Assessment of the diversity of endophytic, rhizospheric and soil bacteria using a targeted metagenomic approach in Mali

ABSTRACT

Aims: This study explores the diversity of endophytic, rhizospheric and soil bacteria species in agricultural ecosystems in Mali through a targeted metagenomics approach of 16S rRNA.
Study design: To carry out this study, the different steps were the conversion of ab1 files into fasta format, the concatenation of sequenced fragments with sense and antisense primers, the alignment of forward and reverse sequences, the creation of consensus sequence between the two fragments, the use of Blast and phylogenetic analysis.
Location and Duration of the Study: The work was carried outin the African Center of Excellence in Bioinformatics and the LaboREM-Biotech at Faculty of Sciences and Techniques, Bamako, Mali.
Methodology: The different bacterial strains were selected from the microbial collection of the LaboREM-Biotech. The 16S rRNA sequencing was performed using the Sanger method, and obtained data were analyzed using bioinformatic tools for species identification and diversity.
Results: The study identified 35 species divided into four phyla and 6 bacterial classes. Among the 35 species identified, the phyla with the largest proportion are Bacillota and Pseudomonadota; *Bacillus* and *Alcaligenes* being the dominant genus. *Bacillus*was the most predominant genus, and a detailed phylogenetic analysis revealed several distinct clades, highlighting the genetic diversity of the species and the evolutionary relationships between them.
Conclusion: This study provides an overview of microbial diversity in agricultural environments in Mali. Future work could explore whole genome sequencing to better understand the functional capabilities of bacteria and extend the study to other types of crops.
Key words: Bacteria, sequencing, clades, diversity.

Keywords: Metagenomic approach, Diversity, Bacteria, sequencing, clades, Diversity, Mali

1. INTRODUCTION

Plants coexist with complex microbial communities, containing bacterial, archaeal, fungal and protistic taxa [1], [2]. However, the soil microbiota plays a fundamental role in the health and productivity of terrestrial ecosystems by acting on key processes such as plant nutrition, disease resistance and organic matter degradation [3], [4]. Thus, endophytic bacteria have a positive impact that can promote plant growth, ensure their protection against pathogen attacks and abiotic stresses such as drought, soil salinity and pollution [5], [6]. Rhizosphere bacteria can also colonize roots and provide them with services such as increasing nutrient availability and plant uptake capacity and can have a positive impact on yields, plant resilience or help fight pathogens [7]. These bacterial species are important components included in the health of agricultural soils and plants, but their diversity and functioning remain largely unknown.

The 16S rRNA sequence is a molecule that has been used to trace the phylogenetic relationships between bacteria and to identify bacteria from various sources, such as environmental or clinical samples. This technology is currently used in clinical laboratories for routine identifications, especially for slow-growing, unusual or difficult-to-cultivate bacteria, as well as for those poorly differentiated by conventional methods [8]. However, phenotypic methods have some intrinsic problems such as high variability between strains of the same species [9]. Identification based on the 16S rRNA sequence is attractive because

the ribosomal SSU ("Small Subunit") is universally present in bacteria and contains regions with species-specific variability. This allows bacteria to be identified down to the genus or species level by comparison with public databases. The molecular approach has been used for bacterial phylogeny and is also of great importance for species identification [10], [11]. For these reasons, 16S rRNA sequencing is still used as a method for pathogen detection in sterile clinical samples in a normal way, or to detect non-culturable species [12]. 16S rRNA is widely used as a marker for the study of bacterial communities due to its diversity and prevalence [13], [14]. Currently, the analysis of microbial communities mainly relies on 16S rRNA for bacteria and archaea, as well as on the ITS (Internal Transcribed Spacer) region for fungal communities. Targeted metagenomics, which emerged from the pioneering work of Karl Woese and David Lane, is based on PCR amplification and sequencing of an orthologous gene considered universal and sufficiently variable to serve as a phylogenetic marker within the microbial communities studied [15].

This is why the present study is therefore committed to using a targeted metagenomic approach to identify and determine the evolutionary relationships between species of endophyte, rhizosphere and soil bacteria, in order to better understand their diversity and role in agricultural ecosystems in Mali. This study therefore aims to (i) identify the species of endophytic, rhizosphere and telluric bacteria, cultivated and isolated at LaboREM-Biotech; (ii) make a taxonomic classification of the species of microorganisms isolated and identified by type of sampling and (iii) determine the diversity and evolutionary relationship of endophytic, rhizosphere and telluric bacteria.

2. MATERIAL AND METHODS

2.1. Study sites and sample collection sources

2.1.1. Samples and sampling sites

During this study, forty-five (45) samples from various sites and sources were collected. Among these different sampling sites, we can cite: Baguineda, Bozola, Niono, Samanko II, Badalabougou, Bendougouba (Kita), Daoudabougou, Icrisat, Diré and Dilly (Nara).

