



Implications of Domestication in *Theobroma cacao* L. Seed-Borne Microbial Endophytes Diversity

Deisy Lisseth Toloza-Moreno¹ · Roxana Yockteng^{1,2} · José Ives Pérez-Zuñiga³ · Cristian Salinas-Castillo⁴ · Alejandro Caro-Quintero⁴

Received: 23 April 2024 / Accepted: 5 July 2024
© The Author(s) 2024

Abstract

The study of plant–microbe interactions is a rapidly growing research field, with increasing attention to the role of seed-borne microbial endophytes in protecting the plant during its development from abiotic and biotic stresses. Recent evidence suggests that seed microbiota is crucial in establishing the plant microbial community, affecting its composition and structure, and influencing plant physiology and ecology. For *Theobroma cacao* L., the diversity and composition of vertically transmitted microbes have yet to be addressed in detail. We explored the composition and diversity of seed-borne endophytes in cacao pods of commercial genotypes (ICS95, IMC67), recently liberated genotypes from AGROSAVIA (TCS01, TCS19), and landraces from Tumaco (Colombia) (AC9, ROS1, ROS2), to evaluate microbial vertical transmission and establishment in various tissues during plant development. We observed a higher abundance of *Pseudomonas* and *Pantoea* genera in the landraces and AGROSAVIA genotypes, while the commercial genotypes presented a higher number of bacteria species but in low abundance. In addition, all the genotypes and plant tissues showed a high percentage of fungi of the genus *Penicillium*. These results indicate that domestication in cacao has increased bacterial endophyte diversity but has reduced their abundance. We isolated some of these seed-borne endophytes to evaluate their potential as growth promoters and found that *Bacillus*, *Pantoea*, and *Pseudomonas* strains presented high production of indole acetic acid and ACC deaminase activity. Our results suggest that cacao domestication could lead to the loss of essential bacteria for seedling establishment and development. This study improves our understanding of the relationship and interaction between perennial plants and seed-borne microbiota.

Keywords Seed-borne microbial endophytes · *Theobroma cacao* · ACC deaminase activity · Indole acetic acid

Introduction

Beneficial microorganisms colonize most plants on the planet, and the composition of these microbial communities is influenced by multiple factors, which vary according to the host species, tissues, plant phenology, geographic location, and environmental conditions [1]. In many cases, these microbial communities play important roles that affect plant health, for example, protection against pathogens, nutrient fixation and translocation, drought resistance, and resistance to herbivory [2]. This evidence shows that the relationship between microorganisms and plants is fundamental in plant biology, and it is essential to understand, for instance, seed germination, plant establishment, and adaptation to stressful conditions [3, 4].

Few studies have focused on determining the effects that influence the formation of a microbial community in a host

✉ Alejandro Caro-Quintero
acarq@unal.edu.co

¹ Centro de Investigación Tibaitatá, Corporación Colombiana de Investigación Agropecuaria (AGROSAVIA), Km 14 Vía Mosquera, Cundinamarca, Colombia

² Institut de Systématique, Evolution, Biodiversité-UMR-CNRS 7205, Muséum National d'Histoire Naturelle, Paris, France

³ Centro de Investigación Palmira, Corporación Colombiana de Investigación Agropecuaria (AGROSAVIA), Sede Popayán, Popayán, Cauca, Colombia

⁴ Departamento de Biología, Facultad de Ciencias, Max Planck Tandem Group in Holobionts, Universidad Nacional de Colombia, Bogotá, Colombia

and how these factors might modify the host-microbe interaction [5]. Among these effects, some can occur by changes in the environmental conditions originated by microorganisms already established in a host, which affect the establishment of new species in the community [6, 7]. This phenomenon is known as priority effects [8]. Within the seed, these priority effects may occur due to the presence of primary symbionts that may affect the ability to form secondary symbiotic relationships in maturing hosts [9] and sometimes exclude other microorganisms due to competition [10]. Therefore, this community of microorganisms might have a role in the composition, assembly, establishment, and structure of the mature plant's microbial community and, consequently, in the plant physiology and ecology [4, 7, 11, 12].

The microbial populations can colonize the external plant surfaces (epiphytic microbiota) or live in the internal plant tissues (endophytic microbiota). Plants would acquire endophytic microorganisms horizontally from the soil and the surroundings or through vertical transmission (seed-borne) [4, 13]. Even though many of these microorganisms' roles are still unknown, the studies suggest that they are not pathogenic and are often beneficial for the host plant [6, 11]. In seeds, fewer studies have evaluated the diversity of endophytes, the mechanisms that structure microbial communities within these [14], and their vertical transmission or establishment in different tissues during plant development [13]. Bacterial endophytes can inhabit various parts of the seed, such as the embryo [15], and after germination, they can colonize the developing tissues [4, 16], such as stems and leaves, and eventually the entire plant endosphere [15]. Certain endophytes can also colonize flowers and seeds and may be transmitted vertically to the next generation [14, 17].

In addition, seed-borne microbial endophytes are particularly interesting as they enhance seed germination and plant growth promotion, so they are important in improving crop productivity and resilience [18–20]. Thus, they can be used as bioproducts in phytoremediation and sustainable agriculture [21]. Also, they might evolve mechanisms that protect the germinating seeds and seedlings from foraging, drought, and fungal seed-borne and soil pathogens [7]. One of the most fascinating interactions between these microorganisms and their host plants is the physiological control of plant growth by bacteria, in which bacteria produce plant hormones or hormone-related signaling molecules that modulate the plant stress response [22]. From these, indole-3-acetic acid (IAA) is an essential hormone in plant-bacterial interactions [12], which improves plant growth under stress conditions [23]. IAA is a plant auxin involved in various plant physiological processes, such as regulation of plant development, induction of plant defense systems, and formation of lateral and adventitious roots [24]. The endophytic bacteria use the tryptophan secreted by plants

in the rhizosphere as a precursor for IAA biosynthesis [24, 25]. They also produce the enzyme 1-aminocyclopropane-1-carboxylate (ACC) deaminase that hydrolyzes ACC, an ethylene precursor regulating its production and metabolism. Ethylene controls the response of plants to abiotic and biotic stresses inhibiting root development [12]. ACC hydrolysis decreases plant stress by improving plant growth under stress conditions and helps to produce deep roots for better plant water absorption [25].

Recent data on multiple cultivated species indicate that the seed microbiota would constitute most cultivated plants' microbial populations [17]. The study of seed-borne endophytes has focused on short-cycle crops (*e.g.*, maize, soybean, rice) where the effect of endophytes on establishment, growth, and resistance to stresses has been evaluated [26–29]. For tropical perennial plants, studies on the association and interaction of plants and seed-borne microbiota are still incipient. A study in wild perennial plants showed that the composition of microbiota changes during the lifetime [30]. For instance, a study in *Quercus robur* reported that under controlled conditions, a significant fraction of the seed microbiome was transferred to the root tissue of developing seedlings [31].

Microbial communities also have an essential role in another perennial species as *Theobroma cacao* L. (cacao), which produces the main ingredient of chocolate. It has been reported that these microbial communities affect the quality of the fermented cacao bean and, eventually, the quality of chocolate, and thus, several studies have focused on determining their importance in the fermentation process [32–35]. Interestingly, some of the observed microbes during the early fermentation times belong to seed endophytes recurrently found in other plants [17], such as *Tatumella* and *Pantoea* bacteria [33, 36]. Studies of microbial endophytes associated with cacao plant tissues and their role are scarce. For instance, Simarmata et al. [37] identified the endophytic bacteria *Brevibacillus brevis* and *Pantoea* sp. as cacao growth promoters, and Hanada et al. [38] found that endophytic fungi of the genera *Trichoderma*, *Pestalotiopsis*, *Curvularia*, *Tolypocladium*, and *Fusarium* showed antagonistic capacity against *Phytophthora palmivora*. Further studies are needed to determine endophytes' acquisition and transmission mechanisms and their possible functions in *T. cacao*.

Here, we present the first study of the composition of the seed-borne endophytic microbial community of *T. cacao*. We germinated seeds under sterile conditions to assess the diversity of the seed-borne microbiota and their establishment in seedling tissues. Bacteria and fungi diversity was identified in different seedling tissues by sequencing molecular markers (16S rRNA and the Internal Transcribed Spacer ITS-5.8S). Additionally, we isolated bacteria and evaluated their plant growth-promoting

capabilities. We analyzed the variability of these communities and the possible effect of domestication by comparing widely used commercial cacao genotypes, AGROSAVIA genotypes, and landraces from Tumaco (Nariño, Colombia). Exploring the diversity of seed-borne endophytes in cacao opens the door not only for elucidating the role of vertical transmission in plant success but also for identifying and modifying the communities involved in cacao bean fermentation and chocolate quality.

Materials and Methods

Germination and Emergence of Cacao Seedlings

Pods of seven genotypes of *T. cacao* were harvested at the AGROSAVIA's El Mira Research Center and cacao-producing farms in Tumaco, Nariño (Colombia), obtained from the commercial genotypes (IMC67, ICS95), recently liberated genotypes from AGROSAVIA (TCS01, TCS19), and landraces from Tumaco (AC9, ROS1, ROS2). Commercial genotypes were collected in 1930 and have been used in breeding programs since 1960; recently liberated genotypes were introduced to farmers in 2014, and landraces have not undergone any breeding improvement and, therefore, have not been massively propagated.

Disinfection of cacao pods and seeds was performed using the methodology described by Falcão et al. [39] with modifications. Cacao pods were washed with soap, disinfected with 2% iodine solution and 70% ethanol, and rinsed with sterile distilled water (sdw). We opened the pods to extract the seeds, manually removed the mucilage using sterile river sand, and rinsed with sdw. Successive washes were performed under constant agitation using 70% ethanol for 2 min, sdw for 2 min, 2.5% sodium hypochlorite for 7 min, and sdw for 2 min. Final disinfection was done with 2.5% sodium hypochlorite for 7 min in a laminar flow cabinet, and seeds were rinsed three times with sdw for 2 min.

Seeds from each cacao genotype were germinated in a humid chamber at room temperature between five and nine days until radicles reached a length of 1.5–2.0 cm. Thirty-six germinated seeds of each cacao genotype were transferred individually to a sealed 500 mL sterile glass vial with wet sterile river sand and placed in a growth room at 26 ± 2 °C and 60% relative humidity. We recorded the emergence and development of seedlings for five weeks, making weekly observations. At the fifth week, we selected ten seedlings per genotype with stems, leaves, and leaflets properly developed. We extracted them from the substrate in a laminar flow cabinet and stored them at -80 °C for further use.

