter asiaticus* (CLas); *Candidatus Liberibacter americanus* (CLam); *Candidatus Liberibacter africanus* (CLaf); *Candidatus Liberibacter solanacearum* CLso-ZC1 (CLso); *Liberibacter crescens* BT-1 (lcc); *Sinorhizobium meliloti* 1021 (sme).

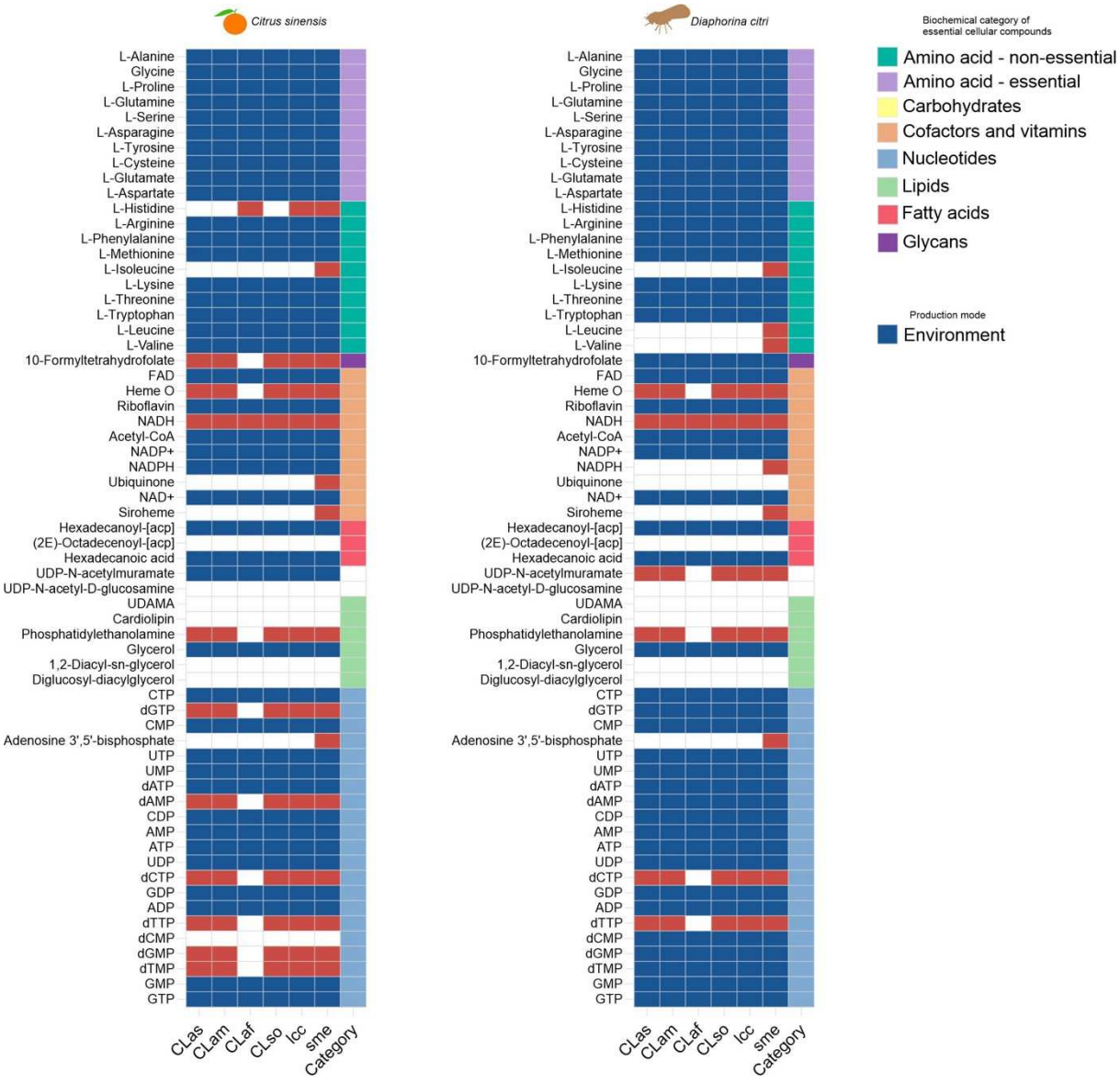


Fig 2. Profiles of the predicted production of cellular building blocks by the *C. sinensis* and *D. citri.* X-axis: *Candidatus Liberibacter asiaticus* (CLas); *Candidatus Liberibacter americanus* (CLam); *Candidatus Liberibacter africanus* (CLaf); *Candidatus Liberibacter solanacearum* CLso-ZC1 (CLso); *Liberibacter crescens* BT-1 (lcc); *Sinorhizobium meliloti* 1021 (sme

Appendix 3

GONÇALVES, Osiel Silva et al. Harnessing Novel Soil Bacteria for Beneficial Interactions with Soybean. **Microorganisms**, v. 11, n. 2, p. 300, 2023. https://doi.org/10.3390/microorganisms11020300

microorganisms

MDPI

139

Article

Harnessing novel soil bacteria for beneficial interactions with soybean

Osiel Silva Gonçalves [1,2,] *, Thamires Santos Souza[2], Guilherme de Castro Gonçalves[2], Alexia Suellen Fernandes[2], Tomás Gomes Reis Veloso[2], Sumaya Martins Tupy[2], Ediones Amaro Garcia[3], and Mateus

1 Institute for Global Food Security, School of Biological Sciences, Queen's University Belfast, Belfast, UK
2 Grupo de Genômica Evolutiva Microbiana, Laboratório de Genética Molecular de Microrganismos, Departamento de Microbiologia, Instituto de Biotecnologia Aplicada à Agropecuária, Universidade Federal de Viçosa, Minas Gerais, Brazil
3 Laboratório de Genética Molecular de Microrganismos, Departamento de Microbiologia, Instituto de Biotecnologia Aplicada à Agropecuária, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil
* Correspondence: o.goncalves@ads.qub.ac.uk (O.S.G.); mateus.santana@ufv.br (M.F.S.)

Abstract: It is claimed that one g of soil holds ten billion bacteria representing thousands of distinct species. These bacteria play key roles in the regulation of terrestrial carbon dynamics, nutrient cycles, and plant productivity. Despite the overwhelming diversity of bacteria, most bacterial species remain largely unknown. Here, we used an oligotrophic medium to isolate novel soil bacteria for positive interaction with soybean. Strictly 22 species of bacteria from the soybean rhizosphere were selected. These isolates encompass ten genera (*Kosakonia*, *Microbacterium*, *Mycobacterium*, *Methylobacterium*, *Monashia*, *Novosphingobium*, *Pandoraea*, *Anthrobacter*, *Stenotrophomonas*, and *Rhizobium*) and have the potential as novel species. Furthermore, the novel bacterial species exhibited plant growth-promoting traits *in vitro* and enhanced soybean growth under drought stress in a greenhouse experiment. We also reported the draft genome sequences of *Kosakonia* sp. strain SOY2 and *Agrobacterium* sp. strain SOY23. Along with our analysis of 169 publicly available genomes for the genera reported here, we demonstrated that these bacteria have a repertoire of genes encoding plant growth-promoting proteins and secondary metabolite biosynthetic gene clusters that directly affect plant growth. Taken together, our findings allow the identification novel soil bacteria, paving the way for their application in crop production.

Keywords: Biocontrol; Drought; Genomics; Soybean.

1. Introduction

Plants are intimately intertwined with microbial communities in which several distinct mechanisms mediate dynamic ecological interactions. Plants release photosynthates belowground through mucilage and exudates, which are used as energy sources by distinct dwelling microbial taxa [1–3]. In return, some specific microbial taxa can promote plant growth and/or offer protection against biotic and abiotic stressors. They occur via the synthesis of phytohormones, acquisition of nutrients, and antagonistic interactions with plant pathogens.

It is estimated that less than 1% of bacterial species have been cultivated under laboratory conditions, a phenomenon known as the "Great Plate Count Anomaly" [4]. Soils are by far the richest environment, containing an extensive and diverse set of bacteria, of which the majority are yet unknown, and are mainly detected by metagenome analysis [5,6]. Hence, the ecological features of most soil bacterial taxa, including their environmental preferences, phenotypes, and metabolic capacities are mostly unknown.

Academic Editor: Firstname Lastname

Received:
Revised: date
Accepted: date
Published: date

Apart from the rhizosphere, most of the soil is considered an oligotrophic environment. This area is distinguished by lower levels of microbial density and activity than those in high-resource environments [7]. The microbiome that inhabits this habitat is classified as a k-strategist, which implies that it can survive in low-nutrient conditions, grow at a slower rate, and has a high tolerance to toxic compounds [8,9]. Yet, most of the cultivation methods used in microbiology rely on nutrient-rich media, which may limit the study of oligotrophic bacteria from soil ecosystems [10]. These bacteria play key roles in regulating terrestrial carbon dynamics, nutrient cycles, and plant productivity. Therefore, novel strategies for assessing this unknown biological diversity are necessary.

Studies have shown the potential for cultivation of previously 'unculturable' bacteria from environmental samples using simple cultivation strategies [11,12]. The cultivation of 'unculturable' bacteria can be improved by combining oligotrophic media, extended incubation periods, and selection of slow-growing bacteria [12,13]. Herein, the goal of this study was to isolate and identify novel microbial species with the potential to plant growth-promoting rhizobacteria (PGPR) for soybean plants, as well as to apply genomics approaches to gain insights into bacteria-plant interactions.