2.1.2. Sources of sample collection

The origin of the different samples treated was diversified; including rhizospheric soils, non-rhizospheric soils, seeds and plant roots. The samples were then sent for culture, microbiological analyses and biochemical characterizations to the Microbiology and Microbial Biotechnology Research Laboratory (LaboREM-Biotech).

2.2. PCR and sequencing of samples

After culture, microbiological and biochemical analysis; the selected microbial strains were sent to Inqaba Biotech laboratory for DNA extraction, PCR and 16S rRNA gene sequencing. The sequencing reaction was performed using the Dye Terminator Kit V3.1, according to the manufacturer's instructions. The fragments were then sequenced using an Applied Biosystems sequencer.

2.3. Identification of endophytic, rhizosphere and soil bacteria species

The data were analyzed using bioinformatics tools and specialized software for data analysis: (i) The files were converted to fasta format and the sequenced fragments were

concatenated with sense and antisense primers in the same file, the forward and reverse sequences were aligned and consensus sequence between the two fragments was created. The Blast method was used to align the consensus sequence to existing 16s sequences in databases to determine the species before the phylogenetic analysis.After sequencing, the 45 sequences were subjected to a quality control which consists in evaluating the chromatograms of the ab1 files. The metagenomic data were processed in order to obtain detailed information on the composition, taxonomy and diversity of microbial communities in our samples.

The generated ab1 files were converted from ab1 format to fasta format using the "seqret" tool [16].The 'cat' command was used to concatenate the forward (F) and reverse (R) files of the same sample into a single file [17] and the multiple alignment tool 'mafft' (Multiple Alignment using Fast Fourier Transform) was then used to align the sequences in the file_FR.fasta [16]. The aligned sequences were saved in the file_FR.msf.To generate a consensus sequence from the multiple alignment, the tool 'cons' of the EMBOSS (European Molecular Biology Open Software Suite) package was used and the generated files were stored in file_FR.msf [18]. The consensus sequence was saved in file FR.cons.To perform analysis of the bacterial sequences obtained, in order to assign them to specific species, the NCBI 'BLAST' (Basic Local Alignment Search Tool) tool was used [19].To show the proportion of bacterial species according to the phylum and class for each species, sampling origins and sampling sites; the R Studio tool was used [20]. The workflow for targeted metagenomic sequencing analysis is shown in figure 1.

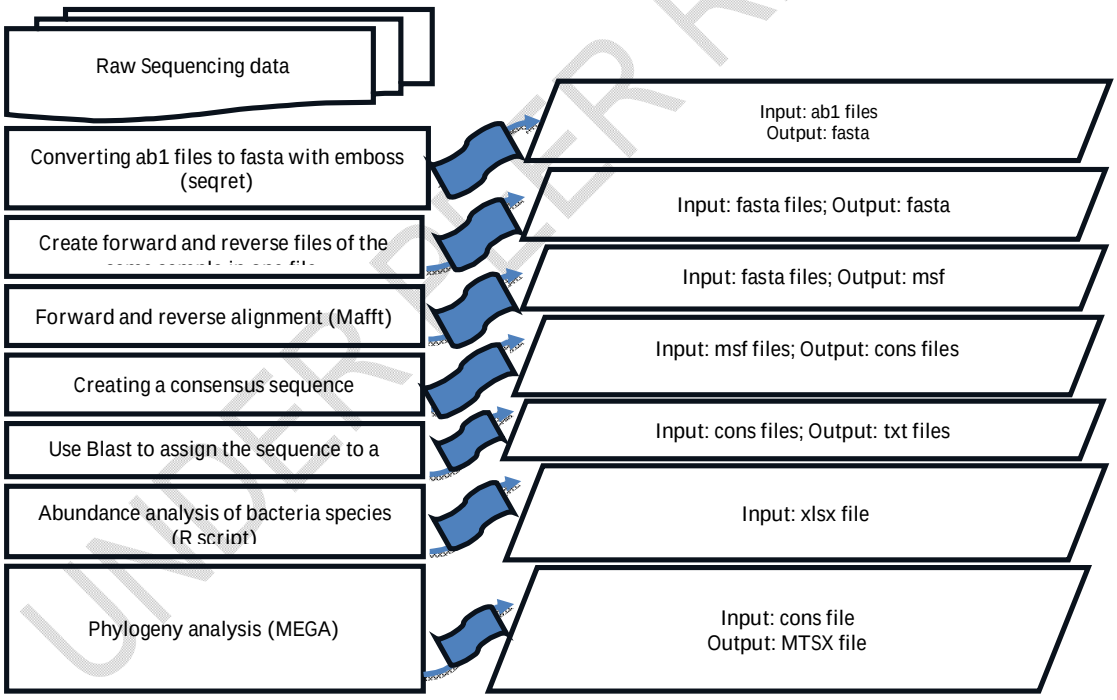

Figure 1. Workflow for targeted metagenomic sequencing analysis [33]