DNA Extraction and Library Preparation

Root, stem, leaf, and cotyledon tissues from the stored seedlings (ten per genotype) were macerated separately in liquid nitrogen. DNA extraction was performed using the DNeasy® Plant Mini Kit (QIAGEN). The quality and concentration of the DNA were verified by visualization in 1% agarose gels stained with SYBR Safe (Invitrogen Carlsbad, CA, USA) and by quantification in a NanoDrop™ 1000 UV spectrophotometer (NanoDrop Technologies, Wilmington, DE, United States), respectively.

The 16S rRNA (515F – 806R) [40] and ITS-5.8S (ITS3_KY02 – ITS4_KY03) [41] regions were amplified to characterize the endophytic bacterial and fungal communities associated with the plant tissues. DNA library preparations followed the protocol described in Caporaso et al. [40] and Caro-Quintero and Ochman [36]. In the first PCR, the primers used to amplify each region have an additional sequence at the 5' end that serves as a linker sequence for the indexing step and sequencing primer region. In the second PCR, indexes and the i5 and i7 Illumina primers were added to each amplicon [32]. The PCR products obtained in each amplification were purified with Agencourt AMPure XP Beads magnetic beads (Beckman Coulter, Brea, CA, USA) and verified by electrophoresis in 1.5% agarose gels stained with SYBR Safe.

For the first PCR, reaction volumes of 20 and 25 µL were used to amplify 16S rRNA and ITS-5.8S regions, respectively. Each reaction for 16S rRNA was composed of 0.4 µL Phire Hot Start II DNA Polymerase (Thermo Scientific), 4.0 µL Buffer 5X, 0.4 µL of dNTPs (10 mM), 0.5 µL of each primer (10 µM), 2.5 µL DNA, and 11.7 µL of ultrapure distilled water (udw) (Invitrogen Carlsbad, CA, USA). The PCR program consisted of an initial denaturation step at 98 °C for 60 s, followed by 35 cycles: denaturation at 98 °C for 30 s, annealing at 60 °C for 10 s, and extension at 72 °C for 15 s, and a final extension step of 72 °C for 120 s. For ITS-5.8S, each reaction was composed of 0.25 µL Phire Hot Start II DNA Polymerase, 5.0 µL buffer 5X, 0.5 µL dNTPs (10 mM), 1.0 µL of each primer (10 µM), 3.0 µL DNA, and 14.25 µL of udw. The PCR program consisted of an initial denaturation at 98 °C for 30 s, followed by 35 cycles: denaturation at 98 °C for 10 s, annealing at 55 °C for 30 s, and extension at 72 °C for 90 s, and a final extension step of 72 °C for 120 s.

For the second PCR, each region was amplified in a 15.0 µL reaction volume. Each reaction was composed of 0.3 µL Phire Hot Start II DNA Polymerase, 3.0 µL buffer 5X, 0.3 µL dNTPs (10 mM), 0.8 µL of each barcode, 3.5 µL of the amplicon of the first PCR, and 6.3 µL of udw. The PCR program consisted of an initial denaturation step at 98 °C for 30 s, followed by 17 cycles of denaturation: 98 °C for 0.05 s, annealing at 50 °C for 16S rRNA and 55 °C for ITS-5.8S for

30 s, and extension at 72 °C for 60 s, and a final extension of 72 °C for 60 s.

Finally, the purified DNA libraries were quantified in a NanoDrop 1000 Spectrophotometer, adjusted to the same concentration, and pooled in one sample. The quality and concentration of pools were verified by automatized electrophoresis in an Agilent 4200 TapeStation System (Agilent Technologies) and sequenced in an Illumina MiSeq System (Universidad del Bosque, Colombia).

Estimation of the Diversity and Structure of the Microbial Community

We assessed the quality of the reads using FastQC v.0.11.5 [42]. We performed demultiplexing, pair-end assembly, and diversity analysis using QIIME2 [43]. Dada2 [44] was the algorithm used to denoise reads and generate Amplicon Sequence Variants (ASVs). Dada2 used the paired-end sequences as input with a truncation length of 200 bp for forward and 240 bp for reverse reads.

Taxonomic classification of bacteria ASVs was performed using the GreenGenes database [45], v-gg-13-8-99-nb [43, 44]. We compared the number of ASVs associated with mitochondria to the total bacteria ASVs per sample to establish the relative abundance of bacteria in the different genetic genotypes and tissues. The fold change ratio between the two was computed ($\log_2(\text{Bacteria/Mitochondria})$) to normalize bacterial abundance per tissue. For the taxonomic classification of fungi, two approaches were used. First, ASVs were classified using the UNITE database (unite-ver8-taxonomy_99_10.05.2021_dev.fasta) [46]. Then, ASVs were searched using BLAST (blastn) against the Internal transcribed spacer region (ITS) from fungi type and reference material from the NCBI (National Center for Biotechnology Information) to determine the closest taxonomic group associated with each library. The ASVs were assigned to the taxonomic group with the best hit score. When multiple taxonomic groups had the best hit score, the ASVs were assigned to the broader consensus taxonomic group.

For the quantification of alpha and beta diversity, we removed the ASVs corresponding to chloroplast and mitochondria from the representative sequences as well as the feature tables using the program "QIIME taxa" with "filter-seqs" and "filter-table" options.

Rarefaction curves were estimated to determine the minimum read depth that reaches species saturation for most libraries. The 16S rRNA libraries were sub-sampled to the minimum depth of 300 reads, while ITS-5.8S libraries were sub-sampled at 900 reads. To estimate species richness (alpha diversity), we quantified the Shannon, Faith, and Phylogenetic Diversity (PD) indices. We

conducted a pairwise Kruskal–Wallis test to assess the statistical differences between genotypes and tissues.

We determined the similarities and differences between samples (beta diversity) by calculating the weighted and unweighted UniFrac metrics [47]. We conducted a Principal Coordinate ordination Analysis (PCoA) using the UniFrac metrics distance matrix to visualize the ordination of samples and how variables, such as the genotype and tissue type might determine the observed clustering. A permutational multivariate analysis of variance, PERMANOVA [48], was performed using the unweighted and weighted UniFrac distance matrix to establish the significant differences between groups of samples.

Isolation and Molecular Identification of Bacteria from Cacao Plant Tissues

We isolated bacteria from seedlings from all genotypes. We separately macerated each plant tissue (seed, root, stem, leaf, and cotyledon) in a laminar flow cabinet. One gram of each macerate was collected and suspended in 9 mL of 0.85% sterile saline solution. Serial dilutions were prepared and plated in triplicate on Luria Bertani and Nutritive Agar culture media. For seeds, 100 µL of 10^{-2} to 10^{-5} dilutions were plated, while for the other tissues, 100 µL of 10^{-3} to 10^{-5} dilutions were plated. Petri dishes were incubated at 30 °C for 24, 48, and 72 h. After incubation, macroscopic (color, elevation, appearance, consistency, edge of colony) and microscopic (Gram-positive or Gram-negative bacteria, cell shape, and arrangement) descriptions of isolated colonies were made, and the identified morphotypes were individualized to obtain pure cultures. From these plates, colony morphotypes were picked and DNA extraction was done using the ZR Fungal/Bacterial DNA MiniPrep™ Kit. The obtained DNA was quantified in NanoDrop™ 1000 UV spectrophotometer. The 16S rRNA region was amplified using the 27F (5'-AGAGTTTGATCCTGGCTCAG-3') and 1492R (5'-GGT TACCTTGTTACGACTT-3') primers. Each PCR reaction was composed of 0.4 µL Phire Hot Start II DNA Polymerase, 4.0 µL Buffer 5X, 0.4 µL dNTPs (10 mM), 1.2 µL of each primer (10 µM), 3.0 µL DNA, and 9.8 µL of ultrapure distilled water. The PCR program consisted of an initial denaturation step at 98 °C for 60 s, followed by 35 cycles: denaturation at 98 °C for 30 s, annealing at 60 °C for 10 s, and extension at 72 °C for 15 s, and a final extension step of 72 °C for 120 s. The PCR products were verified by electrophoresis in 1.5% agarose gels stained with SYBR Safe and then purified with Exo-CIP™ Rapid PCR Cleanup Kit (New England BioLabs) and sequenced by the Sanger method (Universidad Nacional de Colombia).

Assessing the IAA Production and ACC Activity of Bacterial Isolates

The potential of each morphotype for plant growth promotion was evaluated quantifying their capability, *in vitro*, to produce Indole Acetic Acid (IAA) and 1-aminocyclopropane-1-carboxylic acid (ACC) deaminase activity.

First, to evaluate IAA production, an inoculum of each morphotype was suspended in 1 mL of 0.85% sterile saline, and the absorbance was measured at 600 nm and adjusted to an optical density of 0.2. Then, 100 μ L of this suspension were inoculated by triplicate in 5 mL of a minimal medium M9 (composition $\text{g}\cdot\text{L}^{-1}$: 3 g KH_2PO_4 , 1 g NH_4Cl , 0.5 g NaCl, 0.25 g MgSO_4 , 3 g filtered casamino acids) supplemented with 2.5 mM tryptophan and incubated at 30 °C for 72 h with agitation at 180 rpm in the dark. Then, 1 mL of each culture was centrifuged at 14000 rpm for 5 min, and 100 μ L of the supernatant was taken and added to a microplate with 100 μ L of Salkowski's reagent [49]. The microplate was left in the dark with agitation at room temperature for 30 min. Finally, the absorbance was measured by triplicate at 530 nm and compared with a calibration curve of IAA in concentrations from 1 to 45 ppm (part per million) dissolved in ethanol. The strain *Azospirillum brasiliensis* SP7 was used as a positive control, as it is an ATCC reference strain known to produce IAA.