2. Materials and Methods

2.1. Soil sampling, preparation of media, and isolation procedures

Five equidistant samples of rhizosphere soil from soybean (*Glycine max*) were obtained from the experimental field at the Universidade Federal de Viçosa, Minas Gerais, Brazil (20°46'01.1"S 42°52'10.2"W). The map was built using QGIS 3.16 Hannover (DATUM: SIRGAS2000, UTM zone 22S) (Figure 1A). 10g of each sample was diluted in series with sterile saline. Aliquots of each dilution were inoculated in the following media: VL55 medium (3.0 g 2-[N-morpholino]ethanesulfonic acid (MES), 0.4 mM $MgSO_4$, 0.6 mM $CaCl_2$, 0.4 mM $(NH_4)_2HPO_4$, 2 ml of tungstate/selenite solution (composition in 1 L of distilled water: 0.5 g of NaOH, 3 mg of $Na_2SeO_3.5H_2O$, 4 mg of $Na_2WO_4.2H_2O$ [14] and 2 ml of SL-10 trace elements (composition in 1 L of distilled water: 10 ml of HCl [25%, vol/ vol], 1.5 g of $FeCl_2.4H_2O$, 70 mg of $ZnCl_2$, 100 mg of $MnCl_2.4H_2O$, 6 mg of H_3BO_3, 190 mg of $CoCl_2.6H_2O$, 2 mg of $CuCl_2.2H_2O$, 24 mg of $NiCl_2.6H_2O$ and 36 mg of $Na_2MoO_4.2H_2O$. The pH was adjusted to 5.5 with a mixture of 200 mM NaOH and 100 mM KOH. This base medium was autoclaved at 121 °C for 20 min and cooled to 56 °C. A 10 mL aliquot of 5% (w/v) xylan from beechwood (Fluka), 2 mL of vitamin solution (see below), and 2 mL of an SL-10 trace element solution [15] were added per 1 L. Selenite/tungstate, vitamin, and trace element solutions was sterilized by filtration. The vitamin solution contained (per 1 L of distilled water) 2 mg (+)-biotin, 2 mg of folic acid, 10 mg of pyridoxamine hydrochloride, 5 mg of thiamine chloride, 5 mg of riboflavin, 5 mg of nicotinic acid, 5 mg of hemicalcium D-((+)-pantothenate, 0.1 mg of cyanocobalamin, 5 mg of 4-aminobenzoate and 5 mg of DL-6,8-thiotic acid. Nystatin (10 mg L^{-1}) was added. Plates with five replicates per sample were incubated in polyethylene bags to prevent desiccation at 25 °C in the dark for two weeks. The colonies were collected and transferred to fresh *DMSZ acidiphilium* medium agar (composition in 1 L of distilled water:2 g $(NH_4)_2SO_4$, 0.1 g KCl, 0.5 g K_2HPO_4, 0.5 g $MgSO_4$ x $7H_2O$, 0.3 g yeast extract, 1 g D-glucose, 15 g agar-agar) plates for purification. The cells were grown in liquid DMSZ *acidiphilium* medium for cryostock preparation with stocks prepared with glycerol and maintained at -80 °C.

2.2. DNA fingerprinting

For molecular characterization of bacterial isolates, molecular markers BOX, ERIC, and GTG. PCR was performed for fingerprint analysis. BOX GTG regions were amplified by PCR using primers BOXA1R (5'-CTACGGCAAGGCGACGCTGACG-3') [26,27], and ERIC1R (5'-ATGTAAGCTCCTGGGGATTCAC-3'), ERIC2 (3'-AAGTAAGTGACTGGGG TGAGCG-5') [27], and GTG5 (GTGGTGGTGGTGGTG). PCR reactions were adjusted to a volume of 25 µL containing 5.0 uL of Buffer, 2.5 uL of $MgCl_2$, 2.5 µL of primer (BOX and GTG), 2.0 uL of genomic DNA 1.0 uL of dNTP, 0.25 ul of Taq DNA polymerase and 11.75 ul of H_2O. The regions were amplified with an initial

denaturation step at 95°C for 5 min, followed by 30 cycles at 94°C for 1 min, 51°C for 1 min, and 65°C for 8 min, and a final extension at 65°C for 8 min. The amplification products were analyzed by electrophoresis on 1.5% agarose gel, and the fingerprint patterns were analyzed using BioNumerics (Applied Maths) software, and the same pattern fingerprint was considered a clone.

2.3. DNA Extraction and 16S rRNA sequencing, processing, and analysis

The isolates were inoculated into tubes with 20 ml of liquid DSMZ medium for one week at 28 °C and shaken on an orbital shaker at 180 rpm. Cells were collected by centrifugation and Genomic DNA was extracted using the Wizard® Genomic DNA Purification Kit (Promega Corp.), as recommended by the manufacturer. DNA quality was checked using a NanoDrop 2000 (Thermo Scientific) and subjected to gel electrophoresis (0.8% of agarose). The 16S rRNA genes were amplified using the 27F and 1492R ([21]. PCR reactions were adjusted to a volume of 25 µL containing 0.1 µM of each 16SrRNA primer, 25 ng of genomic DNA, 0.2 mM of each dNTP, and 1.25 U of Taq DNA polymerase. The regions were amplified under the following conditions: initial denaturation at 95 °C for 5 min, followed by 34 cycles of 94 °C for 50 s, 60 °C for 1 min, 72 °C for 1 min 30 s, and a final extension at 72 °C for 8 min. Amplification products were analyzed by electrophoresis on a 1.5% agarose gel. The amplicon was sequenced in the ABI 3730xl System (Macrogen, Inc.).

The sequences were trimmed and assembled using Geneious Prime 2022.0.1 (Biomatters). Next, the sequences were compared to the National Center for Biotechnology Information using BLASTn [22] against the 16S ribosomal RNA sequence (Bacteria and Archaea) database. We retrieved whole 16S rRNA sequences for each species' family taxa and built an in-house database. These databases were aligned and subjected to phylogenetic tree inference using the neighbor-joining method in MAFFT version 7 (https://mafft.cbrc.jp/alignment/server/large.html) [23]. The 16S rRNA sequences were then directly aligned against closely related strains in the Reference RNA sequence (Refseq) database retrieved for phylogenetic analysis. Next, the sequences were aligned using the ClustalW algorithm [24] in Mega11 [25]. The best-fit substitution model was calculated in Mega11, and phylogenetic trees were constructed using the maximum likelihood tree (1000 bootstrap replicates) and the substitution model General Time Revesible+gamma distribution with invariable sites (G+I).

2.4. Screening of bacterial isolates for in vitro plant growth-promotion traits and biocontrol

2.4.1. Bacterial growth under reduced water availability

Isolates were grown in Tryptone Soya Agar (TSA) medium (10%) with additional sorbitol at five different concentrations (0 g/L^{-1}; 85 g/L^{-1}, 285 g/L^{-1}, and 660 g/L^{-1}) to simulate water stress at 25°C [16].

2.4.2. Exopolysaccharide production

The isolates were inoculated onto 5 mm diameter paper discs disposed of in a DSMZ medium. The production of exopolysaccharide was checked by slime appearance and mixing a portion of the mucoid substance in 2 mL of absolute ethanol, in which the formation of a precipitate indicated the presence of EPS [17].

2.4.3. Indole acetic acid (IAA) production

Aliquots of 100 µL of bacteria were initially grown in 10 mL of TSA medium (10%) for 48 h in the dark at 28 °C. Next, colonies were transferred to fresh plaque-containing TSA medium (10%) supplemented with 5 mM L-tryptophan. The colonies were covered with a cellulose filter (0.45 µm pore size), and incubated in the dark at 25 °C. After 48h, the membranes were washed with Salkowski reagent (50 mL of perchloric acid (35%) and 1 mL of $FeCl_3$ solution (0.5 M) for 30 min in the dark [18]. The appearance of pink to red indicates IAA production.

2.4.4. Phosphate solubilization assay

Bacteria were inoculated into tubes with 10 ml of Tryptone Soya Broth (TSB) medium (10%) for one week at 28 °C and shaken on an orbital shaker at 180 rpm. Cells were collected by centrifugation and washed twice with 0.8% NaCl solution, and 20 μL of this suspension was spotted on the National Botanical Research Institute's phosphate growth medium (NBRIPM) containing per liter 15 g agar, 10 g glucose, 5 g $Ca_3(PO_4)_2$, 5 g $MgCl_2{\cdot}6H_2O$, 0.25 g $MgSO_4{\cdot}7H_2O$, 0.2 g KCl, and 0.1 g $(NH_4)_2SO_4$ [19]. The plates were incubated for 15 days at 25 °C. Positive phosphate solubilization was confirmed by the appearance of clear zones around spots.

2.4.5. Siderophore production

The isolates were selected for their ability to produce siderophores in CAS medium [20]. Bacteria were collected from TBS, resuspended, and washed twice with phosphate-buffered saline pH 6.5. A 20 μL aliquot of bacterial suspension was spotted on CAS agar plates. The production of siderophores was checked daily for color change from blue to red around each colony.

2.4.6. Biocontrol test

For the biocontrol test with the phytopathogenic fungus *Fusarium oxysporum* f. sp. *phaseoli*, mycelial discs of the fungus were collected and placed under a plate (1.5 cm) from the edge. On the opposite side of the plate, a streak of isolates was inoculated 1.5 cm from the edge. The plates were maintained at 25°C for one week. The percentage of growth inhibition was calculated using the formula (R1-R2)/R1 x 100, where R1 is the radial distance of the *F. oxysporum* f. sp. *phaseoli* mycelium in the absence of the antagonist from the center to the edge of the plate (measured in mm) and R2 is the growth distance of *F. oxysporum* f. sp. *phaseoli* from the center of the plate to the bank toward the isolate.

2.5. Plant growth promotion in greenhouse experiment

The seeds of soybean genotype Conquista were kindly provided by Thalita Avelar Monteiro from the Department of Plant Pathology of UFV. Seeds were surface-sterilized and inoculated with SOY2, SOY5, and SOY23 isolates by mixing for 2 h in the inoculum (10^8 CFU mL^{-1} (DO550 = 0.1). A control treatment was achieved by mixing the seeds with sterilized saline solution (0.85%). The seedlings were grown in plastic trays containing 500 g of a mixture of soil, sand, and manure (3:2:1). The plants were grown under natural sunlight in a greenhouse with an average daytime temperature of 12–33 °C. Soybean plants were watered daily with the same volume until the first trifoliate (stage V1) leaves emerged, after which a water restriction treatment was imposed. The soybean plants were subjected to the following two water treatments: soil relative water content of 30% (control) and 5% (drought stress). The soil water levels were monitored daily. An evaluation was performed 30 d after sowing. The leaf area, number of nodes, shoot and root lengths and shoot, and root dry biomass were determined. The greenhouse experiments were conducted using a completely randomized design. The data were subjected to one-way ANOVA and the Skott-Knott clustering algorithm.