3. RESULTS AND DISCUSSION

3.1. Results

3.1.1 Distribution of bacterial Phyla and Classes among sampling sources

Among the forty-five (45) samples analyzed, thirty-five (35) passed the quality control. The bacterial species were assigned to the thirty-five sequences in the NCBI (National Center for Biotechnology Information) 16S rRNA nucleotide sequence database with identity scores ranging from 80% to 100%. The remaining ten (10) sequences could not be processed by bioinformatics analysis due to their poor sequence quality. The thirty-five (35) species identified belonged to four (04) main phyla (Bacillota [Firmicutes], Pseudomonadota (Proteobacteria), Bacteroidota and Actinomycetota) (Figure 2A) and six (06) main classes (Bacilli, Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria, Flavobacteriia and Actinomycetes) (Figure 2B)

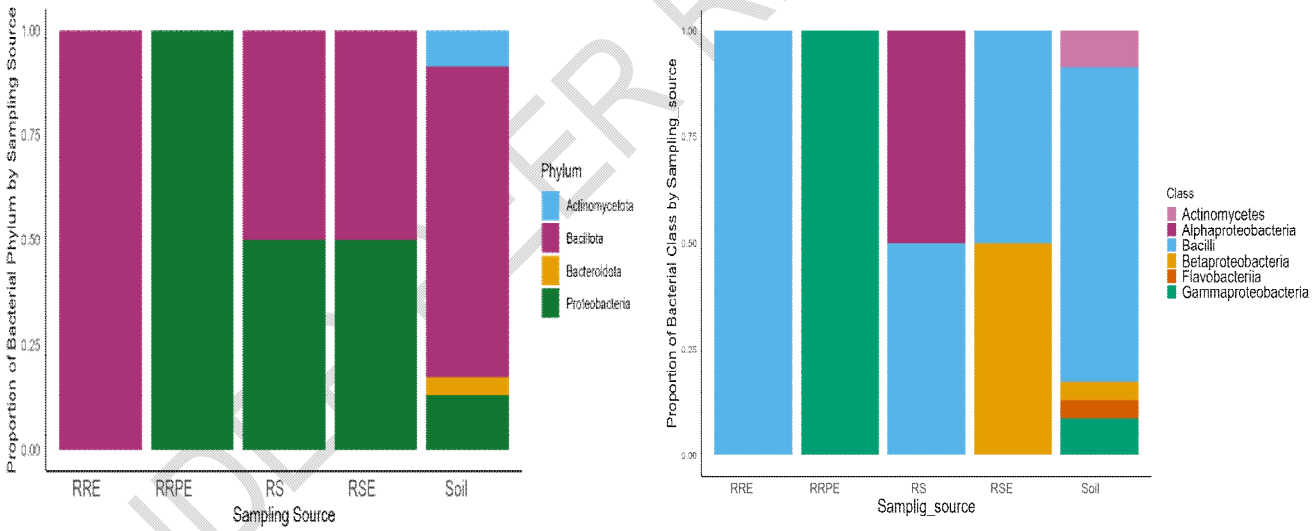

Figure 2: Distribution of bacterial phyla (A) and classes (B)by sampling source.

Bacillota (Firmicutes) followed by Pseudomonadota (Proteobacteria) were respectively the phyla with the largest proportion (Figure 2). In the phylum Firmicutes, Bacillus was the only genus identified in which the species *Bacillus pumilus*, *Bacillus cereusand Bacillus sp* were respectively the most predominant species. The genus *Alcaligenes* with one predominant species (*Alcaligenes faecalis*) in the class Betaproteobacteria was the most dominant among the Proteobacteria (Figure 2).

Soil samples had a higher species composition (Figure 3A), with eighteen 18 species divided into six (06) classes, seven (07) genera and five (05) phyla. Among the endophytes, root tissues and rice seeds had the greatest profusion of bacterial species. These same species were also the predominant species in the soil samples. *Xanthomonas*sp and

Chryseobacteriumcamelliae were the only pathogenic species among the root endophytes and soil samples respectively. In addition, *Bacillus pumilus* followed *by Bacillus*sp were the most predominant species in the soil samples respectively. Bacillus cereus was also the predominant species among the root endophytes (Figure 3A). In these samples at the different sampling sites, the genus *Bacillus* was the most important in which *Bacillus cereus* followed by *Bacillus pumilus* were the predominant species (Figure 3B). The majority of our samples were collected from soils in Baguineda (simple and rhizospheric) and Dilly in the Nara circle (Figure 4).