The ACC deaminase initial screening was done following the protocol reported previously by Habib et al. [50]. Then, the isolates' ability to produce ACC deaminase was evaluated using the M9 medium. This medium was prepared by mixing 20 mL of a sterile solution of salts (8.5 g Na_2HPO_4 , 1.5 g KH_2PO_4 , 0.25 g NaCl dissolved in 100 mL of double-distilled water) with sterile solutions of 0.2 mL 1 M MgSO_4 , 10 μ L of 1 M CaCl_2 , 2 mL of 20% glucose p/v, and distilled water to complete a final volume of 100 mL. The ACC deaminase assay was conducted in a 96-well microplate, one isolate per row. We deposited 120 μ L of M9 medium in each well. We added 15 μ L of 0.1 M MgSO_4 solution to the first four wells of each row and 15 μ L of 3 mM ACC solution to the following four wells of each row. A volume of 15.0 μ L of bacterial inoculum was added in wells two to four and six to eight. Wells one and five were used as blanks, and an additional 15 μ L 1 M MgSO_4 was added to have the same total volume in all wells. The microplates were incubated at 30 °C with continuous agitation of the medium, and the absorbance was measured at a 600 nm wavelength every two h for two days using a BioTek Synergy HTX Reader (Agilent Technologies). The ACC deaminase activity was positive if the growth curve of the isolates with ACC was higher than the observed with 0.1 M MgSO_4 . *Pseudomonas putida* UW4, with recognized ACC activity, was used as a positive control, while a mutant of *P. putida* UW4 *acdS* was used as a

negative control. As a blank, a solution without nitrogen was prepared (0.1 M MgSO_4).

To identify the isolates that produced ACC deaminase, we calculated the difference between the cell growth in the medium with ACC and in the medium with MgSO_4 . With this information, we clustered the isolates using the Manhattan distance and visualized the results as a heatmap using the package heatmap from R.

Results

Seedling Development

The different cacao genotypes' germination, emergence, and development were evaluated under *in vitro* conditions. Over 75.0% of all genotypes' seeds germinated in five to nine days. The germination percentage was higher for IMC67, AC9, and ROS2 with 97.2%, followed by TCS01, TCS19, and ROS1, with 96.9, 88.6, and 77.8%, respectively. ICS95 was the cacao genotype with the lowest germination percentage (75.0%). Meanwhile, seedlings' emergence was higher for TCS01, with 71.0%, followed by ROS2, and AC9, with 65.7 and 62.9%, respectively. Cacao genotypes with the lowest seedling emergence percentages were TCS19, IMC67, ROS1, and ICS95, with 51.6, 51.4, 42.9, and 40.7%, respectively (Table S1).

Library Preparation and Sequencing for 16S rRNA and ITS-5.8S Regions

By sequencing 16S rRNA and ITS-5.8S regions, we examined the microbial community composition for the different cacao genotypes. We obtained successful amplifications in 84.8% of the samples. The success in library construction varied between sequenced regions, genotypes, and tissue. For 16S rRNA, we constructed libraries for 93.2% of the samples, while for ITS-5.8S, the success of generated libraries was for 76.4% of the samples (Fig. S1).

For the 16S rRNA region, 266 libraries were successfully sequenced, with an average sequence depth of 5257.18 (± 2891.29) reads per sample. After performing the quality control, 263 libraries were retained, with 69.6% of the original reads (Table S2). For the ITS-5.8S region, sequences for 221 samples were obtained with an average sequence depth of 5718.12 (± 3123.20) reads per sample. Only 0.6% of the reads were lost after performing quality control, allowing the retention of 99.4% of reads. The sequencing depth for all the libraries ranged between 285 and 15385 reads, with an average of 5686.68 (Table S3).

Genotypic and Tissue-Specific Variations in Bacterial Abundance

Taxonomic classification allowed us to separate the fraction of reads belonging to bacterial 16S rRNA from those belonging to plant organelles (mitochondria and chloroplast). Most reads were classified as mitochondrial or chloroplast 16S rRNA, suggesting that the abundance of bacteria within seeds was low compared to the number of bacteria in other tissues.

A closer examination of the data sets from seedlings showed that the relative abundance of reads classified as bacteria, chloroplast, and mitochondria was similar within tissues (*e.g.*, more similar within roots of all genotypes),

showing that the relative abundance of bacteria is tissue dependent. To compare the taxonomic diversity patterns of bacteria between tissues and their relative abundance, we used the mitochondrial 16S rRNA as an internal marker. This marker allowed us to normalize the abundance of bacteria in each sample. To explore these differences, we calculated the fold change between the number of bacteria and plant mitochondria 16S rRNA reads and evaluated the relative abundance of bacteria in each plant tissue for each cacao genotype (Fig. 1).

The bacterial to mitochondrial proportion of 16S rRNA reads reveals tissue-specific patterns of bacterial abundance. Roots tissue had the highest abundance of bacteria reads compared to mitochondria, in the case of TCS01, bacterial

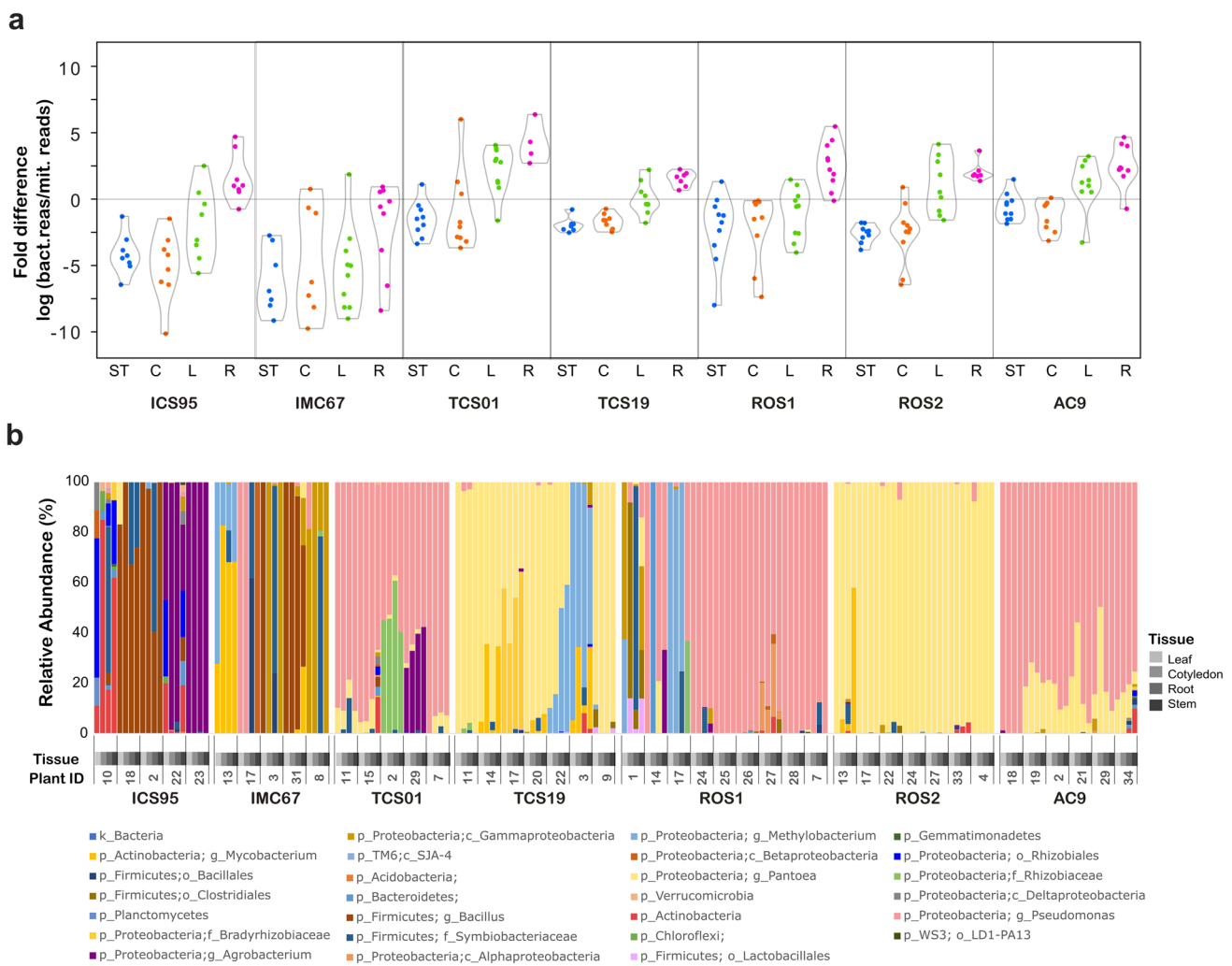


Fig. 1 Relative abundance and taxonomic classification of seed-borne bacterial communities in different *T. cacao* genotypes and tissues. **a** The relative abundance of bacteria to mitochondria is presented per genotype and tissue (ST: Stem, C: Cotyledon, L: Leaf, R: Root). The fold difference represents the log₂ of the ratio between bacterial and mitochondrial reads. Positive values represent tissues with a higher number of reads associated with bacteria than mitochondria, while

negative values represent higher mitochondrial reads. **b** The corresponding taxonomic classification of bacterial reads is shown per genotype, seedling, and tissue. Seedlings replicates are identified by their assigned numbers. The different tissues, leaf, cotyledon, root, and stem are shown in a scale from light to dark gray. The genotypes were ordered left to right, from more to less domesticated. Only seedlings with data for the four analyzed tissues are shown

reads in root tissues were 31.7 times higher than mitochondria, and similar tendencies were found in ROS1, AC9, ICS95, ROS2, and TCS19 genotypes with 11.3, 9.3, 6.7, 4.8 and 3.0 times more bacterial reads. Only IMC67 had a lower number of bacterial to mitochondrial reads with 0.87 times. The leaves tissue was the second with the highest bacterial abundance, in this case, TCS01, ROS2, ROS1 AC9, and TCS19 genotypes had a higher number of bacteria with 7.7, 4.7, 3.7, and 1.5 times the number of mitochondria while ICS95 and IMC67 had lower abundance with 0.98 and 0.41 times. Cotyledon and stem tissues had a lower abundance of bacteria in all genotypes as shown by the negative values of the fold difference (Fig. 1a), exceptions were observed for three seedlings of the TCS01 genotype.

The fold difference analysis also shows differences among genotypes. The AC9, ROS2, ROS1, TCS01, and TCS19 genotypes had a higher average value of bacterial abundance, with 3.6, 2.5, 3.3, 12, and 1.3 times the number of bacterial to mitochondrial reads. This contrasts with the observed in the most domesticated genotypes IMC67 and ICS95 with average values of 0.37 and 0.96. In all evaluated tissues, the IMC67 genotypes had the lower overall bacterial abundance; for instance, tissues such as the IMC67 stem had almost 32 times more mitochondria than bacteria on average (a fivefold difference, Fig. 1a).