2.6. DNA extraction and whole-genome sequencing

Kosakonia sp. strain SOY2 and *Agrobacterium* sp. strain SOY23 were grown into a flask containing 20 ml of liquid DSMZ medium for one week at 28 °C and shaken on an orbital shaker at 180 rpm. Cells were collected by centrifugation and Genomic DNA was extracted using the Wizard® Genomic DNA Purification Kit (Promega Corp.), as recommended by the manufacturer. DNA was checked for quality using a NanoDrop 2000 (Thermo Scientific) and subjected to gel electrophoresis (0.8% of agarose). The whole genome was sequenced using the DNBseq Sequencing platform at BGI, Inc.

2.7. Genome assembly

Raw data with adapter or low-quality sequences were filtered. We first went through a series of data processing to remove contamination and obtain valid data. This step was completed using the bynSOAPnuke software. SOAPnuke software filter parameters: " -n 0.01 -l 20 -q 0.4 --adaMis 3 --outQualSys 1 –minReadLen 150" [28]. The genome was assembled using a de novo assembler implemented as an initial assembly graph from short reads in Unicycler [29], and the assembly metrics were evaluated using QUAST v4.6 [30]. The completeness and contamination of all MAGs were estimated using CheckM (v1.0.11) [31]. The assemblies were annotated using the Prokka v. 1.14.6 [32]. The genomes were submitted to the National Center for Biotechnology Information (NCBI) GenBank. The genome sequence data were uploaded to the Type (Strain) Genome Server (TYGS), a free bioinformatics platform available at https://tygs.dsmz.de, for a whole genome-based taxonomic analysis [33].

2.8. Data retrieving from a public database and bioinformatics analysis

A total 169 genome sequences were retrieved from the NCBI database (last accessed in May 2020) (Table S1). These sequences were manually checked for their association with soil, plants, and rhizosphere according to the BioSample database. The proteome of the genomes was used to predict plant growth-promoting traits (PGPTs) through PLaBAse (v1.01, http://plabase.informatik.uni-tuebingen.de/pb/plabase.php) [34]. We also mined biosynthetic gene clusters (BGCs) using antiSMASH v5.1 [35]. Networks using similarity Minimum Information about a Biosynthetic Gene cluster database (MIBiG), using a locally installed version of the BiG-SCAPE software [36] with the local option enabled and a distance cut-off score of 0.3. The generated network was imported into Cytoscape version 3.7.2 and analyzed using default algorithms [37].

3. Results

3.1. The selection of distinct rhizobacteria with the potential for novel taxonomic species

VL55 medium, an oligotrophic medium commonly employed for the selection of slow-growing bacteria, was used to isolate rhizobacteria. We strictly selected 22 isolates rhizobacterial colonies with morphologically distinct characteristics, including shape, color, size, and texture, from soybean soil samples collected in the experimental field (Figure 1A). The isolates were assigned the acronym SOY (soybean) followed by a sequential number. A DNA fingerprint analysis was performed to examine the genetic profiles of the isolates. Three molecular markers were tested: ERIC, BOX, and GTG. The gel displayed various band patterns employing the markers, with BOX and GTG markers displaying more determinant characteristics for isolate identification (Figure S1). As a result, these two markers were chosen for the genetic profiling of the 22 isolates. A dendrogram revealed three large clusters for the BOX marker (Figure S2A) and three large clusters for the GTG marker (Figure S2B), each with several ramifications (Figure S2B). In general, most isolates had distinct genetic profiles; nevertheless, comparable profiles, such as isolated SOY19 and SOY20 for both markers, were detected, suggesting that the two isolates were genetically related (Figure S2). These findings suggests that this selection strategy allowed for the isolation of genetically distinct isolates.

Taxonomic identification of the isolates was based on sequencing of the gene encoding 16S rRNA. MEGAX was used to create a phylogenetic tree of the sequences. The retrieved sequences were compared with those in NCBI database in terms of coverage and identity. Species closely related to ten genera were found: *Kosakonia, Microbacterium, Mycobacterium, Methylobacterium, Monashia, Novosphingobium, Pandoraea, Arthrobacter, Stenotrophomonas,* and *Rhizobium* (Table S2, Figure 1B). The 16S phylogenetic tree analysis also found ten groupings related to the genera. Thirteen of the 22 isolates analyzed had sequence identities lower than 96% when compared to the database sequences, indicating that these taxa may belong to new genera or species. This finding was supported by phylogenetic analysis, which showed that while the isolates were related to these taxa, different clades were formed

(Figure 1B). Based on the comparison of isolates by NCBI and phylogeny, the proposed classification of these isolates is shown in Table S2.

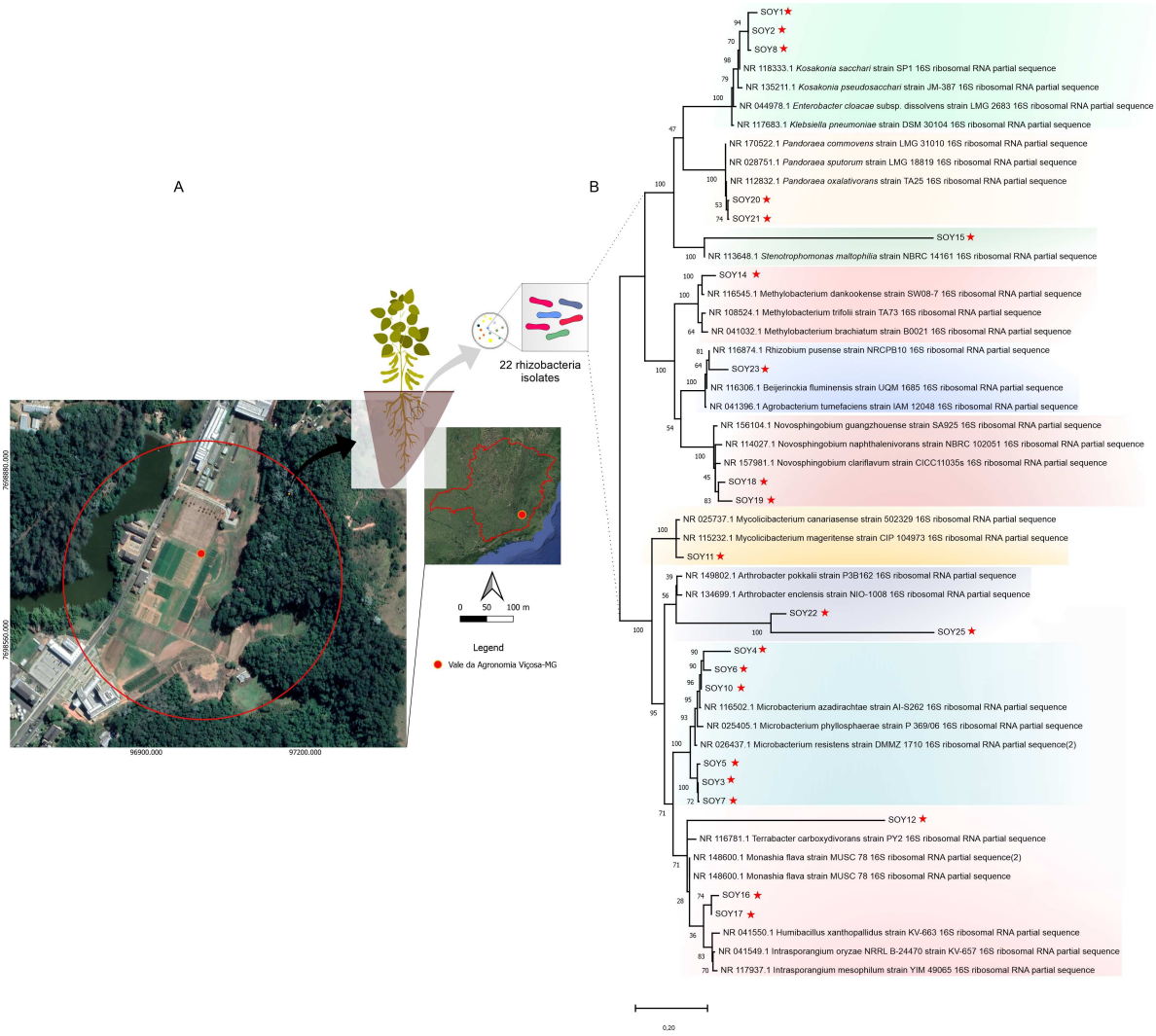


Figure 1. Study site of sample collection and phylogenetic tree of 22 rhizobacteria isolated from soybean. (A) The experimental field at Universidade Federal de Viçosa, Minas Gerais, Brazil (20°46'01.1"S 42°52'10.2"W), indicated by the sampling sites in the red dot. The map was built using QGIS 3.16 Hannover (DATUM: SIRGAS2000, UTM Zone 22S). (B) Likelihood phylogenetic tree of 22 rhizobacteria. The evolutionary history was inferred using the maximum likelihood tree (1000 bootstrap replicates) and the substitution model general time-revesible + gamma distribution with invariable sites (G + I). The scale bar at the bottom indicates the number of differences in base composition among the sequences. Red stars indicate the 22 isolates described herein.

3.2. The novel bacterial species exhibited plant growth-promoting traits

The isolates were analyzed for their capacity to generate growth-promoting characteristics such as IAA, EPS, and phosphate solubilization. Seventy-two percent (n=16) of the 22 rhizobacterial isolates produced IAA, 68% (n=12) solubilized phosphate, and 45% (n=10) produced EPS (Figure 2A). Furthermore, the isolates were tested for their ability to grow in a medium with low water activity. DSMZ medium containing four amounts of sorbitol (0 g/L^{-1}, 285 g/L^{-1}, 520 g/L^{-1}, and 660 g/L^{-1}) was used (Figure 2B). Approximately 91% (n=22) showed positive growth at 285 g/L^{-1} sorbitol concentration, 75% (n=18) showed positive growth at 520 g/L^{-1} sorbitol concentration, and only *Microbacterium* sp. strain

SOY5 exhibited positive growth at all sorbitol concentrations. The lower the water activity, the higher the sorbitol content. Therefore, we observed that bacterial growth was minimal at the highest concentration of sorbitol. Taken together, these findings suggest that the majority of the isolates possessed one or more growth-promoting properties and that these bacteria may enhance plant development under water stress.