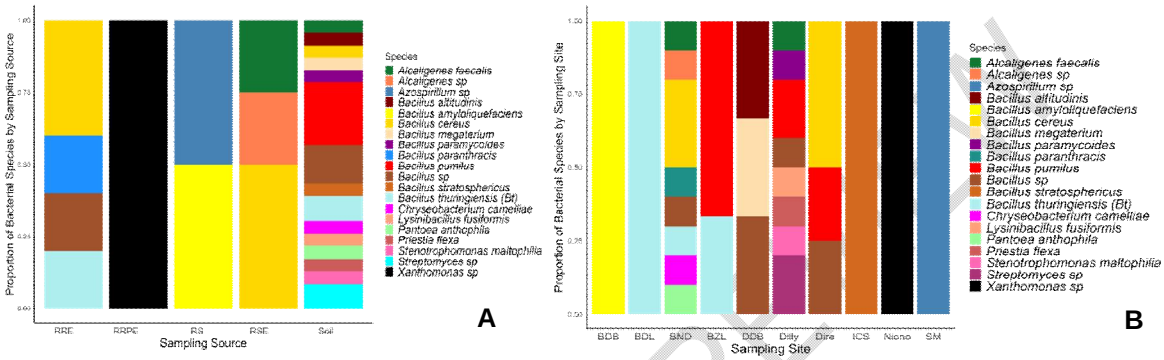

Figure 3: Distribution of bacterial species by sampling source (A) and sampling site (B)

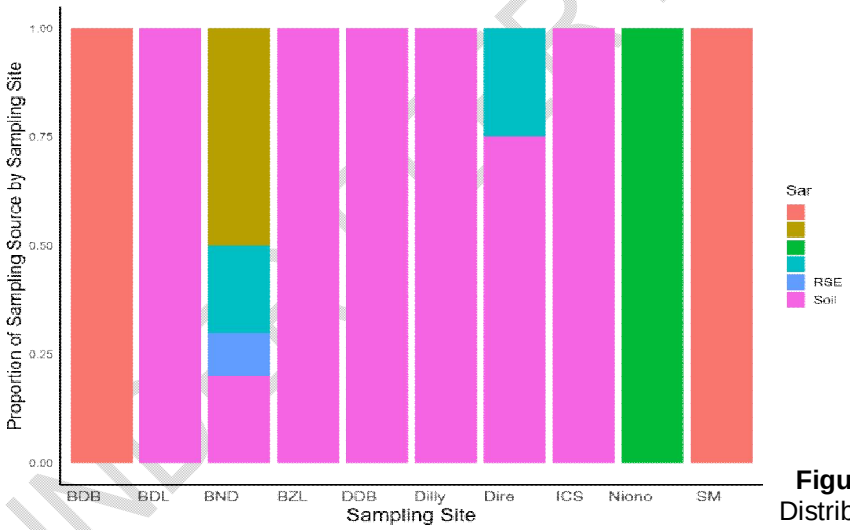

Figure 4: Distribution of sample sources by sampling site.

The dendrogram in Figure 5 illustrates the taxonomic classification of soil-borne and endophytic bacteria after their identification.