Bacterial and Fungal Taxonomic Composition

Bacterial composition was evaluated after excluding 16S rRNA plant mitochondria and chloroplast sequences. We obtained 2,514,150 reads belonging to 240 libraries. These reads were classified into taxonomic groups using the GreenGenes database. Most of the 16S rRNA reads were assigned to the *Proteobacteria* (86.3%) and the *Firmicutes* phyla (11.3%, Fig. 1b). A few other reads were assigned to the phylum *Actinobacteria* (2.2%).

At a finer resolution, we found 87 groups at the genus level in all samples. Of these 87, only 61 had more than 1.0% abundance. These genera were not shared by all samples; for instance, only 33 were in more than one sample. The genus *Pseudomonas* was the most prevalent, being identified in 107 plant tissues of the 280 evaluated, followed by the genus *Pantoea*, which was found in 102 tissues. *Pseudomonas* was dominant in the tissues of AC9, ROS1, and TCS01 genotypes, while *Pantoea* was more prevalent in ROS2 and TCS19. The abundance of these two genera was negatively correlated ($r = -0.47$, $p < 0.05$), indicating possible ecological exclusion. In the case of ICS95 seedlings, *Agrobacterium* and *Bacillus* were predominant. In contrast, in IMC67, we did not observe the prevalence of one specific bacteria genus. Instead, we detected the presence of multiple genera and families, such as *Methylobacterium*, *Mycobacteriaceae*, *Symbiobacteriaceae*, and *Bradyrhizobiaceae* (Fig. 1b).

Some differences were observed between seedlings from the same genotype. For instance, seedling 34 of the AC9 genotype presented a greater diversity of bacteria than the other replicates. In ROS1, three of the seedlings evaluated presented taxonomic groups that were more prevalent than those found in the other replicates. For seedling 17, the genus *Methylobacterium* predominated, except in the root where *Bacillus* and *Pseudomonas* genera were detected. In seedling 14, only the phylum *Bacteroidetes* was presented in the stem, and in seedling 27, a dominant group belonging to the class *Alphaproteobacteria* was observed (Fig. 1b).

For the fungi libraries, we obtained 1,256,757 ITS reads after filtering and trimming, and we used them to evaluate the taxonomic composition. After the QIIME2 analysis, 213 ASVs were obtained of which 203 were retained for fungi classification. The results obtained with the UNITE database were too unspecific; therefore, we used BLAST to attempt a better classification. This analysis showed that the great majority of the fungal reads in cacao tissues belonged to the phylum *Ascomycota* (99.9%), and only a few reads belonged to the phyla *Basidiomycota* (0.02%) and *Mucoromycota* (0.01%). Within the *Ascomycota*, 99.7% of the reads belonged to the subphylum *Pezizomycotina* and few to the subphylum *Saccharomycotina* (0.3%) (Fig. 2).

Most genotypes and tissues had a similar fungal community (Fig. 2). The genus *Penicillium* was the most dominant, with 78.8% of reads (Fig. 2). Few samples had other fungi from the orders *Pleosporales* and *Hypocreales* and subclass *Chaetothyriomycetidae*. Some seedlings presented other dominant taxa than *Penicillium*. For instance, tissues from individuals 34 of AC9 and 18 of ICS95 had more abundance of fungi of the subclass *Chaetothyriomycetidae* (Fig. 2).

Species Richness Analysis

The bacterial and fungal species richness indices were calculated for each sample to determine if there were significant differences between genotypes and tissue types (Tables S4, S5). The ITS species richness analysis did not show significant differences between genotypes or tissues ($p > 0.05$). The analysis of bacterial ASVs (with relative abundance $> 1\%$) showed that the seedlings host a small number of species (on average, four ASVs per tissue). For some seedling tissues, we did not retrieve any bacterial reads. An average of six ASVs were found in the ITS libraries. The ICS95 rhizosphere libraries had a higher number of 16S rRNA ASVs (34 ASVs). To account for the phylogenetic distance among ASVs, we calculated the Faith phylogenetic diversity (PD) index [51]. There were significant differences between the diversity values of AC9 and the commercial genotypes (ICS95 and IMC67) as well as between AC9 and the recently liberated genotypes from AGROSAVIA (TCS01 and TCS19), no significant differences were observed when

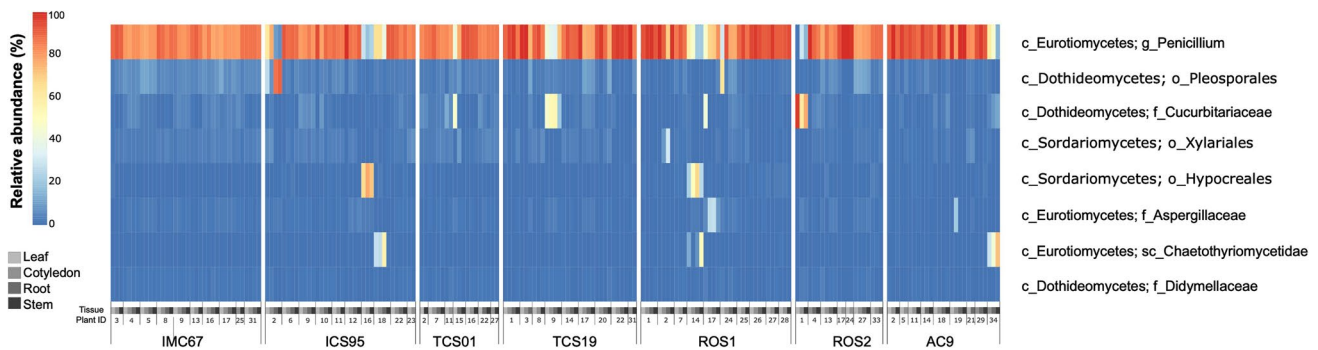


Fig. 2 Relative abundance and taxonomic classification of dominant fungi communities in *T. cacao* commercial genotypes (IMC67 and ICS95), recently liberated genotypes from AGROSAVIA (TCS01

and TCS19), and landraces (ROS1, ROS2, and AC9). The lowest taxonomic classification is shown for each ASVs, the letters denote: g (genus), o (order), f (family), sc (subclass), and c (class)

compared to other landraces. Similarly, the diversity values of ROS2 were significantly different to TCS19, IMC67, TCS01, and ICS95 ($p < 0.05$) (Table S4).

Regarding the tissues, the different diversity indices (Faith-PD and observed features) showed that the number of observed species and phylogenetic diversity was, on average, higher in roots than in the other tissues. Statistical evaluation of Faith-PD distances shows that there are significant differences ($p < 0.05$) between root tissue with cotyledon and leaf tissues (Table S5).

Interestingly, when the frequency of seedling emergence is mapped against the phylogenetic diversity index, the most diverse genotypes had a lower seedling emergence, except for ROS1, in which seed emergence was low (42.9%). For instance, the genotypes with the highest emergence values, TCS01 (71.0%), ROS2 (65.7%), and AC9 (62.9%), had the least diverse endophytic communities of bacteria, as

reflected in Faith distance distributions. In contrast, IMC67 (51.4%) and TCS19 (51.6%), with lower emergence values, had more diverse communities. Similarly, ICS95 (40.7%) was the genotype with the most diverse bacterial community and had the lowest emergence values (Table S1). A significant negative correlation (Pearson's r) is observed between Faith PD and seed germination ($r = -0.86$, $p < 0.05$) and emergence ($r = -0.82$, $p < 0.05$) when the values for genotype ROS1 are excluded from the analysis.

Bacterial and Fungal Composition of Different Genotypes and their Tissues

A Principal Coordinate Ordination Analysis (PCoA) of the seed-borne bacteria was constructed with a matrix of weighted UniFrac distances between samples (Fig. 3). The PCoA was used to visualize the similarities or differences

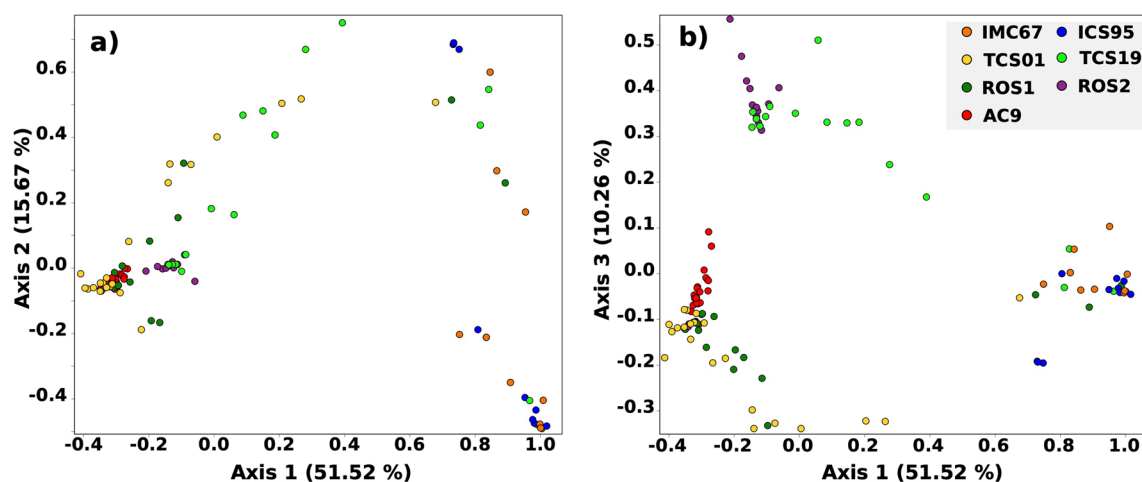


Fig. 3 Principal Coordinate ordination Analysis of the seed-borne bacterial community present in different tissues of *T. cacao* genotypes. The IMC67 and ICS95 are the commercial genotypes, TCS01 and TCS19 are recently liberated genotypes from AGROSAVIA, and

ROS1, ROS2, and AC9 are landraces. The PCoA was constructed using a weighted UniFrac distance matrix. **a** Axes 1 and 2 explain 67.2% of the variability. **b** Axes 1 and 3 explain 61.8%

between the endophytic bacterial communities and to evaluate if the distribution was structured based on the genotype or plant tissues. The first, second, and third axes explained 51.5, 15.7, and 10.3% of the variability, respectively. When the samples were differentiated by genotype, three clustering groups were observed one composed of TCS01, AC9, and ROS1, a second of ROS2 and TCS19, and a third consisting primarily of IMC67 and ICS95 (Fig. 3). In contrast, when samples were analyzed by tissue, there was no clear clustering (data not shown).

