16S Metagenomics investigation of Bacterial diversity and prediction of its functionalities in Moroccan phosphate mine ecosystem.

Salah Eddine Azaroual

MASCIR: Moroccan Foundation for Advance Science Innovation and Research

Yassine Kasmi

MASCIR: Moroccan Foundation for Advance Science Innovation and Research

Abderrahim Aasfar

MASCIR: Moroccan Foundation for Advance Science Innovation and Research

Hicham El Arroussi

MASCIR: Moroccan Foundation for Advance Science Innovation and Research

Youssef Zeroual

OCP Group: Groupe OCP

Youssef El Kadiri

Institut National d'Hygiene du Maroc

Abdelali Zrhidri

Institut National d'Hygiene du Maroc

Elmostafa Elfahime

CNRST: Centre National de la Recherche Scientifique et Technologique

Abdelaziz Sefiani

Institut National d'Hygiene du Maroc

Issam MEFTAH KADMIRI (✉ i.kadmiri@mascir.ma)

MASCIR: Moroccan Foundation for Advance Science Innovation and Research <https://orcid.org/0000-0003-1067-1652>

Research Article

Keywords: phosphate mine, 16S metagenomics sequencing, bacterial communities

Posted Date: June 28th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-659872/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Native plants in extreme environments may harbor some unique microbial communities with particular functions to sustain their growth and tolerance to harsh conditions. The aim of this study was to investigate the bacterial communities profiles in the Moroccan phosphate mine using 16S metagenomics sequencing. Taxonomic resolution of six hypervariable regions of the 16S rRNA was assessed in the collected samples and mock standard community. The bacterial functions enriched in the samples were also predicted to assess the contribution of the proposed low cost approach in the description of bacterial diversity and functionality in soil and plant.

Methods

The Ion 16S™ metagenomics kit was used to compare the discriminatory potential of the six targeted hypervariable regions of the 16S rRNA gene in the description of the bacterial communities. Alpha and beta diversity analysis were also performed accordingly in the studied samples. The later were collected from the bulk phosphate mine and the soils attached to roots of three different wild plants, that natively grow in this environment. Samples from the rhizosphere of wheat plants in cropping system were also studied. *Actinobacteria* related sequences were compared between the samples to investigate their possible origin between the bulk mine and the wild plants rhizosphere. The prediction of 82 bacterial functions related to nitrogen and phosphate metabolism along with stress was performed using Tax4fun bioinformatics tool.

Results

The rhizosphere of the three wild plants in the Moroccan phosphate mine is characterized by interesting bacterial diversity including *Proteobacteria* (62,24%, 71,15% and 65,61%), *Actinobacteria* (22,53%, 15,24%, 22,30%), *Bacteroidetes* (7,57% ; 4,23% ; 7,63%), and *Firmicutes* (5,82%; 1,17% ; 2,83%). Broad taxonomic diversity of minor phyla was also found in the native plants samples and included *Acidobacteria*, *Armatimonadetes*, *Chloroflexi*, *Cyanobacteria*, *Gemmatimonadetes* and *Verrucomicrobia*. The bulk phosphate mine samples were dominated by *Actinobacteria* with average relative abundance of 97,73% and *Proteobacteria* was Less abundant phylum in these samples. V3, V4 and V67 regions performed better in the taxonomic resolution at different levels e.g., phylum, family and genus. However, V8 and V9 regions gave the lesser taxonomic resolution and V9 region identified only *Proteobacteria* phylum in the mock standard sample. Alpha and beta diversity analysis indicated low abundance of observed OTUs and diversity in the bulk samples compared with the rhizosphere samples of the native plants in the phosphate mine. Moreover, significant ($p = 0,021$) differences in bacterial diversity were observed among at least three kind of samples : The bulk phosphate mine, the rhizosphere of the native wild plants of the phosphate mine and the rhizosphere of the wheat crop. The results indicated that both plant genotype

and mainly soil conditions may be involved in the shaping of bacterial diversity. Such indication was also confirmed by the prediction of functional profiles that showed enrichment of many functions related to biological nitrogen fixation in the rhizosphere of native plants and the stress related functions in the bulk phosphate mine in comparison with the wheat rhizosphere samples