Five isolates with positive results for IAA, phosphate solubilization, EPS, and growth in reduced water were selected: *Kosakonia* sp. strain SOY1, *Kosakonia* sp. strain SOY2, *Microbacterium* sp. strain SOY4, *Monashia* sp. strain SOY12, and *Agrobacterium* sp. strain SOY23. The isolates were initially evaluated to determine their growth pattern over time (in hours) for the plant test. The bacteria were grown for 10 h, during which time a typical growth curve was observed, with the bacteria reaching the initial stationary phase (Figure S3). After 3 h of incubation, an aliquot of the culture medium at OD 1.0 was plated on a nutritional agar medium for cell counting and viability analysis. The plate cell count for all isolates showed a cell density of 10^8 CFU/mL, thus validating the optimal spot on the growth curve for plant inoculation. Soybean seeds were treated with the isolates and stored for two weeks. Plants treated with the bacterial suspensions of the isolates grew at a faster rate at the end of the first week. Given that the initial leaflets were visible in all treatments, a rise in the stems of the plants treated with bacterial suspensions was generally observed under laboratory conditions.

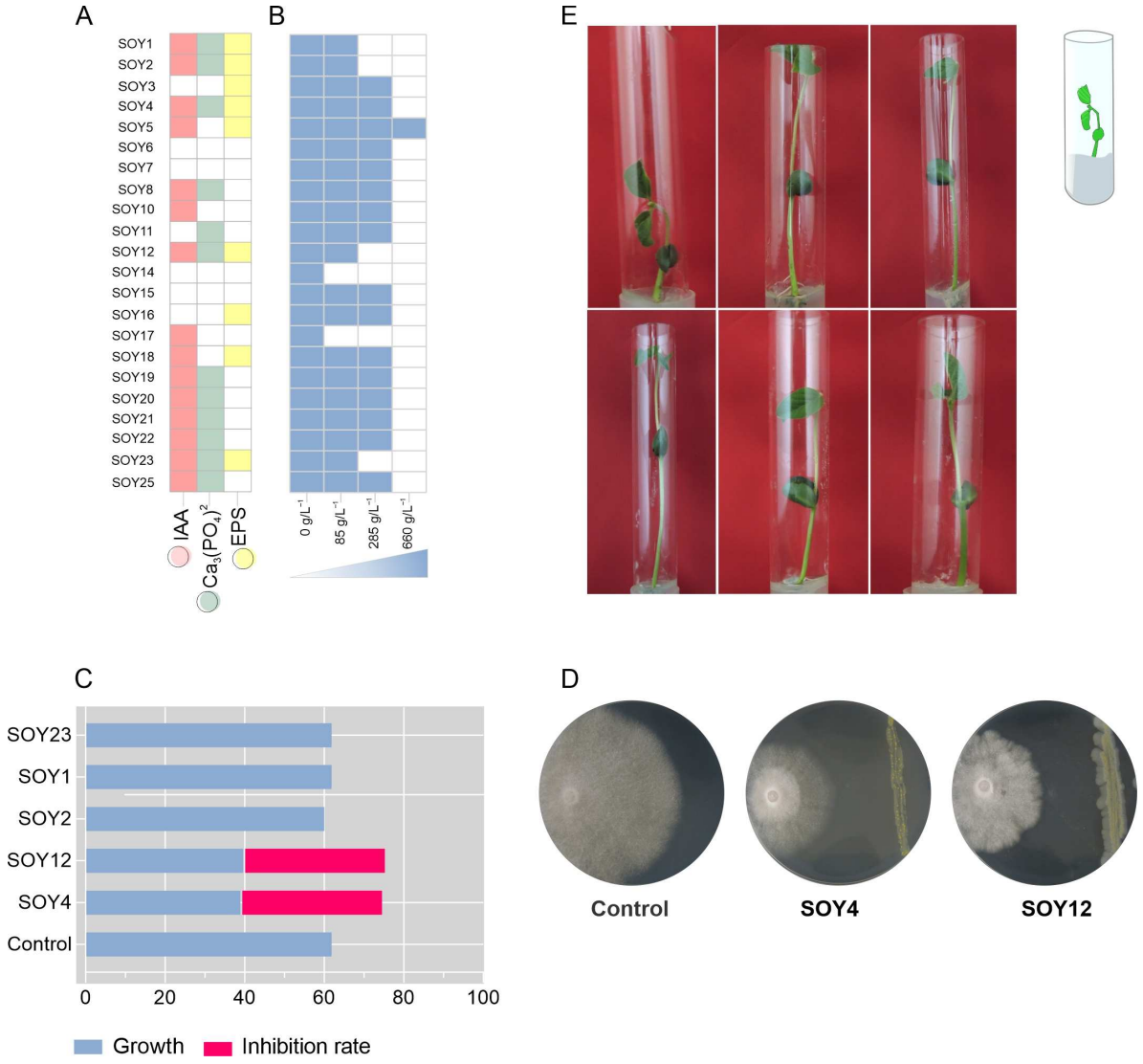


Figure 2. *In vitro* features for plants promoting growth and biological control. (A) The production of features for plants promotes growth. Left to right: synthesis of indole acetic acid (IAA) in pink; phosphate solubilization ($Ca_3(PO_4)_2$ in green; exopolysaccharide (EPS) production in yellow. Colored

squares indicate the growth of each isolate. (B) Ability of bacteria to grow in a medium with reduced water availability containing different sorbitol concentrations at increasing levels. The colored squads indicate the growth of each isolate. (C) Biological phytopathogen control. On the x-axis, the growth rate in centimeters by the fungus and the isolates are plotted on the y-axis. The growth of the fungus is shown in blue, and the inhibition rate of the fungal growth is shown in pink. (D) Paired culture test with *Microbacterium* sp. strain SOY4 and *Monashia* sp. strain SOY12 against *Fusarium oxysporum* f. sp. *phaseoli* sp. (E) *In vitro* soybean growth-promotion tests; left to right: control plant with PBS buffer, plant inoculated with SOY1, plants inoculated with SOY2, SOY4, and SOY12 plants inoculated with SOY23.

Finally, the five isolates previously chosen for growth promotion were examined for biocontrol efficacy against the phytopathogen *Fusarium oxysporum* f. sp. *phaseoli*, which causes plant wilting in common beans [38]. *Fusarium oxysporum* f. sp. *phaseoli* was chosen as the model pathogen because *F. oxysporum* is a globally dispersed disease with several hosts. Paired *in vitro* culture activity research revealed that *Microbacterium* sp. strain SOY4 and *Monashia* sp. strain SOY12 inhibited plant pathogen development by 35%, indicating that these two isolates also have the ability to control phytopathogens.

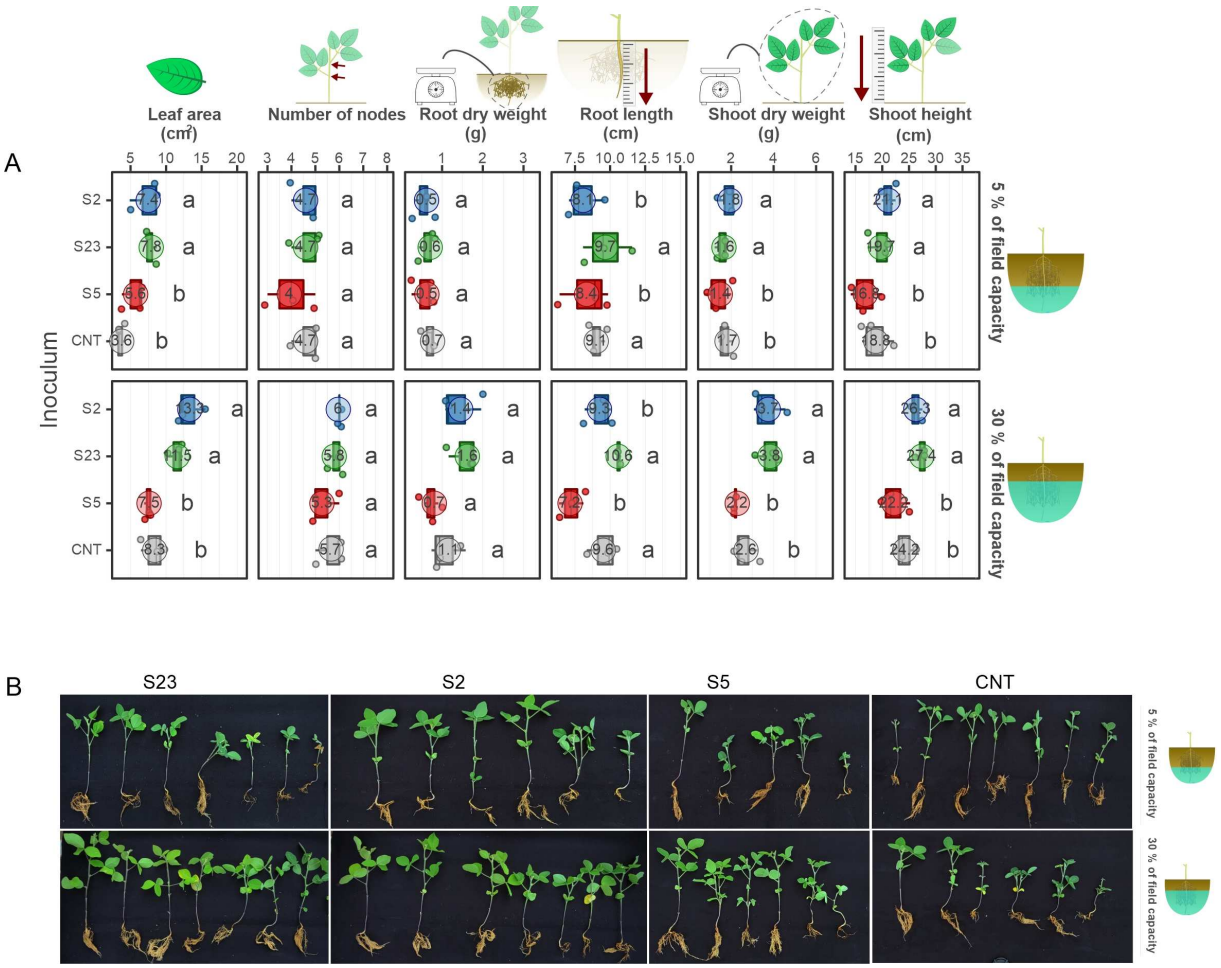


Figure 3. Soybean growth under drought in greenhouse conditions. (A) Phenotypic traits were measured in soybean plants under two water deficit conditions (5% and 30% field capacity) and inoculation with three bacterial isolates. Treatments that showed the same lowercase letter within each water deficit condition did not show statistically significant differences according to the skott-knott test. (B) Non-inoculated soybean plants (control) and plants inoculated with different bacteria (SOY2, SOY5, and SOY23) and under drought stress.