We conducted a PERMANOVA analysis using the weighted and unweighted UniFrac distances to establish the significance of the genotype clustering (Tables S6, S8). Read subsampling was set to 500 reads because, at this depth, the species diversity was saturated based on the rarefaction curves. When analyzing the samples based on the weighted UniFrac distance, microbial communities were significantly different between genotypes except for commercial genotypes IMC67 and ICS95 ($p < 0.05$, Tables S6, S8). Interestingly, the weighted UniFrac average distance, having AC9 as a reference, showed a clear pattern where less domesticated genotypes had a shorter distance between them (AC9, ROS1, and ROS2), slightly higher distance with recently liberated ones (TCS01 and TCS19), and the most considerable distance with the commercial genotypes (IMC67 and ICS95, Fig. S2). Moreover, the comparison of plant tissue distances showed a significant difference between root and leaves, and between leaves and cotyledon ($p < 0.05$, Table S7). The comparison of tissue distances showed a significant difference between roots and other tissues indicating that the most differentiated endophytic community is in roots ($p < 0.05$, Table S9).

The Principal Coordinate Ordination Analysis (PCoA) of the fungi community was constructed with a matrix of weighted UniFrac distances and no clustering based on tissues or genotypes was observed (data not shown). Statistical analysis of the fungal community diversity between tissues showed no significant differences. Evaluation of genotypes showed few differences, the IMC67 genotype was significantly different from the others based on the unweighted UniFrac ($p < 0.05$, Table S10) and ROS1 based on the weighted UniFrac ($p < 0.05$) except with AC9 and ICS95 (Table S11).

Isolation of Endophytic Bacteria and Evaluation of their Plant Growth-Promoting Capabilities

Bacteria endophytes were isolated from seeds and seedlings to determine their ability to promote plant growth. A total of 35 morphotypes were isolated from seedlings; 45.7% of the bacteria were isolated from the TCS01 genotype, 20.0% from the AC9, 11.4% from the TCS19, and 22.9% from the IMC67 genotype. Only two endophytic morphotypes were

isolated directly from seeds of the AC9 genotype. We identified 33 of the isolated bacteria based on their 16S rRNA sequence. Most bacteria belonged to the *Pseudomonas* (42.4%) and *Bacillus* (30.3%) genera, while others belonged to *Stenotrophomonas* (12.1%), *Pantoea* (9.1%), and 6.1% to other genera (Table S12).

The evaluation of the potential for plant growth promotion of the bacterial isolates was assessed to determine the 1-aminocyclopropane-1-carboxylic acid (ACC) deaminase activity and indole acetic acid (IAA) production. Twelve morphotypes (P2501, P1906, P1909, P2510, AC902, P1901, P1504, P1610, P2509, P1101, P2102, and P1103) were found to have ACC deaminase enzyme activity with average growth curves in ACC deaminase presence above the growth curve of the MgSO_4 control treatment (Fig. 4a). A spectrophotometric calibration curve was done to estimate the IAA concentration. This calibration curve showed an r^2 value of 0.99, indicating the method's reliability for estimating the production of IAA of the different morphotypes (Table S13). Among the evaluated morphotypes, *P. agglomerans* P1901, *Pseudomonas* sp. P2502 and strain P2104 (unidentified) produced more than 10.0 ppm of IAA, outperforming the control strain *Azospirillum brasiliensis* SP7, which generated 9.48 ppm after 72 h of growth (Fig. 4b). Furthermore, 11 isolates displayed IAA production levels exceeding 7.0 ppm after 72 h. These findings highlight the potential of these isolates to generate substantial amounts of IAA, suggesting their promising role in plant growth promotion.

Discussion

Most studies on cacao endophytes are focused on vegetative tissues [13, 37] and only a few studies have addressed the diversity of seed-borne endophytes using culture-dependent approaches [52, 53]. Here, we explored vertically transmitted microbes in cacao by studying landraces and commercial cacao genotypes germinated under sterile conditions and we characterized these communities using amplicon sequencing.

The microbial diversity in cacao tissues showed that the roots (in all genotypes) presented the highest diversity and abundance of bacteria. Previous studies in cacao's roots [54], also recorded more bacteria in this tissue, and this pattern has also been observed in other crops, such as sugar cane [55]. This increased bacterial abundance in roots might be related to access to soil nutrients or abundant root exudate, since the root apical meristem usually has a significant concentration of exudates, providing energy and nutrient sources that promote bacterial cell division [55]. The second tissue with the highest bacteria abundance was the leaf tissue, which is to be expected, as they are among the most metabolically active tissues due to processes such as

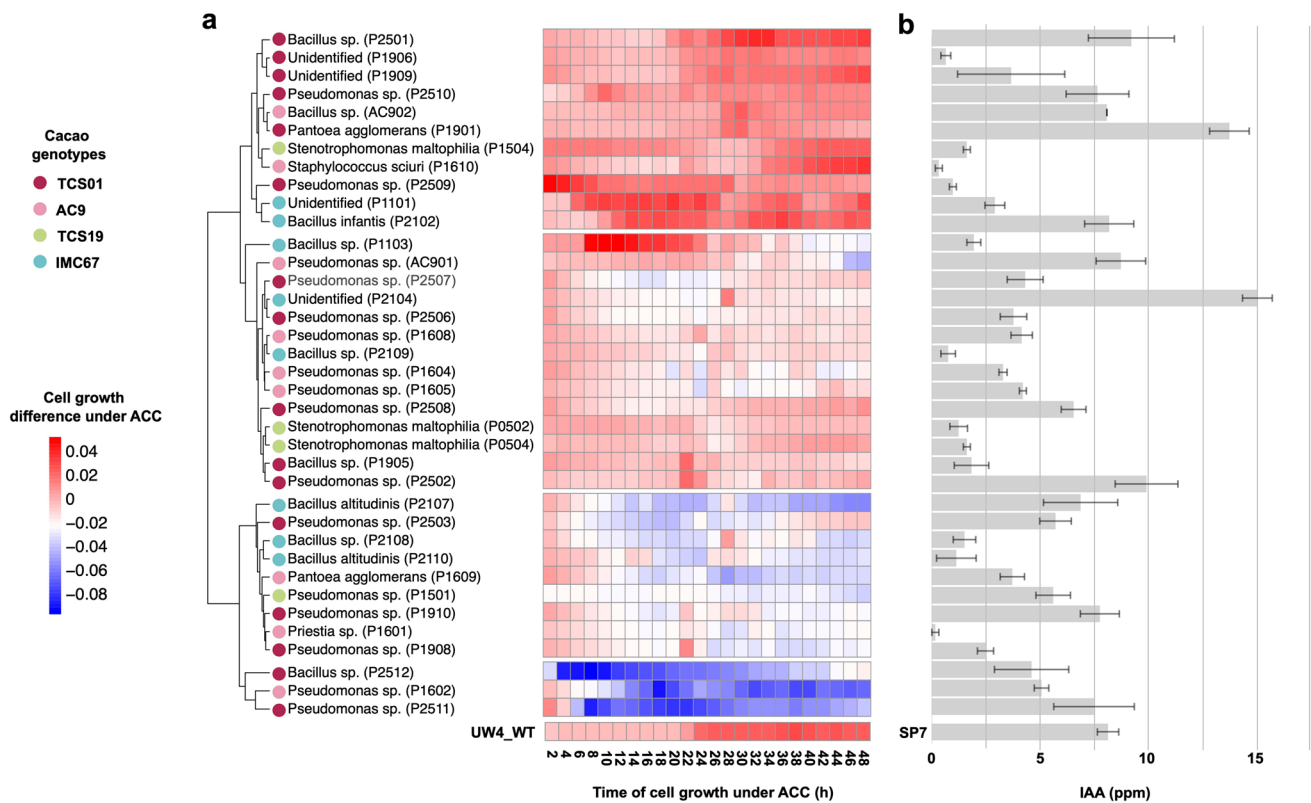


Fig. 4 Evaluation of the plant growth-promoting capabilities of *T. cacao* seed-borne bacterial isolates. The ACC deaminase and indole acetic acid production (IAA) was evaluated for 37 bacterial isolates. **a** The heatmap represents the cell growth difference in a medium with ACC deaminase compared to a medium with $MgSO_4$. Cell growth was monitored through absorbance (600 nm) every 2 h for 48 h. *Pseudomonas putida* UW4 WT was used as positive control for ACC

deaminase production. The taxonomic classification based on 16S rRNA analysis is shown for each isolate, and the circles' colors show the isolate's genotype origin. The clades to the left of the heatmap cluster the isolates based on their cell growth profile. **b** The average IAA production and the standard deviations are shown for each isolate. *Azospirillum brasiliensis* SP7 was used as a positive control

photosynthesis, evapotranspiration, and second metabolite excretion. In contrast, we found few reads associated with bacteria in seed tissues. One possible reason for this low abundance is that bacteria might be dormant or have little metabolic activity. For instance, a study on *Miscanthus* [56] indicated that endophytic bacteria in the seed are induced to become dormant and stay in low numbers. Once the seed matures and becomes more active upon imbibition, the endophytic bacteria multiply in the seedling tissue.