Conclusion

The proposed amplicon sequencing approach combined to bioinformatic prediction of bacterial functions could be considered as low-cost method to study the bacterial communities in soils and plants. The application of such approach to samples from the phosphate mines and wild plants revealed the need to consider at least V3, V4 and V6-7 hypervariable regions of the 16S rRNA gene in the taxonomic resolution of bacterial diversity in soil samples. This approach revealed also interesting bacterial diversity governed by the plant genotypes and the local conditions with interesting predicted functionalities involved in plant growth and development and stress tolerance in extreme environments. Further studies could target some interesting identified phyla and predicted bacterial functionalities, such studies could also deeply investigate the origin of the bacterial diversity in these native plants to develop new generation of bacterial inoculants for sustainable agriculture.

Background

Plant associated microbiota have attracted many researchers worldwide and continue to be one of the most studied research areas in microbiology over the last decades. In a recent revision and review of the concept definition and challenges in microbiome research, Berg et al [1] proposed an explicit definition for the term microbiome based on the previous definitions with some newly introduced amendments. The proposed definition can be summarized as the theatre of activities of microorganisms living in a giving ecosystem. According to this definition, plant microbiomes are not passive actors [2]. They can, rather, interact with host development, physiology and systematic defense (e.g., mycorrhizal symbiosis with wheat plants [3]), induce disease resistance (e.g. bacterial endophyte *Enterobacter* sp. “M6” that confers resistance to fungal pathogen *Fusarium graminearum* in Finger Millet [4]), confer fitness advantages to plant host (e.g. Association of microbiome phenotypes with Arabidopsis genes involved in immunity, cell-wall integrity, root and root hair development [5]) and increase plant tolerance to stress and drought [6-8]. In fact, plant Microbiome was identified as a key for the next green revolution [9], and numerous products and microbiome management strategies were developed in agriculture including (i) microbiome transplants, (ii) microbial inoculants, (iii) microbial and plant extracts, (iv) methods to change environmental conditions and (V) microbiome engineering [10-12]

The analysis of the plant associated microorganisms, their composition and their activities and functions are still challenging. Traditionally, microbiological investigations of microbiota relies on the isolation and the cultivation of single isolates followed by morphological, biochemical and molecular characterizations. However, despite the continuous improvements in cultivation-dependent techniques, an

often cited estimate indicated that only 1% of extant bacteria can be isolated in culture and as much as 99% or more of the microorganisms present in many natural environments are not readily culturable [13,14]. Although this traditional approach is still required in order to isolate single strains that can be applied as inoculum, the newly developed next-generation sequencing (NGS) molecular methods are, nowadays, widely used to profile microbial communities and their dynamic in different plants and plant compartments [15]. Sequencing platforms (e.g., Illumina®, PacBio®, Oxford nanopore® and Ion Personal Genome Machine (PGM) ®) have known interesting advances in wet laboratory protocols, kits and bioinformatics pipelines in plant Microbiome related applications and studies. Such advances allow cost reduction, time preservation and high accuracy in deciphering the structure and function of microbial communities in plant microbiome analysis [16-17]. The microbiome analysis with NGS platforms relies mainly on the amplicon sequencing methods that target the 16S rRNA gene [18]. Its sequences serve as a marker of choice in order to infer the microbiome composition. 16S rRNA gene is universally present in bacteria and archaeobacteria, it contains conserved, variable and hypervariable regions allowing taxonomical classification and separation. These characteristics of the 16S rRNA gene allow the development of numerous primer combinations that are used to amplify hypervariable regions and to generate amplicon of variable lengths suitable for subsequent sequencing with different NGS platforms (e.g., Illumina, Oxford Nanopore, Pacific Biosciences, Ion PGM). Illumina sequencing platforms are considered as the major player among NGS platforms and the MiSeq variant is considered as the default choice for metagenomics studies [18, 19]. In the illumina systems, the variable regions V3 and V4 of the 16S rRNA are predominantly sequenced and analyzed in microbial diversity studies. Similarly, Thermo scientific™ developed Ion 16S™ metagenomics kit that has the ability to amplify and sequence six hypervariable regions (V2, V3, V4, V6-7, V8 and V9) simultaneously [20]. However, this NGS system is rarely used in metagenomics analysis of 16S rRNA gene mainly in the context of plant – microbiome interactions. In this work we used the Ion 16S™ metagenomics approach to investigate the microbial diversity in the rhizosphere and bulk soil of native plants in Moroccan phosphate mine.