3.3. Two isolates showed the ability to enhance soybean growth under drought in greenhouse conditions

The ability to enhance soybean growth under drought in greenhouse conditions was tested for the *Kosakonia* sp. strain SOY2, *Microbacterium* sp. strain SOY5, and *Agrobacterium* sp. strain SOY23. We found that the SOY2 and SOY23 bacteria showed potential in water stress mitigation. Under both water stress conditions, plants whose seeds were inoculated with these bacteria generated more dry matter in the shoot (Figure 3) and had a smaller reduction in leaf area than the other treatments (control without inoculation and inoculation with SOY5) (severe and moderate). There was no statistically significant change in the root system mass across treatments (Figure 3A). Plants inoculated with these two bacteria (SOY2 and SOY23) had a shorter root length than the other treatments. Under water stress, the plant's root system usually changes, such as the production of more sharp angles between, to make it deeper [39], allowing the plant to use the water available in deeper soil strata. However, changing the architecture of the root system consumes energy that the plant may otherwise employ. Thus, when compared to SOY2 and SOY23, the increase in root length in SOY5 and the control treatments may imply less stress mitigation Non-inoculated plants showed a 56% reduction in leaf area (Figure 3A), whereas SOY2 and SOY23 bacterial treatments reduced leaf area by 25 and 32%, respectively. In other words, compared to the control, inoculation with these bacteria reduced the leaf area loss by more than half. Plants with larger leaf areas have a larger surface area for catching light, which implies that photo-assimilates (sugars) are produced at a higher rate than plants with smaller leaf areas, allowing for more grain filling during the reproductive phase (Figure 3B).

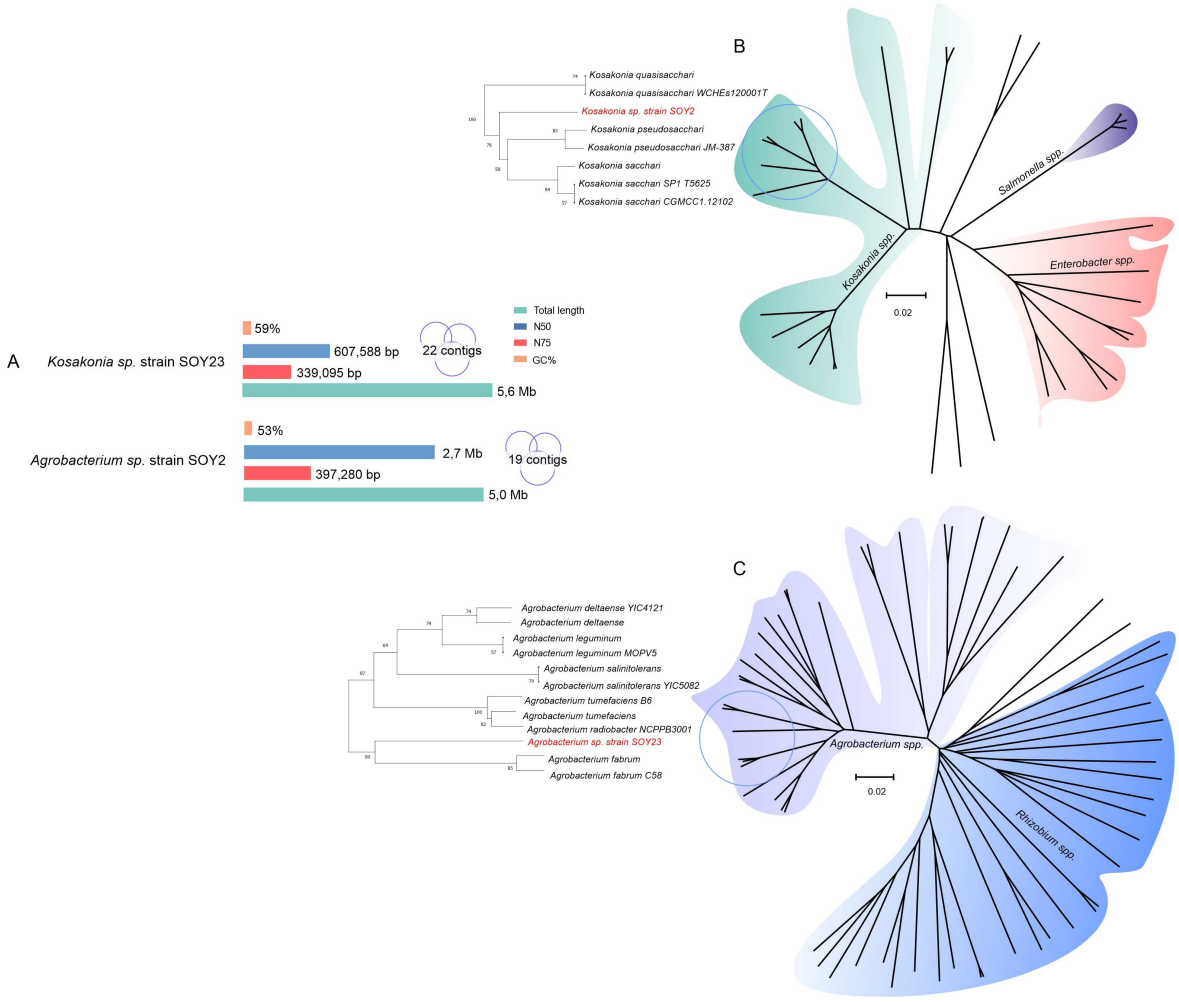


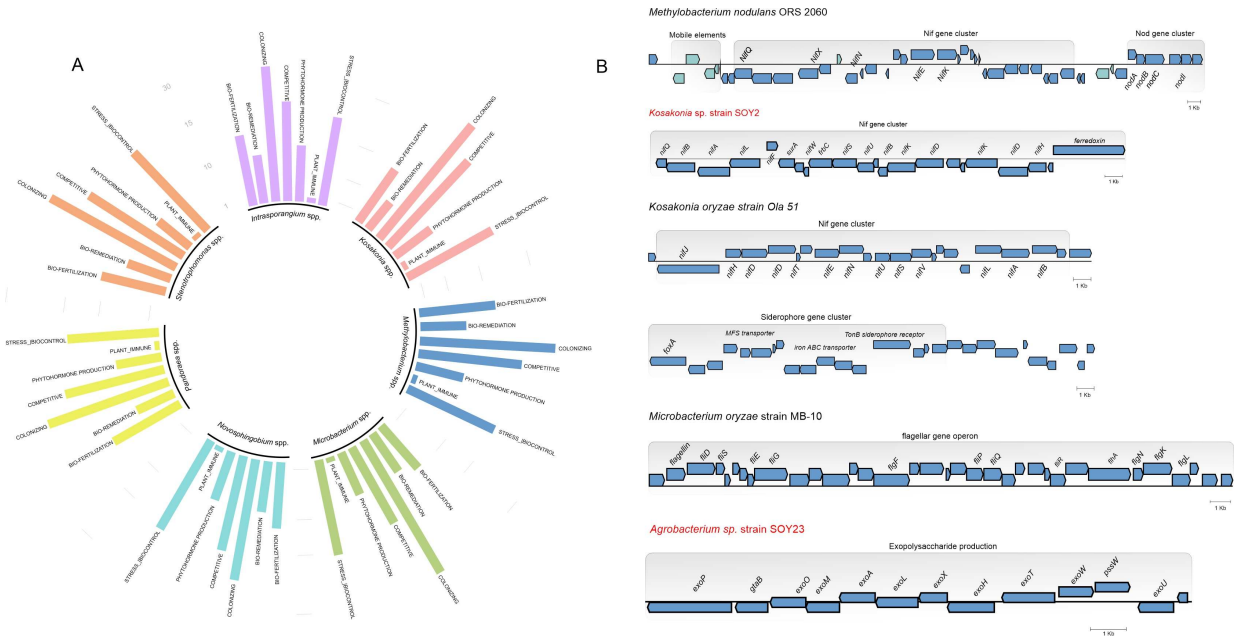
Figure 4. Whole-genome sequences of SOY2 and SOY23 rhizobacteria. (A) Genomic features of *Kosakonia* sp. strain SOY2 and *Agrobacterium* sp. SOY23. (B) Phylogenomic tree of *Kosakonia* sp. strain SOY2 inferred with FastME 2.1.6.1 from GBDP distances calculated from genome sequences. (C) Phylogenomic tree of *Agrobacterium* sp. strain SOY23. The branch lengths were scaled in terms of the

GBDP distance formula d_5. The numbers above are GBDP pseudo-bootstrap support values of > 60 % from 100 replications, with an average branch support of 80.8 %. The tree was rooted at its midpoint. An essential operon encoding plant growth-promoting genes was found. The *nif* gene cluster in *Methylobacterium nodulans*, shows the *nif* cluster surrounded by mobile genetic elements and nod genes organized in a single operon with six ORFs. We also found a conservative nif operon in *Kosakonia* sp. strain SOY2 and *K. oryzae*, as well as a single operon for siderophore synthesis. In addition, a flagellar operon observed in *M. oryzae* responsible for plant system colonization, and a single cluster for EPS synthesis in *Agrobacterium* sp. strain SOY23 was also mapped.

3.4. The genomic sequences of Kosakonia and Agrobacterium

Given the potential *of Kosakonia* sp. strain SOY2 and *Agrobacterium* sp. strain SOY23 to positive interact with soybean *in vitro* and *in planta*, here we report a draft genome sequence of these two strains to gain a better understanding of this interaction with the plant. The assembly of *Kosakonia* sp. strain SOY2 consisted of 19 contigs with a total sequence length of 5,035,089 bp, an N50 value of 2,707,390 bp, and a GC content of 53.81% (Figure 4A). The closest placement taxonomy result was *Kosakonia* sp000410515, with an average nucleotide identity (ANI) of 95.63 %. According to the whole-genome phylogeny, this strain is related to plant-associated *Kosakonia* spp. However, the genome was not assigned to any species (Figure 4B). A total of 4,748 CDS, 3 rRNAs (5S, 16S, and 23S), 71 tRNAs, and 1 tmRNA were found in the *Kosakonia* sp. strain SOY2 genome.