Using different cacao genotypes allowed us to assess the relationship between endophytic bacterial diversity (abundance and composition) and different domestication levels. Previous works have explored these patterns, finding contrasting results. For instance, some reports have shown that domestication reduces endophytic bacterial diversity in *Agave* plants [57], *Hordeum vulgare* [58], and *Malus domestica* [59], while in plants such as *Lactuca sativa*, *Triticum durum*, and *Phaseolus vulgaris* domestication has increased endophytic bacterial diversity [60, 61]. In our study, the tissues of commercial genotypes IMC67 and ICS95 had higher

bacterial diversity than landraces. At first hand, the cacao pattern seems to fall into the increased diversity scenario; however, when normalized by mitochondria reads, the bacteria abundance is higher in landraces and exceptionally low in commercial genotypes. One possible explanation for this pattern is that crop management practices may affect the endophyte microbiome [4]. For instance, systemic fungicides can affect the endophytic community, which is vertically transmitted through seeds [62]. Plants of the IMC67 genotype, commonly used as rootstock, are treated with periodic applications of specific systemic and contact fungicides (e.g., Copper Oxychloride and Ridomil), which would probably affect the composition and abundance of the microbiome. Another clear difference between commercial genotypes and landraces is the propagation mechanisms. In landraces, propagation occurs as a natural reproduction process. In contrast, in commercial genotypes, at least those used as clones, propagation occurs by grafting, which disrupts the natural process of plant reproduction, which might alter the transmission of endophytes to seeds and fruits.

The dynamics of microbial populations within plant tissues provide insights into the complex interplay of seed-borne endophytes. In the case of landraces and AGROSA-VIA genotypes, most seedlings' tissues are dominated only by ASVs associated with *Pantoea* and *Pseudomonas* genera. These genera are commonly found in plant tissues and are known to coexist [63]. Some species of these genera are considered unharmed or beneficial endophytes, while others are reported as pathogenic. *Pantoea* is recognized as one of the main endophytic genera; it has been isolated from various tissues, such as maize seeds [64, 65] and wheat rhizosphere [66], and some of these isolates have shown to improve nutrient uptake, modulate growth, and stress-related phytohormones, among other benefits [63–65].

Similarly, *Pseudomonas* species can promote plant growth and provide biological control against fungal pathogens [67]. For example, *Ps. chlororaphis* has been found to have antagonistic activity against *Phytophthora palmivora*, the causal agent of black pod disease [67, 68]. Our results indicate that either *Pantoea* or *Pseudomonas* dominate in almost all tissues, and their abundance may vary depending on ecological factors. Ecological competition may occur in cacao seed-borne tissues when one species is abundant, the other is not. In some cases, *P. agglomerans* can coexist with *Ps. syringae* [69] but compete with *Ps. savastanoi* for nutrients and space, inhibiting its growth through antibiosis activity [66].

Furthermore, our study found an inverse relationship between bacterial diversity and the emergence of cacao seedlings of different genotypes, where the higher abundance of *Pseudomonas* and *Pantoea* was associated with the highest seed emergence values, as observed for TCS01, ROS2, and AC9. Meanwhile, genotypes like IMC67, ROS1, and TCS19, with lower values of emergence success, had other taxonomic groups. Interestingly, ICS95, the genotype with the most diverse bacterial community, had the lowest emergence values. In this regard, Chesneau et al. [70] and Barret et al. [71] found a negative relationship between ASVs richness and seedling emergence and fitness in beans, reflected by a high proportion of non-germinated seeds and abnormal development of seedlings.

In contrast to the patterns observed for bacterial diversity, the ITS-5.8S amplicon study showed that the diversity of endophytic fungi in *T. cacao* was unaffected by the different domestication levels. The ITS-5.8S libraries for most seedling tissues were dominated by members of the phylum *Ascomycota*, compared to other phyla such as *Basidiomycota*. The presence of asexual morphs (hyphomycetous and coelomycetous) in *Ascomycota* fungi provides a good ability to colonize the internal tissues of plants, therefore they are reported more often as endophytes than other fungi taxa [72]. The genus *Penicillium* (order *Eurotiales*) was found in all plant tissues and was dominant in most *T. cacao*

seedlings. Species of this fungal genus have been reported as endophytes of various plants, acting as growth promoters, improving disease resistance, and protecting against stress conditions [73, 74].

To elucidate the functional roles of the seed-borne bacterial communities, we isolated bacteria from the seedling tissues of the different evaluated genotypes. The characterization showed that most bacteria isolates have plant growth-promoting capabilities (67.6%). Most isolates corresponded to the *Bacillus* and *Pseudomonas* genera, consistent with the 16S rRNA amplicon characterization. Isolation of *B. subtilis* and *Pseudomonas* has been reported previously in cacao leaves [75], fruits, and seeds [53, 76]. Interestingly, most of our isolates display IAA production and ACC deaminase activity. We identified ten isolates that had an equivalent or higher IAA production and ACC deaminase activity than the reference strains; these were isolated principally from the TCS01 genotype (six isolates), which presented the highest germination of seeds (96.9%) and the highest emergence of seedlings (71.0%) under in vitro conditions. It has been reported that the ACC deaminase-producing bacterium improves seed germination and plant growth in different crops even under stress conditions [23, 25, 77, 78]. In turn, this also coincides with the fact that the TCS19 genotype presented the lowest germination percentage among those evaluated for bacterial isolates, and from which only one endophytic bacteria with this ACC deaminase activity was isolated. In cacao, the production of IAA and ACC deaminase by bacteria might stimulate the root hair and cotyledon cell expansion during seedling development and improve plant growth under stress conditions, as has been observed in other plants [23, 25, 77].

In addition, our approach for the normalization of bacteria relative abundance using organelle 16S rRNA reads facilitated the comparison of microbial abundance across tissues and genotypes. It provided consistent values among replicates and tissues of the same genotypes. Although a large proportion of 16S rRNA reads corresponded to organelle, the results showed that our method is suitable and practical for cases with low microbial complexity as saturation of species diversity was observed. There are alternative techniques that block the amplification of organelle 16S rRNA, thereby increasing the proportion of bacterial reads in the libraries, such as the Peptide Nucleic Acid—PNA primers [17, 79]. However, these techniques also prevent read normalization, which is a trade-off to consider.

Conclusions

The investigation of seed-borne endophytes in cacao is relatively uncharted territory, with limited studies delving into their vertical transmission and establishment across various

plant tissues during development. By exploring the diversity of these endophytes, we gain insights into the microbial communities established by vertical transmission. Our current study digs into this intriguing aspect, examining the composition of the seed-borne endophytic microbial community associated with *Theobroma cacao*. We further scrutinize the variability within these communities and the impact of domestication, drawing comparisons between widely used commercial, recently liberated genotypes, and landraces.

Our findings highlight a complex relationship between crop domestication and endophytic bacterial diversity, showing the intricate dynamics of microbial populations within plant tissues, particularly highlighting the dominance of *Pantoea* and *Pseudomonas* genera in less domesticated genotypes (landraces). These endophytes might play significant roles in growth modulation as we observed a better emergence of cacao seedlings where these two bacterial groups were dominant.

Our results also showed that commercial genotypes had a reduced bacteria abundance, probably influenced by agricultural practices and propagation mechanisms, paving the way for colonizing a more diverse range of taxonomic groups. These observations underscore the importance of incorporating abundance quantification methods, such as qPCR or read normalization using mitochondrial or chloroplast 16S rRNA copies, to fully understand the effect of agricultural practices and domestication on the plant microbiome.

On the other hand, the endophytic fungal diversity in *T. cacao* seedlings is predominantly composed of Ascomycota members, with the genus *Penicillium* being ubiquitous across all plant tissues. Further studies would be necessary to establish if these fungi might play crucial roles in growth promotion or increasing resistance against pathogens, holding promise as biocontrol agents.

In conclusion, our study illuminates the intricate interplay between domestication, crop management, and microbiome diversity. These insights are paramount to agriculture, as they elucidate the profound impact of human-driven practices on crop microbiomes. This knowledge could potentially inform strategies to enhance plant health, productivity, propagation alternatives, and resilience against diseases.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00248-024-02409-9>.

Acknowledgements The authors would like to acknowledge the AGROSAVIA Plant Production laboratory team for their collaboration in the in vitro plant multiplication tests and Danna Araujo for preparing fungal DNA libraries.

Author Contributions DLT-M: conceived, designed, and performed the experimental procedures, analysis of the data, producing the results, authored drafts of the paper, and approved the final draft. RY and AC-Q: conceived and designed the experimental procedures, analysis of the data, production of the results, authored drafts of the paper, and approved the final draft. JIPZ: conceived the experimental procedures,

collected the plant material, wrote the materials and methods, and approved the final draft. CSC: wrote the materials and methods of the biochemical characterization of the strains and performed these experiments.

Funding Open Access funding provided by Colombia Consortium. This project was funded by the joint call for research, development, and innovation projects in the agricultural sector between AGROSAVIA and the Universidad Nacional de Colombia – 2020 (Announcement CI20200014) with internal code 1001663.

Data Availability The bacterial and fungal sequence data generated in this study using MiSeq have been deposited and are available in the NCBI Sequence Read Archive (SRA) under BioProject PRJNA1011283.

Declarations

Competing Interests The authors declare no competing interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Compant SA, Samad A, Faist H, Sessitsch A (2019) A review on the plant microbiome: Ecology, functions, and emerging trends in microbial application. *J Adv Res* 19:29–37. <https://doi.org/10.1016/j.jare.2019.03.004>
2. Doornbos RF, Van Loon LC, Bakker PA (2012) Impact of root exudates and plant defense signaling on bacterial communities in the rhizosphere. A review. *Agron Sustain Dev* 32(1):227–243. <https://doi.org/10.1007/s13593-011-0028-y>
3. Berg G, Raaijmakers JM (2018) Saving seed microbiomes. *ISME J* 12(5):1167–1170. <https://doi.org/10.1038/s41396-017-0028-2>
4. Nelson EB (2018) The seed microbiome: Origins, interactions, and impacts. *Plant Soil* 422(1–2):7–34. <https://doi.org/10.1007/s11104-017-3289-7>
5. Fitzpatrick C, Copeland J, Wang PM, Guttman D, Kotanen P, Johnson MTJ (2018) Assembly and ecological function of the root microbiome across angiosperm plant species. *Proc. Natl Acad Sci U S A* 115(6):E1157–E1165. <https://doi.org/10.1073/pnas.1717617115>
6. White JF, Chang X, Kingsley KL, Zhang Q, Chiaranunt P, Micci A et al (2021) Endophytic bacteria in grass crop growth promotion and biostimulation. *Grass Res* 1(1):1–9. <https://doi.org/10.48130/gr-2021-0005>
7. Bever JD, Mangan SA, Alexander HM (2015) Maintenance of plant species diversity by pathogens. *Annu Rev Ecol Evol Syst* 46:305–325. <https://doi.org/10.1146/annurev-ecolsys-112414-054306>