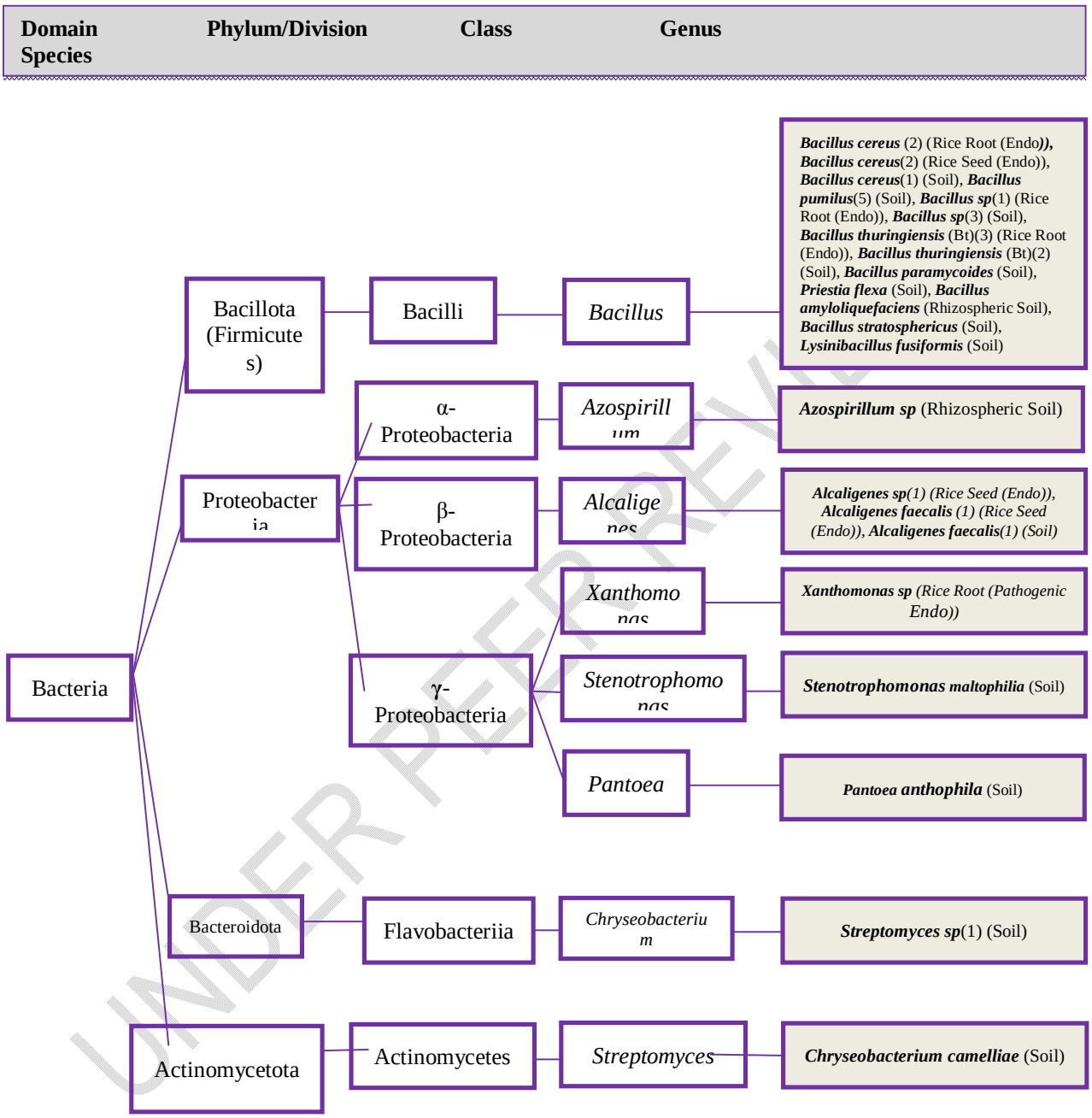

Figure 5: Taxonomic classification of identified soil bacteria and endophytes.

3.1.2 Diversity and evolutionary relationship between bacterial species

Several significant clades were identified in the constructed phylogenetic tree. A clade being defined as a group of organisms comprising a common ancestor, each represents a distinct genetic lineage with ecological and functional implications; " clades were observed:

- Bacillus clade: This genus is widely represented in the tree with robust branches. The species *Bacillus pumilus, Bacillus stratosphericus*and *Bacillus amyloliquefaciens* form a clade supported by relatively high Bootstrap values.
- Alcaligenes clade: It is noted that the species Alcaligenes faecalis appears twice, in different contexts (soil and rice seed).
- Endophyte clade: The bacterial species identified in rice roots and seeds, indicated as endophytes, form distinct clades. An interesting case is that of Bacillus cereus, which appears several times in the tree with different origins (soil, roots, seeds).

The genus Bacillus presents a great genetic diversity with several distinct subgroups. These species are found in various environments, including in soil and as endophytes in rice roots and seeds.

Endophytes such as *Alcaligenes faecalis* and *Bacillus* showed an affinity for plant tissues (roots, seeds), while other species such as *Bacillus pumilus* and *Stenotrophomonas* are present in soil.