The assembly of *Agrobacterium* sp. strain SOY23 consisted of 22 contigs with a total sequence length of 5,626,101 bp, N50 value of 607,588 bp, and GC content of 59.59%. Surprisingly, the genome was not assigned to any closest species, as it was outside the predefined ANI radius (Table S2, Figure 4B). CheckM was employed to estimate the completeness and contamination of the genome, which were 99.94% and 0.35%, respectively. A total of 5,225 CDS, 46 tRNAs, and 1 tmRNA were found in the *Agrobacterium* sp. strain SOY23 genome.



Figue 5. *In silico* identification and comparison of plant growth-promoting genes. (A) Circular plot illustrating the average number of plant growth-promoting genes found in each of the seven genera. (B) Genes encoding plant colonization and plant growth promotion associated with nitrogen fixation, siderophores, flagella, and exopolysaccharides found in the genomes.

3.5. A genome mining analysis of novel s*pecies of soil bacteria revealed several proteins with traits that promote plant growth*

To better understand the potential of all species described here, we gathered publicly available genomes from the NCBI database and performed *in-silico* identification and comparison of plant growth-promoting genes. We first sought plant growth-promoting genes in 169 genome sequences associated with soil, plants, and rhizosphere (Table S1). We found that the majority of these genomes encode proteins with direct effects on the plant, such as biofertilization, iron acquisition, P and K solubilization, phytohormone synthesis, and nitrogen fixation. These genomes possess heavy metal resistance genes such as cobalt, copper, nickel, and selenium, which might be exploited for bioremediation. Furthermore, these bacteria encode traits that aid in plant system colonization, such as chemotaxis proteins and motility, and they may utilize plant-derived amino acids, sugars, and peptides. In addition, it neutralizes abiotic stresses, such as high and low temperature and acidic, osmotic, and salinity stress. The average number of each gene class per genome is shown in Figure 5 and Table S3.

We also explored the repertoire of secondary metabolite biosynthetic gene clusters (BGCs) encoded within these genomes using AntiSMASH. In total, 506 putative BGC regions were identified (Table S4). The majority of the BGC were found in the *Methylobacterium* (146) followed by *Arthrobacter* (120). There were 93 terpene-containing gene clusters found, mostly in *Methylobacterium*, 47 NAPAA clusters from three genera (*Arthrobacter*, *Methylobacterium*, and *Microbacterium*), 43 betalactone clusters from 76 phyla, and 40 Type III polyketide synthases clusters. In general, the BGC regions in length varied from 2 kb to 76kb, the largest size an NRP + Polyketide: Modular type I found in the *Microbacterium amylolyticum* DSM 24221. In addition, we used BiG-SCAPE to construct a BGC sequence similarity network. We discovered that the majority of BGCs were taxonomically distributed among the same genera and architecturally heterogeneous (Figure S4, Figure S5). In general, nearly none of the BGCs found here were grouped with the MIBiG reference BGCs, indicating a repertoire of novel secondary metabolite BGCs.

4. Discussion

In this study, 22 soybean growth-promoting rhizobacteria were identified. 78% of the 22 bacteria isolated exhibited positive results for water stress growth, indicating their potential for application in crops with a shortage of water. The isolates produced IAA, the primary auxin found in plants, which is synthesized in the apical system of the stem and transported to the roots; plant-associated microbes can also synthesize it. Its primary effect is the growth of roots and stems [40,41]. An *in vitro* test of soybean revealed this elongation effect. Furthermore, the isolates were able to solubilize phosphorus, an important element in plant metabolism. Plant growth is hampered by its absence; nevertheless, the least available nutrients for the plant since it is held by the precipitation of other soil elements, resulting in insoluble inorganic phosphates with the lowest available for the plant [42,43]. As a result, bacteria with the capacity to solubilize insoluble inorganic phosphate sources, increasing the soluble phosphorus level in the soil solution and, plant availability, play a crucial role in the phosphorus biogeochemical cycle [44].

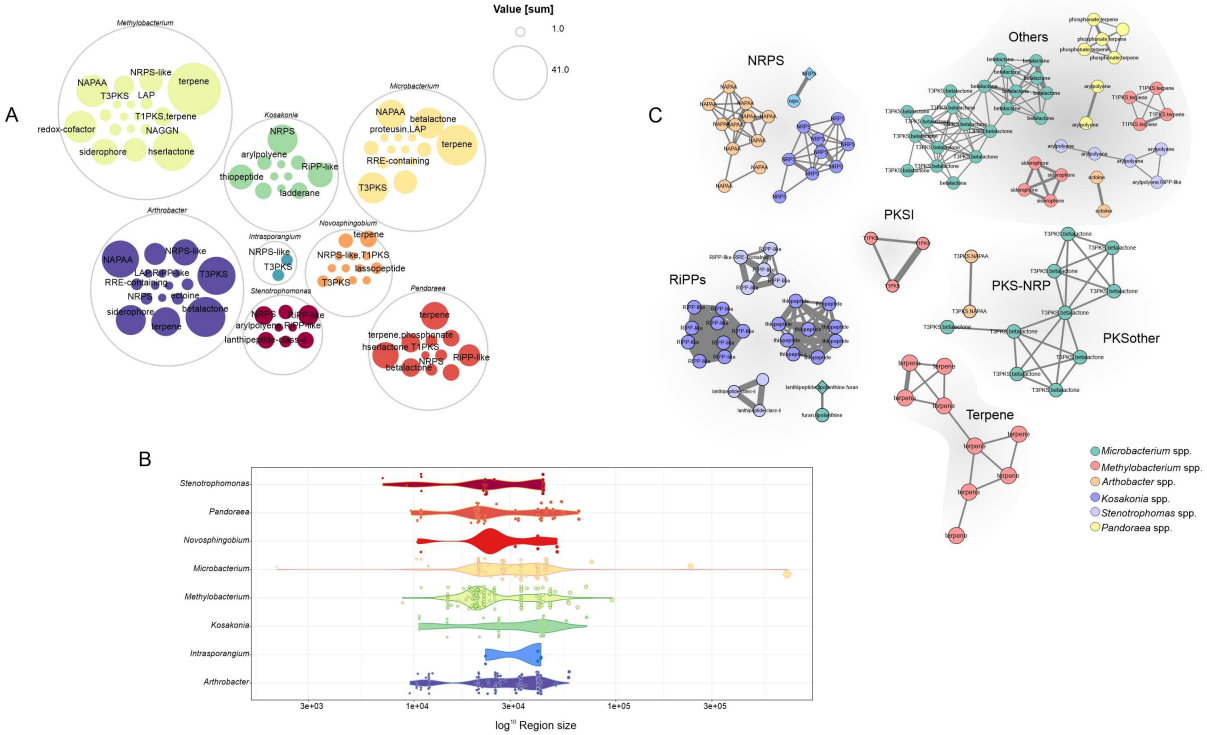


Figure 6. Identification and comparison of biosynthetic gene clusters. (A) Genera-level BGCs distribution. (B) Length of BGCs across genera. (C) Sequence similarity networks of the BGCs.

The genera *Bacillus*, *Pseudomonas*, and *Burkholderia* have been commonly identified as widely prevalent growth-promoting bacteria in soybeans [45–47]. Here, by contrast, we discovered *Kosakania* sp., *Microbacterium* sp., *Mycobacterium* sp., *Methylobacterium* sp., *Novosphingobium* sp., *Anthrobacter* sp., *Stenotrophomonas* sp., *Monashia* sp., and *Pandoraea* sp. This could be due to the VL55 isolation medium used in this study. VL55 is a defined medium that is low in nutrients and has pH in the most acidic range. Because xylan is the only available carbon source, it is commonly employed for the selection of bacteria that were previously classified as "unculturable" in soil [11,48]. Although the isolates identified in this study were grouped close to the aforementioned genera, phylogenetic analysis revealed that they may belong to new genera and/or species.

Some of the genera shown here are related to plant growth promotion activities, such as *Microbacterium* in Neem growth promotion [49]. Another example is the genus *Methylobacterium*, which absorbs carbon molecules during endophytic associations and releases biopolymers, organic acids, coenzymes, vitamins, and toxins that aid in disease management [50,51]. Yet, few studies have been conducted to investigate the interactions of this genus with plants. In addition to species already known to promote plant growth, genera such as *Novosphingobium* are known to produce enzymes capable of degrading aromatic compounds [52,53]; *Arthrobacter* is used in many industrial applications and have the potential to be used in bioremediation [54]. Surprisingly, genera known to cause diseases in humans, such as *Pandoraea* and *Stenotrophomonas* were isolated. However, based on the phylogenetic distance between the species, these isolates may belong to a distinct group of bacteria unrelated to human diseases.

Microbacterium sp. strain SOY4 and *Monashia* sp. strain SOY12 were effective in suppressing the phytopathogen *Fusarium oxysporum* f. sp. *phaseoli*. Both isolates were Actinobacteria, a phylum recognized for producing secondary metabolic metabolites with broad antifungal properties [55]. In addition, we have showed the ability of two isolates SOY2 (*Kosakonia* sp.) and SOY23 (*Agrobacterium* sp.) to enhance soybean growth under drought in greenhouse conditions. Plants seeds inoculated with these bacteria generated more dry matter in the shoots and had a smaller reduction in leaf area than the other treatments. Although *Agrobacterium* has been identified as a plant pathogenic bacterium, certain studies

have shown that specific tumor-inducing *Agrobacterium* strains can stimulate plant growth in non-susceptible plant hosts[56,57].

The use of genomics techniques in investigations aimed at the potential of plant growth-promoting bacteria has provided evidence of the genetic characteristics that promote the microbe-plant interaction [58]. Here, we gained insight into the genomic potential of two bacteria of our isolates SOY2 and SOY23 along with the bacteria genus mentioned in this study. We confirmed that SOY2 and SOY23 did not belong to any of the nearest species and represented two novel bacteria species, and together with 169 genomes, have a repertoire of genes encoding for plant growth-promoting proteins with direct effects on the plant, such as bio-fertilization, iron acquisition, P and K solubilization, phytohormone synthesis, and nitrogen fixation. In addition, secondary-metabolite BGCs analysis revealed a variety of novel secondary metabolite BGCs to be explored.