8. Fukami T (2015) Historical contingency in community assembly: integrating niches, species pools, and priority effects. *Annu Rev Ecol Syst* 46:1–23. <https://doi.org/10.1146/annurev-ecolsys-110411-160340>
9. Ridout ME, Schroeder KL, Hunter SS, Styer J, Newcombe G (2019) Priority effects of wheat seed endophytes on a rhizosphere symbiosis. *Symbiosis* 78(1):19–31. <https://doi.org/10.1007/s13199-019-00606-6>
10. Newcombe G, Harding A, Ridout M, Busby PE (2018) A Hypothetical Bottleneck in the Plant Microbiome. *Front Microbiol* 9:1–10. <https://doi.org/10.3389/fmicb.2018.01645>
11. White JF, Kingsley KL, Zhang Q, Verma R, Obi N, Dvinskikh S et al (2019) Review: Endophytic microbes and their potential applications in crop management. *Pest Manag Sci* 75(10):2558–2565. <https://doi.org/10.1002/ps.5527>
12. Afzal I, Shinwari ZK, Sikandar S, Shahzad S (2019) Plant beneficial endophytic bacteria: Mechanisms, diversity, host range and genetic determinants. *Microbiol Res* 221:36–49. <https://doi.org/10.1016/j.micres.2019.02.001>
13. Shahzad R, Khan AL, Bilal S, Asaf S, Lee IJ (2018) What is there in seeds? Vertically transmitted endophytic resources for sustainable improvement in plant growth. *Front Plant Sci* 9:1–10. <https://doi.org/10.3389/fpls.2018.00024>
14. Samreen T, Naveed M, Nazir M, Asghar H, Khan M, Zahir Z et al (2021) Seed associated bacterial and fungal endophytes: Diversity, life cycle, transmission, and application potential Appl. Soil Ecol 168:104191. <https://doi.org/10.1016/j.apsoil.2021.104191>
15. Kandel SL, Joubert PM, Doty SL (2017) Bacterial endophyte colonization and distribution within plants *Microorganisms* 5(4): 9–11. <https://doi.org/10.3390/microorganisms5040077>
16. Shade A, Jacques MA, Barret M (2017) Ecological patterns of seed microbiome diversity, transmission, and assembly. *Curr Opin Microbiol* 37:15–22. <https://doi.org/10.1016/j.mib.2017.03.010>
17. Johnston-Monje D, Gutiérrez JP, Lopez-Lavalle LAB (2021) Seed-transmitted bacteria and fungi dominate juvenile plant microbiomes. *Front Microbiol* 12:737616. <https://doi.org/10.3389/fmicb.2021.737616>
18. Bintarti AF, Sulesky-Grieb A, Stopnisek N, Shade A (2022) Endophytic microbiome variation among single plant seeds. *Phytobiomes J* 6(1):45–55. <https://doi.org/10.1094/PBIOMES-04-21-0030-R>
19. Rochefort A, Simonin M, Marais C, Guillerme-Erckelboudt AY, Barret M, Sarniguet A (2021) Transmission of seed and soil microbiota to seedling. *mSystems* 6(3):e0044621. <https://doi.org/10.1128/mSystems.00446-21>
20. Santos RM, Desoignies N, Rigobelo EC (2022) The bacterial world inside the plant. *Front Sustain Food Syst* 6:830198. <https://doi.org/10.3389/fsufs.2022.830198>
21. Xavier ML, Bosch AR, Karp SG, Socol CR (2021) Endophytic bacteria: a possible path towards a sustainable agriculture. *Bio-technol Res Innov* 5(2):e2021008. <https://doi.org/10.4322/biori.21050203>
22. Wang YL, Zhang HB (2023) Assembly and function of seed endophytes in response to environmental stress. *J Microbiol Biotechnol* 33(9):1119–1129. <https://doi.org/10.4014/jmb.2303.03004o>
23. Shameer S, Prasad TNKV (2018) Plant growth promoting rhizobacteria for sustainable agricultural practices with special reference to biotic and abiotic stresses. *Plant Growth Regul* 84(3):603–615. <https://doi.org/10.1007/s10725-017-0365-1>
24. Espinosa-Zaragoza S, Sánchez-Cruz R, Sanzón-Gómez D, Escobar-Sandoval M, Yañez-Ocampo G, Morales-Constantino M et al (2021) Identification of endophytic bacteria of seeds from *Cedrela odorata* L. (Meliaceae) with biotechnological characteristics *Acta Biol. Colomb* 26(2):196–206. <https://doi.org/10.15446/abc.v26n2.85325>
25. Grobelak A, Kokot P, Świątek J, Jaskulak M, Rorat A (2018) Bacterial ACC deaminase activity in promoting plant growth on areas contaminated with heavy metals. *J Ecol Eng.* 19(5):150–157. <https://doi.org/10.12911/22998993/89818>
26. Dos Santos LF, Fernandes Souta J, de Paula Soares C, Oliveira da Rocha L, Luiza Carvalho Santos M, Grativol C et al (2021) Insights into the structure and role of seed-borne bacteriome during maize germination. *FEMS Microbiol Ecol* 97(4):fiab024. <https://doi.org/10.1093/femsec/fiab024>
27. Kim J, Roy M, Ahn SH, Shanmugam G, Yang JS, Jung HW et al (2022) Culturable endophytes associated with soybean seeds and their potential for suppressing seed-borne pathogens. *Plant Pathol J* 38(4):313–322. <https://doi.org/10.5423/PPJ.OA.05.2022.0064>
28. Walitang DI, Roy A, Lee Y, Choi G, Jeong B, Jamal AR et al (2023) The Endophytic plant growth promoting methyllobacterium oryzae CBMB20 integrates and persists into the seed-borne endophytic bacterial community of rice. *Agric* 13(2). <https://doi.org/10.3390/agriculture13020355>
29. Kim J, Ahn S, Yang J, Choi S, Jung H, Jeon J (2023) Plant protective and growth promoting effects of seed endophytes in soybean plants. *Plant Pathol J* 39(5):513–521. <https://doi.org/10.5423/PPJ.OA.06.2023.0091>
30. Wagner MR, Lundberg DS, Del Rio TG, Tringe SG, Dangel JL, Mitchell-Olds T (2016) Host genotype and age shape the leaf and root microbiomes of a wild perennial plant. *Nat Commun* 7:2–15. <https://doi.org/10.1038/ncomms12151>
31. Abdelfattah A, Tack AJM, Lobato C, Wassermann B, Berg G (2023) From seed to seed: the role of microbial inheritance in the assembly of the plant microbiome. *Trends Microbiol* 31(4):346–355. <https://doi.org/10.1016/j.tim.2022.10.009>
32. Pacheco-Montealegre ME, Dávila-Mora LL, Botero-Rute LM, Reyes A, Caro-Quintero A (2020) Fine resolution analysis of microbial communities provides insights into the variability of cocoa bean fermentation. *Front Microbiol* 11:1–15. <https://doi.org/10.3389/fmicb.2020.00650>
33. Herrera-Rocha F, Cala MP, Aguirre JL, Rodríguez-López C, Chica MJ, Olarte HH et al (2021) Dissecting fine-flavor cocoa bean fermentation through metabolomics analysis to break down the current metabolic paradigm. *Sci Rep* 11(1):1–18. <https://doi.org/10.1038/s41598-021-01427-8>
34. Mota-Gutierrez J, Ferrocino I, Giordano M, Suarez-Quiroz ML, Gonzalez-Ríos O, Coccolin L (2021) Influence of taxonomic and functional content of microbial communities on the quality of fermented cocoa pulp-bean mass. *Appl Environ Microbiol* 87(14):e0042521. <https://doi.org/10.1128/AEM.00425-21>
35. Tigrero-Vaca J, Maridueña-Zavala MG, Liao HL, Prado-Lince M, Zambrano-Vera CS, Monserrate-Maggi B et al (2022) Microbial diversity and contribution to the formation of volatile compounds during fine-flavor cacao bean fermentation. *Foods* 11(7):915. <https://doi.org/10.3390/foods11070915>
36. Caro-Quintero A, Howard O (2015) Assessing the unseen bacterial diversity in microbial communities. *Genome Biol Evol* 7(12):3416–3425. <https://doi.org/10.1093/gbe/evv234>
37. Simarmata R, Widowati T, Dewi T, Lekatompessy S, Antonius S (2020) Isolation, screening and identification of plant growth-promoting endophytic bacteria from *Theobroma cacao*. *Biosaintifika J Biol Biol Educ* 12(2):155–162. <https://doi.org/10.15294/biosaintifika.v12i2.21280>
38. Hanada RE, Pomella AWV, Costa HS, Bezerra JL, Loguercio LL, Pereira JO (2010) Endophytic fungal diversity in *Theobroma cacao* (cacao) and *T. grandiflorum* (cupuaçu) trees and their potential for growth promotion and biocontrol of black-pod disease. *Fungal Biol* 114(11–12):901–910. <https://doi.org/10.1016/j.funbio.2010.08.006>
39. Falcão L, Silva-Werneck J, Vilarinho B, da Silva J, Pomella A, Marcellino L (2014) Antimicrobial and plant growth-promoting