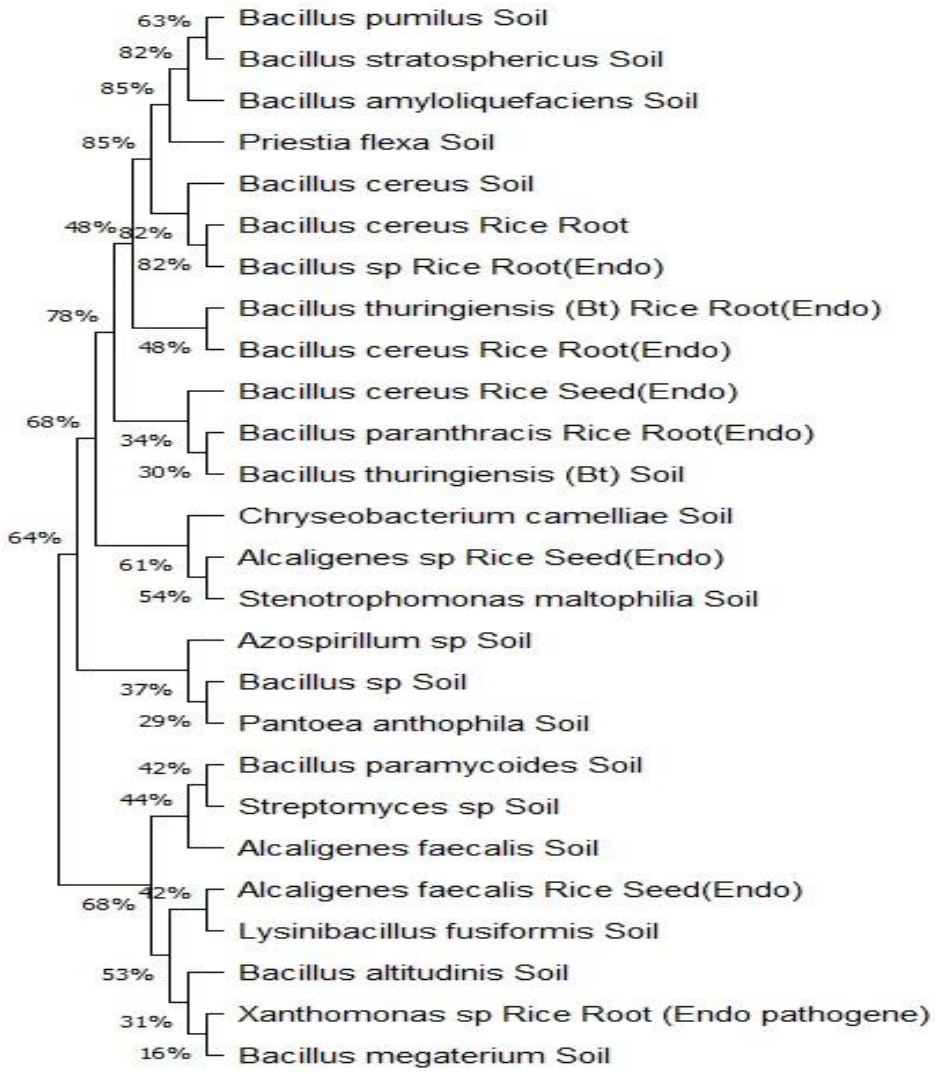

Figure 6: Evolutionary history was inferred using the Neighbor-Joining method. The optimal tree is shown. Evolutionary distances were calculated using the composite maximum likelihood method and are expressed in units of the number of base substitutions per site. The proportion of sites where at least one unambiguous base is present in at least one sequence for each descendant clade is shown next to each internal node in the tree. This analysis was conducted on 26 nucleotide sequences. All ambiguous positions were removed for each pair of sequences (pairwise removal option). There were a total of 1742 positions in the final dataset. Evolutionary analyses were performed in MEGA11.

3.2 Discussion

The taxonomic classification of bacteria often relies on the study of hypervariable regions of the 16S rRNA gene, which are widely used to identify and differentiate species [34], [35]. These 16S genes were selected in this study because of their nucleotide heterogeneity, thus providing an effective ability to identify and classify bacterial species present in samples. This study is similar to the results obtained by Raimi and Adeleke [36]; who reported a

predominance of the phyla Proteobacteria followed by Firmicutes with greater diversity at the genus level, with Pseudomonas and Bacillus being the most frequent genera among almost all plant species and organs. Compared to the results of this study, where Bacillus represents the dominant genus in soil, a similar observation has been made in other works, including that of Xu et al. [37] who reported a higher relative proportion of *Bacillus* in soil compared to the rhizosphere. Thus, the study by Xu et al. [37], presented a higher proportion of Streptomyces in the rhizosphere than in ordinary soil, a contrast to our observations where Bacillus dominated in rhizosphere soil samples.

These differences could reflect variations specific to sample sources or particular environmental conditions influencing the microbial composition.
Phylogenetic analysis revealed several significant clades, including those of the genera Bacillus, Alcaligenes, as well as endophytes such as *Bacillus cereus*, thus aligning our results with previous studies that have also highlighted the diversity and genetic robustness of these bacterial groups. Stable groupings within the genus Bacillus have also been highlighted in recent studies, such as that of Zeng et al. [38], where species such as *Bacillus altitudinis, Bacillus pumilus* and *Bacillus amyloliquefaciens* demonstrated strong genetic cohesion despite their adaptation to diverse environments. Alcaligenes faecalis, repeatedly detected in diverse environments such as soil and rice seeds, was also mentioned in the work of Felestrino et al. [39], where species of the genus Alcaligenes showed a wide ecological distribution, indicating strong adaptability to diverse habitats.

The study of Zhao et al. [40] showed that endophytes MQ23 and MQ23R, although originating from different environments (healthy *S. alopecuroides* plants in northwest China and nodules under greenhouse conditions), exhibited almost perfect genetic homology (100%), belonging to the same Bacillus subclade. In contrast to this homogeneity, our analysis of *Bacillus cereus* revealed distinct clades according to the various environmental contexts (soil, roots, seeds). This divergence shows that, despite different contexts of origin, endophytes can sometimes share high genetic stability, confirming their potential for phylogenetic robustness and similar ecological roles in diverse environments. In this study, although the identification of bacteria is based on 16S rRNA genes, whole genome sequencing could provide a more comprehensive view, revealing more about the functional capacities of bacteria in their various habitats and improving the understanding of their ecological role in agriculture. Finally, the construction of the phylogenetic tree, although efficient, is based on assumptions such as the homogeneity of substitution rates, which may not apply uniformly to all taxonomic groups, thus influencing the interpretation of evolutionary links.

4. CONCLUSION

This study made it possible to characterize the diversity of endophytic, rhizosphere, and telluric bacterial species present in different agricultural sites in Mali. Thanks to a targeted metagenomic approach, thirty-five (35) bacterial species were identified, of which ten sequences could not be analyzed due to sequencing failure. These identified sequences belong to four major phyla, demonstrating a rich microbial diversity. The results revealed a large proportion of Bacillota (Firmicutes) and Pseudomonadota (Proteobacteria), respectively, with genera such as Bacillus and Alcaligenes being the most predominant. Phylogenetic analysis also highlighted significant clades, highlighting the evolutionary links between the studied species.