Plants are very diverse and occupy most of terrestrial ecosystems, where they play a key role as primary producers. They are also present in extreme and uncommon environments such as deserts, contaminated soils and ores. Many researchers postulated that native plants from these extreme environments may harbor some unique microbial communities with particular functions to sustain their growth and tolerance to harsh conditions [21-22]. Investigations of the microbiome of plants growing in unfavorable environments and stresses may provide insight into microbial and plant traits that allow them to alleviate such stresses (e.g., drought, salinity, pests and pathogens). In this sense, Pérez- Jaramillo et al. [23] proposed a 'back to the roots' framework that comprises the exploration of the microbiomes of native plants vegetation growing in their native habitats for the identification of microbial traits to reinstate beneficial association that may have been lost during plant domestication. For examples, in Chile, the rhizospheres of *Atriplex* sp. and *Stipa* sp. (shrubs) grown in the Atacama Desert were dominated by *Gammaproteobacteria* as revealed by 454–pyrosequencing studies [21]. Zhang et al. [22] described and compared abundance and composition of bacterial communities associated with leaves and roots of plants grown in Atacama desert (*Distichlis spicate* (Poacea) and *Pluchea absinthioides* (Asteraceae)).

Authors revealed the abundance of *Proteobacteria*, *Firmicutes*, *Actinobacteria* and *Bacteroidetes* among the endophytic bacterial communities. Interestingly, most of operational taxonomic units (OTUs) were not shared between endospheres of these indigenous plants indicating the effect of plant genotype on the bacterial communities distribution. The ability of some native isolates from Atacama desert to act as plant growth promoting bacteria (PGPB) was confirmed in wheat plants, they also showed greater protection against salt stress [24].

Moroccan phosphate mines, mainly in the center of Morocco (Ben Guerir Location), are characterized by high temperature in the summer, aridity and rains scarcity, very low organic matter and high phosphate mineral contents. This environment can be considered as extreme and the native plants growing in this habitat could represent the optimal place to look for PGPB with interesting and unique traits that confer resistance to crops. Unfortunately, very few microbiological studies on phosphate mines have been carried out. In one of our recent studies, we isolated and characterized bacilli strains from rock phosphate mines that showed interesting plant growth promotion traits such as phosphate solubilization, phytohormone and ACC-deaminase production [25]. Previously, Hamdali et al. [26,27] studied *Actinobacteria* (e.g., *Streptomyces griseus* and *Micromonospora antiaca*) native in Moroccan phosphate mines that showed abilities to solubilize rock phosphate and other multiple plant growth properties under *in vitro* and greenhouse conditions. The above mentioned studies followed the traditional cultivation-dependent techniques, and to the best of our knowledge there is no study applying the NGS methods to describe the microbial communities in this environment. In this work, we proposed the metagenomics analysis of the 16S rRNA gene in the rhizosphere of wild native plants in phosphate mines along with the bulk mine soil. We compared the composition and abundance of the analyzed bacterial communities with the wheat rhizosphere in cropping natural system. We used in this study the Ion 16STM metagenomics kit (Thermo Fisher Scientific Inc., Genome Biotechnologies, Morocco) in order to compare the discriminatory potential of the six targeted hypervariable regions of the 16S rRNA gene in the description of the bacterial communities in the studied samples. Barb et al. [28] analyzed the six hypervariable regions of the 16S marker gene by using the Ion 16STM metagenomics kit to study the oral microbiome diversity. They developed a novel analytical pipeline with the V4 region enabled the best resolution of bacterial communities profiling up to both family and genus levels. In a recent study, Tamošiūnė et al. [29] assessed the effect of cold Plasma treatment on the plant-associated Microbiome using the Ion 16STM metagenomics kit which seems to be the first study applying this kit on plant-associated microorganisms. However in this study, authors did not give indication related to the best 16S rRNA hypervariable region that gives a robust identification and classification of taxa.