- properties of the cacao endophyte *Bacillus subtilis* ALB629. *J Appl Microbiol* 116:1584–1592. <https://doi.org/10.1111/jam.12485>
40. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ et al (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci* 108(1):4516–4522. <https://doi.org/10.1073/pnas.1000080107>
 41. Toju H, Tanabe AS, Yamamoto S, Sato H (2012) High-coverage ITS primers for the DNA-based identification of ascomycetes and basidiomycetes in environmental samples. *PLoS ONE* 7:e40863. <https://doi.org/10.1371/journal.pone.0040863>
 42. Andrews S, Krueger F, Segonds-Pichon A, Biggins L, Krueger C et al (2012) FastQC: a quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
 43. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA et al (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 37(8):852–857. <https://doi.org/10.1038/s41587-019-0209-9>. Accessed 5 Jul. 2021
 44. Callahan B, Mcmurdie P, Rosen M, Han A, Johnson A, Holmes S (2016) DADA2: High resolution sample inference from Illumina amplicon data. *Nat Methods* 13(7):581–583. <https://doi.org/10.1038/nmeth.3869>
 45. McDonald D, Price M, Goodrich J, Nawrocki E, DeSantis T, Probst A et al (2012) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J* 6(3):610–618. <https://doi.org/10.1038/ismej.2011.139>
 46. Abarenkov K, Allan Z, Timo P, Raivo P, Filipp I, Henrik N, Urmak K (2021) UNITE QIIME release for Fungi. Version 10.05.2021. UNITE Community. <https://doi.org/10.15156/BIO/1264708>
 47. Lozupone C, Lladser M, Knights D, Stombaugh J, Knight R (2011) UniFrac: An effective distance metric for microbial community comparison. *ISME J* 5(2):169–172. <https://doi.org/10.1038/ismej.2010.133>
 48. Anderson MJ (2017) Permutational Multivariate Analysis of Variance (PERMANOVA). In: Wiley StatsRef Stat. Ref. Online, Balakrishnan N, Colton T, Everitt B, Piegorsch W, Ruggeri F, Teugels JL (eds) pp 1–15. <https://doi.org/10.1002/9781118445112.stat07841>
 49. De-Bashan L, Antoun H, Bashan Y (2008) Involvement of indole-3-acetic acid produced by the growth-promoting bacterium *Azospirillum* spp. in promoting growth of *Chlorella vulgaris*. *J Phycol* 44(4):938–947. <https://doi.org/10.1111/j.1529-8817.2008.00533.x>
 50. Habib S, Kausar H, Saud H (2016) Plant growth-promoting rhizobacteria enhance salinity stress tolerance in okra through ROS-scavenging enzymes. *Biomed Res Int* 2016:1–10. <https://doi.org/10.1155/2016/6284547>
 51. Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biol Conserv* 61(1):1–10. [https://doi.org/10.1016/0006-3207\(92\)91201-3](https://doi.org/10.1016/0006-3207(92)91201-3)
 52. Tchinda RA, Boudjeko T, Simao-Beaunoir AM, Lerat S, Tsala É, Monga E, Beaulieu C (2016) Morphological, physiological, and taxonomic characterization of actinobacterial isolates living as endophytes of cacao pods and cacao seeds. *Microbes Environ* 31(1):56–62. <https://doi.org/10.1264/jsme2.ME15146>
 53. Florez A (2021) What is inside of the cacao seed we eat? Approach to mining the cacao microbiome *Theobroma cacao*. McGill University, Canada
 54. Alves-Júnior M, de Sousa FO, Silva TF, Albino UB, Garcia MG, Moreira SMC et al (2021) Functional and morphological analysis of isolates of phylloplane and rhizoplane endophytic bacteria interacting in different cocoa production systems in the Amazon. *Curr Res Microb Sci* 2:100039. <https://doi.org/10.1016/j.crmicr.2021.100039>
 55. Silva M, Freire F, Andrade M, Kuklinsky-Sobral J, Paes da Costa D, Lira-Cadete L (2012) Isolamento e prospecção de bactérias endofíticas e epifíticas na cana-de-açúcar em áreas com e sem cupinicida. *R Bras Ci Solo* 36(4):1113–1121. <https://doi.org/10.1590/S0100-06832012000400006>
 56. Cope-Selby N, Cookson A, Squance M, Donnison I, Flavell R, Farrar K (2017) Endophytic bacteria in *Miscanthus* seed implication for germination, vertical inheritance of endophytes, plant evolution and breeding. *GCB Bioenergy* 9:57–77. <https://doi.org/10.1111/gcbb.12364>
 57. Coleman-Derr D, Desgarennes D, Fonseca-Garcia C, Gross S, Clingenpeel S, Woyke T et al (2016) Plant compartment and biogeography affect microbiome composition in cultivated and native *Agave* species. *New Phytol* 209:798–811. <https://doi.org/10.1111/nph.13697>
 58. Bulgarelli D, Garrido-Oter R, Münch Philipp C, Weiman A, Dröge J, Pan Y et al (2015) Structure and function of the bacterial root microbiota in wild and domesticated barley. *Cell Host Microbe* 17(3):392–403. <https://doi.org/10.1016/j.chom.2015.01.011>
 59. Abdelfattah A, Tack AJM, Wasserman B, Liu J, Berg G, Norelli J et al (2022) Evidence for host–microbiome co-evolution in apple. *New Phytol* 234(6):2088–2100. <https://doi.org/10.1111/nph.17820>
 60. Cardinale M, Grube M, Erlacher A, Quehenberger J, Berg G (2015) Bacterial networks and co-occurrence relationships in the lettuce root microbiota. *Env Microbiol* 17(1):239–252. <https://doi.org/10.1111/1462-2920.12686>
 61. Abdullaeva Y, Ambika B, Honermeier B, Schnell S, Cardinale M (2021) Domestication affects the composition, diversity, and co-occurrence of the cereal seed microbiota. *J Adv Res* 31:75–86. <https://doi.org/10.1016/j.jare.2020.12.008>
 62. Gutierrez A, Grillo MA (2022) Effects of domestication on plant – microbiome interactions special issue – mini review. *Plan Cell Physiol* 00(00):1–13
 63. Hosni T, Moretti C, Devescovi G, Suarez-Moreno ZR, Fatmi M, Guarnaccia C et al (2011) Sharing of quorum-sensing signals and role of interspecies communities in a bacterial plant disease. *ISME J* 5:1857–1870. <https://doi.org/10.1038/ismej.2011.65>
 64. Sheibani-Tezerji R, Naveed M, Jehl MA, Sessitsch A, Rattei T, Mitter B (2015) The genomes of closely related *Pantoea ananatis* maize seed endophytes having different effects on the host plant differ in secretion system genes and mobile genetic elements. *Front Microbiol* 6:1–16. <https://doi.org/10.3389/fmicb.2015.00440>
 65. Gao JL, Xue J, Yan H, Tong S, Khan MS, Wang LW et al (2019) *Pantoea endophytica* sp. nov., novel endophytic bacteria isolated from maize planting in different geographic regions of northern China. *Syst Appl Microbiol* 42(4):488–494. <https://doi.org/10.1016/j.syapm.2019.06.001>
 66. Cherif-Silini H, Thissera B, Bouket AC, Saadaoui N, Silini A, Eshelli M et al (2019) Durum wheat stress tolerance induced by endophyte *Pantoea agglomerans* with genes contributing to plant functions and secondary metabolite arsenal. *Int J Mol Sci* 20(3989):1–36. <https://doi.org/10.3390/ijms20163989>
 67. Miguelez-Sierra Y, Acebo-Guerrero Y (2019) *Pseudomonas chlororaphis* CP07 strain reduces disease severity caused by *Phytophthora palmivora* in genotypes of *Theobroma cacao*. *Eur J Plant Pathol* 155:1133–1143. <https://doi.org/10.1007/s10658-019-01842-8>
 68. Acebo-Guerrero MEJ, Hernández-Rodríguez A, Vandeputte O, Miguélez-Sierra Y, Heydrich-Pérez M, Ye L et al (2015) Characterization of *Pseudomonas chlororaphis* from *Theobroma cacao* L. rhizosphere with antagonistic activity against *Phytophthora*

- palmivora* (Butler). J Appl Microbiol 119:1112–1126. <https://doi.org/10.1111/jam.12910>
69. Sabaratnam S, Beattie GA (2003) Differences between *Pseudomonas syringae* pv *syringae* B728a and *Pantoea agglomerans* BRT98 in Epiphytic and Endophytic Colonization of Leaves. Appl Environ Microbiol. 69(2):1220–1228. <https://doi.org/10.1128/AEM.69.2.1220>
 70. Chesneau G, Laroche B, Prévieux A, Marais C, Briand M, Marolleau B et al (2022) Single Seed Microbiota: assembly and transmission from parent plant to seedling. MBio 13(6). <https://doi.org/10.1128/mbio.01648-22>
 71. Barret M, Briand M, Bonneau S, Prévieux A, Valière S, Bouchez O et al (2015) Emergence shapes the structure of the seed microbiota. Appl Environ Microbiol 81(4):1257–1266. <https://doi.org/10.1128/AEM.03722-14>
 72. Rashmi M, Kushvee JS, Sarma VV (2019) A worldwide list of endophytic fungi with notes on ecology and diversity. Mycosphere 10(1):798–1079. <https://doi.org/10.5943/mycosphere/10/1/19>
 73. Toghueo RMK, Boyom FF (2020) Endophytic *Penicillium* species and their agricultural, biotechnological, and pharmaceutical applications. Biotech 10(3):107. <https://doi.org/10.1007/s13205-020-2081-1>
 74. Zuo Y, Li X, Yang J, Liu J, Zhao L, He X (2021) Fungal endophytic community and diversity associated with desert shrubs driven by plant identity and organ differentiation in extremely arid desert ecosystem. J Fungi 7:578. <https://doi.org/10.3390/jof7070578>
 75. Ishak Z, Mohd MK, Russman AH, Ahmad MJ, Ernie RR, Wan S et al (2016) Plant growth hormones produced by endophytic *Bacillus subtilis* strain LMK-BK isolated from cocoa. Malaysian Cocoa J 9(1):127–133
 76. Vera-Loor M, Bernal-Cabrera A, Vera-Coello D, Leiva-Mora M, Rivero-Aragón A, Morales-Díaz L (2021) Arbol filogenético y diversidad de bacterias endófitas asociadas a *Theobroma cacao* L. en una zona de la Provincia de Esmeraldas, Ecuador. Bioagro 33(3):223–228. <https://doi.org/10.51372/bioagro333.8>
 77. Chen B, Luo S, Wu Y, Ye J, Wang Q, Xu X et al (2017) The effects of the endophytic bacterium *Pseudomonas fluorescens* Sasm05 and IAA on the plant growth and cadmium uptake of *Sedum alfredii* hance. Front Microbiol 8:1–13. <https://doi.org/10.3389/fmicb.2017.02538>
 78. Ahmad J, Sher MA, Hussain MB, Saddique MA (2021) 1-aminocyclopropane 1-carboxylic acid deaminase producing bacteria inoculation for improving the maize seed germination and seedling growth. Agric Sci J 3(1):46–55. <https://doi.org/10.56520/asj.v3i1.65>
 79. Sakai M, Ikenaga M (2013) Application of peptide nucleic acid (PNA)-PCR clamping technique to investigate the community structures of rhizobacteria associated with plant roots. J Microbiol Methods 92(3):281–288. <https://doi.org/10.1016/j.mimet.2012.09.036>