DISCLAIMER

The authors hereby declare that no generative AI technologies such as large language models (ChatGPT, COPILOT, etc.) and text-to-image generators were used in the writing or editing of this manuscript.

REFERENCES

[1] Riva, V., Mapelli, F., Bagnasco, A., Mengoni, A., & Borin, S. (2022). A meta-analysis approach to defining the culturable core of plant endophytic bacterial communities. *Applied and Environmental Microbiology*, *88*(6), e02537-21.

[2] Middleton, H. (2023). *Implication des microARNs dans la communication plante-microbiote rhizosphérique* (Doctoral dissertation, Université de Rennes).

[3] Husson, O., Sarthou, J. P., & Duru, M. (2023). Référentiels et nouveaux indicateurs pour fonder une agriculture régénératrice.

[4] Alabouvette, C., & Cordier, C. (2018). Fertilité biologique des sols: des microorganismes utiles à la croissance des plantes. *Innovations Agronomiques*, *69*, np.

[5] Ren, Z., Tang, S., Jiang, Y., Jiang, M., Zheng, S., Liu, W., ... & Yin, M. (2018). High-throughput sequencing analysis of endophytic bacteria diversity in fruits of white and red pitayas from three different origins. *Polish Journal of Microbiology*, *67*(1), 27-35.

[6] Khan, M. A., Asaf, S., Khan, A. L., Adhikari, A., Jan, R., Ali, S., ... & Lee, I. J. (2020). Plant growth-promoting endophytic bacteria augment growth and salinity tolerance in rice plants. *Plant Biology*, *22*(5), 850-862.

[7] Saeed, Q., Xiukang, W., Haider, F. U., Kučerik, J., Mumtaz, M. Z., Holatko, J., ... & Mustafa, A. (2021). Rhizosphere bacteria in plant growth promotion, biocontrol, and bioremediation of contaminated sites: a comprehensive review of effects and mechanisms. *International journal of molecular sciences*, *22*(19), 10529.

[8] Mignard, S., & Flandrois, J. P. (2006). 16S rRNA sequencing in routine bacterial identification: a 30-month experiment. *Journal of microbiological methods*, *67*(3), 574-581.

[9] Billy, C. (2003). Détection génotypique des résistances bactériennes: de la phénotypie à la génotypie, deux méthodes complémentaires. *Réanimation*, *12*(3), 192-197.

[10] Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., ... & Claverie, J. M. (2004). The 1.2-megabase genome sequence of Mimivirus. *Science*, *306*(5700), 1344-1350.

[11] Nogales, B., Moore, E. R., Llobet-Brossa, E., Rossello-Mora, R., Amann, R., & Timmis, K. N. (2001). Combined use of 16S ribosomal DNA and 16S rRNA to study the bacterial community of polychlorinated biphenyl-polluted soil. *Applied and Environmental Microbiology*, *67*(4), 1874-1884.

[12] Martineau, M. (2023). *Prévalence, identification et caractérisation des bactéries du genre Mycoplasma détectées dans les affections respiratoires équines* (Doctoral dissertation, Normandie Université).

[13] Větrovský, T., & Baldrian, P. (2013). The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PloS one*, *8*(2), e57923.

[14] Abellan-Schneyder, I., Matchado, M. S., Reitmeier, S., Sommer, A., Sewald, Z., Baumbach, J., ... & Neuhaus, K. (2021). Primer, pipelines, parameters: issues in 16S rRNA gene sequencing. *Msphere*, *6*(1), 10-1128.