Well-known plant growth-promoting rhizobacteria (PGPR) have been widely used in the commercial sector with success; however, the search for new microbes that are resilient to adverse effects such as drought and have the potential to promote the growth of crops plants is becoming increasingly relevant, given climate change and its impact on food production. Chauhan and colleagues [59] reported several novel PGPRs that have not yet been achieved for commercial scales of production.

4. Conclusions

This work adds to the inclusion of new species with significant potential for promoting plant growth. Taken together, here we demonstrated novel soil bacteria with growth-promoting capability that had not previously been reported for soybean. In addition, we demonstrated the importance of coupling more complex medium culture with bioinformatics approaches to select new PGPR. The findings enable the identification of distinct bacteria with a high potential for promoting plant growth, opening the path for future research and uses in agriculture intending to reduce the environmental impact of synthetic industrial pesticides, and fertilizers, and help mitigate drought stress.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, Figure S1: Figure S1. Depicts a Molecular Marker Test. According to the line above, the markers employed are delimiting the wells. M stands for GeneRuler 1Kb Plus DNA Ladder molecular weight marker; Figure S2. Dendrogram analysis of isolate genetic profiles. The closeness value is shown by the colored circles with values at the intersections of the trees. A. Dendrogram made use of a BOX marker. B. dendrogram with the GTG marker. The appropriate 1.5% agarose electrophoresis gels are shown below the dendrograms. The name of the isolate is used to denote each well. The 1Kb and 100bp markers are in the first and last wells, respectively.; Figure S3. Standard growth curve for isolates. A. Dispersion curve. The x-axis shows time in hours, and the y-axis shows absorbance in nm. B. Bar graph depicting plaque growth after 3 hours. The X-axis represents isolates, the y-axis represents colony-forming units (CFU) per ml; Figure S4. Biosynthetic gene cluster types identified from the 169 publicly genomic sequences available for soil bacteria; Figure S5. CORASON phylogeny of Biosynthetic gene cluster types identified from the 169 publicly genomic sequences available for soil bacteria; Table S1: Information regarding the 169 publicly available genomes obtained from NCBI database; Table S2: Taxonomic identification of isolates; Table S3: Classes of plant growth-promoting genes found in each of the genomes; Table S4: Biosynthetic gene clusters identified from the 169 publicly genomic sequences available for soil bacteria.

Author Contributions: O.S.G., Conceptualization, Investigation, Software, Data curation, Methodology, Visualization, Writing - original draft; T.S.S., Investigation, Methodology, Writing - original draft; G. C.G., Investigation, Methodology, Visualization; T.G.R.S., Formal analysis, Methodology, Software. A.S.F., Investigation, Methodology. S.M.T., Software, Data curation. E.A.G., Investigation, Methodology. M.F.S., Conceptualization, resources, Supervision, Writing - review & editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq (Process Nos. 143132/2019-9), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG, Grant # APQ-02381-21) and Coordenacão de Aperfeiçoamento de Pessoal de Nível Superior/Programa de Excelência Acadêmica-Finance Code 001.

Data Availability Statement: 16S rRNA sequences are available under the following accession numbers from OQ054593 to OQ054611 in the Center for Biotechnology Information (NCBI). The genomes sequences are available in the NCBI database under the accession number JAPVXH000000000 for *Kosakonia* sp. strain SOY2 and

JAPWDT000000000 for *Agrobacterium* sp. strain SOY23. The accession numbers of publicly genome sequences used can be found in the Supplementary Material for this manuscript.

Acknowledgments: The authors thank Thalita Avelar Monteiro for providing the soybean seeds used in this study and Ralph Bonandi for drawing the map in Figure 1.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Berg, G.; Grube, M.; Schloter, M.; Smalla, K. Unraveling the plant microbiome: looking back and future perspectives. *Frontiers in Microbiology 2014*, *5*, 148.
2. Dastogeer, K.M.G.; Tumpa, F.H.; Sultana, A.; Akter, M.A.; Chakraborty, A. Plant microbiome–an account of the factors that shape community composition and diversity. *Curr Plant Biol* **2020**, *23*, 100161, doi: https://doi.org/10.1016/j.cpb.2020.100161.
3. Naylor, D.; Coleman-Derr, D. Drought stress and root-associated bacterial communities. *Frontiers in Plant Science 2018*, *8*, 2223.
4. Staley, J.T.; Konopka, A. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol* **1985**, *39*, 321–346, doi:10.1146/annurev.mi.39.100185.001541.
5. Bardgett, R.D.; van der Putten, W.H. Belowground biodiversity and ecosystem functioning. *Nature* **2014**, *515*, 505–511, doi:10.1038/nature13855.
6. Tiedje, J.M.; Asuming-Brempong, S.; Nüsslein, K.; Marsh, T.L.; Flynn, S.J. Opening the black box of soil microbial diversity. *Applied Soil Ecology* **1999**, *13*, 109–122, doi: https://doi.org/10.1016/S0929-1393(99)00026-8.
7. Sokol, N.W.; Slessarev, E.; Marschmann, G.L.; Nicolas, A.; Blazewicz, S.J.; Brodie, E.L.; Firestone, M.K.; Foley, M.M.; Hestrin, R.; Hungate, B.A.; et al. Life and death in the soil microbiome: how ecological processes influence biogeochemistry. *Nat Rev Microbiol* **2022**, *20*, 415–430, doi:10.1038/s41579-022-00695-z.
8. Andrews, J.H.; Harris, R.F. R- and K-Selection and Microbial Ecology BT - Advances in microbial ecology. In Marshall, K.C., Ed.; Springer US: Boston, MA, 1986; pp. 99–147 ISBN 978-1-4757-0611-6.
9. Kielak, A.M.; Cipriano, M.A.P.; Kuramae, E.E. *Acidobacteria* strains from subdivision 1 act as plant growth-promoting bacteria. *Arch Microbiol* **2016**, *198*, 987–993, doi:10.1007/s00203-016-1260-2.
10. Huang, Y.S.; Shen, F.T. Bioprospecting of facultatively oligotrophic bacteria from non-rhizospheric soils. *Applied Soil Ecology* **2016**, *108*, 315–324, doi: 10.1016/J.APSOIL.2016.09.004.
11. Joseph, S.J.; Hugenholtz, P.; Sangwan, P.; Osborne, C.A.; Janssen, P.H. Laboratory cultivation of widespread and previously uncultured soil bacteria. *Appl Environ Microbiol* **2003**, *69*, 7210–7215, doi:10.1128/aem.69.12.7210-7215.2003.
12. Kato, S.; Yamagishi, A.; Daimon, S.; Kawasaki, K.; Tamaki, H.; Kitagawa, W.; Abe, A.; Tanaka, M.; Sone, T.; Asano, K.; et al. Isolation of previously uncultured slow-growing bacteria by using a simple modification in the preparation of agar media. *Appl Environ Microbiol* **2018**, *84*, e00807-18, doi:10.1128/AEM.00807-18.
13. Pulschen, A.A.; Bendia, A.G.; Fricker, A.D.; Pellizari, V.H.; Galante, D.; Rodrigues, F. Isolation of uncultured bacteria from antarctica using long incubation periods and low nutritional media. *Front Microbiol* **2017**, *8*, 1346, doi:10.3389/fmicb.2017.01346.
14. Tschech, A.; Pfennig, N. Growth yield increase linked to caffeate reduction in *Acetobacterium Woodii*. *Arch Microbiol* **1984**, *137*, 163–167, doi:10.1007/BF00414460.
15. Widdel, F.; Kohring, G.-W.; Mayer, F. Studies on dissimilatory sulfate-reducing bacteria that decompose fatty acids. *Arch Microbiol* **1983**, *134*, 286–294, doi:10.1007/BF00407804.
16. Kavamura, V.N.; Santos, S.N.; Silva, J.L. da; Parma, M.M.; Ávila, L.A.; Visconti, A.; Zucchi, T.D.; Taketani, R.G.; Andreote, F.D.; Melo, I.S. de Screening of Brazilian cacti rhizobacteria for plant growth

promotion under drought. *Microbiol Res* **2013**, *168*, 183–191, doi: https://doi.org/10.1016/j.micres.2012.12.002.

17. Paulo, E.M.; Vasconcelos, M.P.; Oliveira, I.S.; Affe, H.M. de J.; Nascimento, R.; Melo, I.S. de; Roque, M.R. de A.; Assis, S.A. de An alternative method for screening lactic acid bacteria to produce exopolysaccharides with rapid confirmation. *Food Science and Technology* **2012**, *32*, 710–714, doi:10.1590/S0101-20612012005000094.
18. Gordon, S.A.; Weber, R.P. Colorimetric estimation of indoleacetic acid. *Plant Physiol* **1951**, *26*, 192–195, doi:10.1104/pp.26.1.192.
19. Nautiyal, C.S. An efficient microbiological growth medium for screening phosphate solubilizing microorganisms. *FEMS Microbiol Lett* **1999**, *170*, 265–270, doi:10.1111/j.1574-6968.1999.tb13383.x.
20. Schwyn, B.; Neilands, J.B. Universal chemical assay for the detection and determination of Siderophores. *Anal Biochem* **1987**, *160*, 47–56, doi:10.1016/0003-2697(87)90612-9.
21. Heuer, H.; Krsek, M.; Baker, P.; Smalla, K.; Wellington, E.M. Analysis of actinomycete communities by specific amplification of genes encoding 16s rRNA and gel-electrophoretic separation in denaturing gradients. *Appl Environ Microbiol* **1997**, *63*, 3233–3241.
22. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic Local Alignment Search Tool. *J Mol Biol* **1990**, *215*, 403–410, doi:10.1016/S0022-2836(05)80360-2.
23. Katoh, K.; Rozewicki, J.; Yamada, K.D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* **2019**, *20*, 1160–1166, doi:10.1093/bib/bbx108.
24. Larkin, M.A.; Wilm, A.; Higgins, D.G.; Valentin, F.; Blackshields, G.; McWilliam, H.; Wallace, I.M.; Thompson, J.D.; Brown, N.P.; McGettigan, P.A.; et al. Clustal W and Clustal X Version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948, doi:10.1093/bioinformatics/btm404.
25. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis version 11. *Mol Biol Evol* **2021**, *38*, 3022–3027, doi:10.1093/molbev/msab120.
26. Koeuth, T.; Versalovic, J.; Lupski, J.R. Differential subsequence conservation of interspersed repetitive *Streptococcus Pneumoniae* BOX elements in diverse bacteria. *Genome Res* **1995**, *5*, 408–418, doi:10.1101/gr.5.4.408.
27. Versalovic, J.; Koeuth, T.; Lupski, J.R. Distribution of repetitive dna sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucleic Acids Res* **1991**, *19*, 6823–6831, doi:10.1093/nar/19.24.6823.
28. Chan, Y.; Chen, Y.; Shi, C.; Huang, Z.; Yong, Z.; Shengkang, L.; Li, Y.; Ye, J.; Yu, C.; Li, Z.; et al. SOAPnuke: A MapReduce acceleration supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* **2017**, *7*, doi:10.1093/gigascience/gix120.
29. Wick, R.R.; Judd, L.M.; Gorrie, C.L.; Holt, K.E. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* **2017**, *13*, e1005595-.
30. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality Assessment Tool for Genome Assemblies. *Bioinformatics* **2013**, *29*, 1072–1075, doi:10.1093/bioinformatics/btt086.
31. Parks, D.; Imelfort, M.; Skennerton, C.; Philip, H.; Tyson, G. CheckM: Assessing the Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes. *Genome Res* **2015**, *25*, doi:10.1101/gr.186072.114.
32. Seemann, T. Prokka: Rapid Prokaryotic Genome Annotation. *Bioinformatics* **2014**, *30*, 2068–2069, doi:10.1093/bioinformatics/btu153.