[15] Fox, G. E., Pechman, K. R., & Woese, C. R. (1977). Comparative cataloging of 16 S ribosomal ribonucleic acid: molecular approach to procaryotic systematics. *International Journal of Systematic and Evolutionary Microbiology*, *27*(1), 44-57.
[16] Madeira, F., Madhusoodanan, N., Lee, J., Eusebi, A., Niewielska, A., Tivey, A. R., ... & Butcher, S. (2024). The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Research*, gkae241.
[17] Didelot, X. (2023). Phylogenetic Analysis of Bacterial Pathogen Genomes. In *Bacterial Pathogenesis: Methods and Protocols* (pp. 87-99). New York, NY: Springer US.
[18] Carver, T. J., & Mullan, L. J. (2005). Jae: Jemboss alignment editor. *Applied bioinformatics*, *4*, 151-154.
[19] Ferrari, I. V., & Patrizio, P. (2021). Study of Basic Local Alignment Search Tool (BLAST) and multiple sequence alignment (Clustal-X) of monoclonal mice/human antibodies. *BioRxiv*, 2021-07.
[20] Dege, D., & Brüggemann, P. (2024). Marketing analytics with RStudio: a software review.
[21] Keklik, G. (2023). Understanding evolutionary relationships and analysis methods through mega software. *INTERNATIONAL JOURNAL OF NEW HORIZONS IN THE SCIENCES*, 83-90.
[22] Fernández, A., Segura-Alabart, N., & Serratosa, F. (2023). The MultiFurcating Neighbor-Joining Algorithm for Reconstructing Polytomic Phylogenetic Trees. *Journal of Molecular Evolution*, *91*(6), 773-779.
[23] Ranwez, V. (2002). *Méthodes efficaces pour reconstruire de grandes phylogénies suivant le principe du maximum de vraisemblance* (Doctoral dissertation, Université Montpellier II-Sciences et Techniques du Languedoc).
[24] McRoberts, R. E., Næsset, E., Hou, Z., Ståhl, G., Saarela, S., Esteban, J., ... & Chirici, G. (2023). How many bootstrap replications are necessary for estimating remote sensing-assisted, model-based standard errors?. *Remote Sensing of Environment*, *288*, 113455.
[25] Prillo, S., Deng, Y., Boyeau, P., Li, X., Chen, P. Y., & Song, Y. S. (2023). CherryML: scalable maximum likelihood estimation of phylogenetic models. *Nature methods*, *20*(8), 1232-1236.
[26] Bérard, S. (2003). *Comparaison de séquences répétées en tandem et application à la génétique* (Doctoral dissertation, Université Montpellier II-Sciences et Techniques du Languedoc).
[27] Kapli, P., Yang, Z., & Telford, M. J. (2020). Phylogenetic tree building in the genomic age. *Nature Reviews Genetics*, *21*(7), 428-444.
[28] Shafir, A., Halabi, K., Escudero, M., & Mayrose, I. (2023). A non-homogeneous model of chromosome-number evolution to reveal shifts in the transition patterns across the phylogeny. *New Phytologist*, *238*(4), 1733-1744.
[29] Mahadani, A. K., Awasthi, S., Sanyal, G., Bhattacharjee, P., & Pippal, S. (2022). Indel-K2P: a modified Kimura 2 Parameters (K2P) model to incorporate insertion and deletion (Indel) information in phylogenetic analysis. *Cyber-Physical Systems*, *8*(1), 32-44.
[30] Alouane, T. (2022). Analyse bioinformatique de l'organisation structurale et fonctionnelle des génomes de deux micro-organismes pathogènes; la bactérie opportuniste Acinetobacter baumannii et le champignon Aspergillus fumigatus.
[31] Xiang, C. Y., Gao, F., Jakovlić, I., Lei, H. P., Hu, Y., Zhang, H., ... & Zhang, D. (2023). Using PhyloSuite for molecular phylogeny and tree-based analyses. *Imeta*, *2*(1), e87.
[32] Gauthy Choc, L. (2023). État des lieux et analyse de l'apport potentiel du Cloud Computing et des techniques de Re connaissance Optique de Caractères dans le métier d'auditeur.
[33] Ladoukakis, E., Kolisis, F. N., & Chatziioannou, A. A. (2014). Integrative workflows for metagenomic analysis. *Frontiers in cell and developmental biology*, *2*, 70.

[34] Lu, J., & Salzberg, S. L. (2020). Ultrafast and accurate 16S rRNA microbial community analysis using Kraken 2. *Microbiome*, *8*(1), 124.
[35] kadhim Nimr, H. (2023). Detection of 16SRNA Gene variation of bacterial strains isolated from water marsh polluted with oil spills southern of IRAQ. *HIV Nursing*, *23*(2), 079-086.
[36] Raimi, A., & Adeleke, R. (2023). 16S rRNA gene-based identification and plant growth-promoting potential of cultivable endophytic bacteria. *Agronomy Journal*, *115*(3), 1447-1462.
[37] Xu, Q., Jiang, D., Zhou, N., Kang, Y., Li, M., Yang, C., ... & Liu, C. (2024). Community structure of soil microorganisms and endophytes of honeysuckle at different ecological niche specificities. *BMC microbiology*, *24*(1), 367.
[38] Zeng, Q., Xie, J., Li, Y., Gao, T., Zhang, X., & Wang, Q. (2021). Comprehensive genomic analysis of the endophytic Bacillus altitudinis strain GLB197, a potential biocontrol agent of grape downy mildew. *Frontiers in Genetics*, *12*, 729603.
[39] Felestrino, É. B., Sanchez, A. B., Caneschi, W. L., Lemes, C. G. D. C., Assis, R. D. A. B., Cordeiro, I. F., ... & Moreira, L. M. (2020). Complete genome sequence and analysis of Alcaligenes faecalis strain Mc250, a new potential plant bioinoculant. *PLoS One*, *15*(11), e0241546.
[40] Zhao, L., Xu, Y., Sun, R., Deng, Z., Yang, W., & Wei, G. (2011). Identification and characterization of the endophytic plant growth prompter Bacillus cereus strain MQ23 isolated from Sophora alopecuroides root nodules. *Brazilian Journal of Microbiology*, *42*, 567-575.