33. Meier-Kolthoff, J.P.; Göker, M. TYGS Is an Automated High-Throughput Platform for State-of-the-Art Genome-Based Taxonomy. *Nat Commun* **2019**, *10*, 2182, doi:10.1038/s41467-019-10210-3.
34. Patz, S.; Gautam, A.; Becker, M.; Ruppel, S.; Rodríguez-Palenzuela, P.; Huson, D.H. PLaBAse: A comprehensive web resource for analyzing the plant growth-promoting potential of plant-associated bacteria. *bioRxiv* **2021**, 2021.12.13.472471, doi:10.1101/2021.12.13.472471.
35. Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S.Y.; Medema, M.H.; Weber, T. AntiSMASH 5.0: Updates to the Secondary Metabolite Genome Mining Pipeline. *Nucleic Acids Res* **2019**, *47*, W81–W87, doi:10.1093/nar/gkz310.
36. Navarro-Munoz, J.; Selem, N.; Mullowney, M.; Kautsar, S.; Tryon, J.; Parkinson, E.; de los Santos, E.L.; Yeong, M.; Cruz-Morales, P.; Abubucker, S.; et al. A Computational Framework for Systematic Exploration of Biosynthetic Diversity from Large-Scale Genomic Data. *Nat Chem Biol* **2018**, 16(1):60-68. doi: 10.1038/s41589-019-0400-9
37. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* **2003**, *13*, 2498–2504, doi:10.1101/gr.1239303.
38. Abawi, G.Samuel.; Pastor Corrales, M.A.; Centro Internacional de Agricultura Tropical. *Root Rots of Beans in Latin America and Africa: Diagnosis, Research Methodologies, and Management Strategies*; Centro Internacional de Agricultura Tropical, 1990; ISBN 958918314X.
39. Fenta, B.A.; Beebe, S.E.; Kunert, K.J.; Burridge, J.D.; Barlow, K.M.; Lynch, J.P.; Foyer, C.H. Field phenotyping of soybean roots for drought stress tolerance. *Agronomy* **2014**, *4*, 418–435, doi:10.3390/agronomy4030418.
40. Glick, B.R. Plant Growth-Promoting Bacteria: Mechanisms and Applications. *Scientifica (Cairo)* **2012**, *2012*, 963401, doi:10.6064/2012/963401.
41. Patten, C.L.; Glick, B.R. Bacterial Biosynthesis of Indole-3-Acetic Acid. *Can J Microbiol* **1996**, *42*, 207–220, doi:10.1139/m96-032.
42. Ehrlich, H.L. Geomicrobiology: Its Significance for Geology. *Earth Sci Rev* **1998**, *45*, 45–60, doi: https://doi.org/10.1016/S0012-8252(98)00034-8.
43. Hayat, R.; Ali, S.; Amara, U.; Khalid, R.; Ahmed, I. Soil beneficial bacteria and their role in plant growth promotion: a Review. *Ann Microbiol* **2010**, *60*, 579–598, doi:10.1007/s13213-010-0117-1.
44. Yang, X.; Post, W.M.; Thornton, P.E.; Jain, A. The distribution of soil phosphorus for global biogeochemical modeling. *Biogeosciences* **2013**, *10*, 2525–2537, doi:10.5194/bg-10-2525-2013.
45. Kuklinsky-Sobral, J.; Araújo, W.L.; Mendes, R.; Geraldi, I.O.; Pizzirani-Kleiner, A.A.; Azevedo, J.L. Isolation and characterization of soybean-associated bacteria and their potential for plant growth Promotion. *Environ Microbiol* **2004**, *6*, 1244–1251, doi: https://doi.org/10.1111/j.1462-2920.2004.00658.x.
46. Moretti, L.G.; Crusciol, C.A.C.; Kuramae, E.E.; Bossolani, J.W.; Moreira, A.; Costa, N.R.; Alves, C.J.; Pascoaloto, I.M.; Rondina, A.B.L.; Hungria, M. Effects of growth-promoting bacteria on soybean root activity, plant development, and yield. *Agron J* **2020**, *112*, 418–428, doi: https://doi.org/10.1002/agj2.20010.
47. Schmidt, J.; Messmer, M.; Wilbois, K.-P. Beneficial microorganisms for soybean (*Glycine Max* (L.) Merr), with a focus on low root-zone temperatures. *Plant Soil* **2015**, *397*, 411–445.
48. Sait, M.; Hugenholtz, P.; Janssen, P.H. Cultivation of globally distributed soil bacteria from phylogenetic lineages previously only detected in cultivation-independent surveys. *Environ Microbiol* **2002**, *4*, 654–666, doi:10.1046/j.1462-2920.2002.00352. x.

49. Madhaiyan, M.; Poonguzhali, S.; Lee, J.-S.; Lee, K.-C.; Saravanan, V.S.; Santhanakrishnan, P. *Microbacterium Azadirachtae* Sp. Nov., a Plant-growth-promoting *Actinobacterium* isolated from the rhizoplane of neem seedlings. *Int J Syst Evol Microbiol* **2010**, *60*, 1687–1692, doi:10.1099/ijs.0.015800-0.
50. Dourado, M.N.; Bogas, A.C.; Pomini, A.M.; Andreote, F.D.; Quecine, M.C.; Marsaioli, A.J.; Araújo, W.L. *Methylobacterium*-plant interaction genes regulated by plant exudate and quorum sensing molecules. *Brazilian Journal of Microbiology* **2013**, *44*, 1331–1339, doi:10.1590/S1517-83822013000400044.
51. Grossi, C.E.M.; Fantino, E.; Serral, F.; Zawoznik, M.S.; Fernandez Do Porto, D.A.; Ulloa, R.M. *Methylobacterium* Sp. 2A Is a plant growth-promoting rhizobacteria that has the potential to improve potato crop yield under adverse conditions. *Front Plant Sci* **2020**, *11*.
52. Liu, Z.-P.; Wang, B.-J.; Liu, Y.-H.; Liu, S.-J. *Novosphingobium Taihuense* Sp. nov., a novel aromatic-compound-degrading bacterium isolated from Taihu Lake, China. *Int J Syst Evol Microbiol* **2005**, *55*, 1229–1232, doi:10.1099/ijs.0.63468-0.
53. Sohn, J.H.; Kwon, K.K.; Kang, J.-H.; Jung, H.-B.; Kim, S.-J. *Novosphingobium Pentaromativorans* Sp. Nov., a high-molecular-mass polycyclic aromatic hydrocarbon-degrading bacterium isolated from estuarine sediment. *Int J Syst Evol Microbiol* **2004**, *54*, 1483–1487, doi:10.1099/ijs.0.02945-0.
54. Westerberg, K.; Elväng, A.M.; Stackebrandt, E.; Jansson, J.K. *Arthrobacter Chlorophenolicus* sp. nov., a new species capable of degrading high concentrations of 4-chlorophenol. *Int J Syst Evol Microbiol* **2000**, *50*, 2083–2092, doi:10.1099/00207713-50-6-2083.
55. Salwan, R.; Sharma, V. Molecular and biotechnological aspects of secondary metabolites in actinobacteria. *Microbiol Res* **2020**, *231*, 126374, doi: 10.1016/J.MICRES.2019.126374.
56. Bruto, M.; Prigent-Combaret, C.; Muller, D.; Moënne-Loccoz, Y. Analysis of genes contributing to plant-beneficial functions in plant growth-promoting rhizobacteria and related Proteobacteria. *Sci Rep* **2014**, *4*, 6261, doi:10.1038/srep06261.
57. Walker, V.; Bruto, M.; Bellvert, F.; Bally, R.; Muller, D.; Prigent-Combaret, C.; Moënne-Loccoz, Y.; Comte, G. Unexpected phytostimulatory behavior for *Escherichia Coli* and *Agrobacterium Tumefaciens* model strains. *Molecular Plant-Microbe Interactions®* **2013**, *26*, 495–502, doi:10.1094/MPMI-12-12-0298-R.
58. Imam, J.; Singh, P.K.; Shukla, P. Plant microbe interactions in post genomic era: perspectives and applications. *Front Microbiol* **2016**, *7*, 1488, doi:10.3389/fmicb.2016.01488.
59. Chauhan, H.; Bagyaraj, D.J.; Selvakumar, G.; Sundaram, S.P. Novel plant growth promoting rhizobacteria—prospects and potential. *Applied Soil Ecology* **2015**, *95*, 38–53, doi: 10.1016/J.APSOIL.2015.05.011.