In this work, we analyzed the effect of the six hypervariable regions in the taxonomic assignment, abundance and distribution of the bacterial communities in the studied samples. In this approach, the prediction of the bacterial functionalities based on 16S rRNA profiles was performed using Tax4fun software package available in R environment [30]. With this bioinformatics prediction, we gave insights into the functional capabilities of the microbial communities associated with natives plants in phosphate mines, in a cost-effective way in comparison with the shotgun metagenomics approach.

Results

Sequencing yield and pre-processing

To investigate the microbiome associated with native plants that grow in phosphate mines in Morocco and its comparison with the phosphate bulk soil samples and the rhizosphere of wheat crop, a total of 15 samples were subjected to 16S rRNA metagenomics analysis using Ion Torrent PGM™ sequencing platform. Total obtained reads accounted for 2 782 897. they were filtered for chimeras, quality score and copy number. Reads were classified according to 16S rRNA variable regions for each sample and only reads with high number of copies were mapped against 16S rRNA databases (Supplementary materials Table S1 and Table S2).

Taxonomic resolution of the primers targeting variable regions of the 16S rRNA

Variable regions of the 16S rRNA were amplified in the DNA derived from the studied samples and the ZymoBIOMICS community standard using two separate primer sets with the Ion 16S™ metagenomics kit. Read Classification to taxonomic profiles at four levels e.g., phylum, family, genus and species was compared between the consensus and each region according to the number of taxa identified. In this section, results obtained for the standard ZymoBIOMICS DNA were also presented.

Figure 1 showed that at the phylum level, V2, V3, V4, V67 and V8 performed well for the identification of the phyla present in the ZymoBIOMICS standard. However, V9 region identified only *Proteobacteria* phylum. Taxonomic resolution at more lower levels became challenging and V9 region gave less resolution up to family, genus and species levels. V67 and V8 regions gave better resolution than V9, nevertheless, new families and genera were found with these regions that are not present in the standard. V2, V3 and V4 regions gave the best resolution up to both family and genus levels (Figure 1).

The same analysis was performed for the collected soil samples (Figure 2). However, since the bacterial composition of these samples is unknown compared to the standard ZymoBIOMICS, the maximum number of each taxonomic level present in each sample was taken as the target resolution. For each region, resolution percentages were calculated accordingly and presented as heatmap. The overall analysis of the figure 2 indicated variability in the taxonomic resolution of different 16S rRNA regions according to the taxonomic level and the sample types. However, some regions gave the lesser taxonomic resolution e.g., V9 and V2 whereas V3, V4, V67 and V8 performed better in the taxonomic resolution mainly at the phylum level. At higher taxonomic levels e.g., Genus and Species, the number of taxa recovered by every hypervariable region decreased for almost all the regions except the V3 region.

Bacterial diversity analysis by consensus sequences in the studied samples:

16S Metagenomics analysis included samples from bulk phosphate mine, samples from the rhizosphere of three native wild plants in phosphate mine and samples from the wheat rhizosphere. Figure 3 presented the distribution and abundance of bacterial phyla, families and genera in the studied samples

using the consensus sequences from the all variable regions of the 16s rRNA gene. Results of the bacterial composition clearly showed differences according to sample types and the plant species. Bacterial composition of the samples obtained from bulk phosphate mine in the vicinity of the wild plants was dominated by *Actinobacteria* with average relative abundance of 97,73% (figure 3a). *Proteobacteria* was Less abundant in these samples compared to *Actinobacteria*. At the family and genus levels, an average of 94,54% and 94,29% of OTUs were assigned to *Nocardioidaceae* and *Nocardioides*, respectively in these samples. This result indicated the selective presence and dominance of *Nocardioides* members in this particular environment among the bacterial diversity. As showed in figure 3, the bacterial composition was completely different in the rhizosphere samples of the three native plants. *Proteobacteria* phylum was almost enriched than other phyla with relative abundance of 62,24%, 71,15% and 65,61%, in the rhizosphere *Alternanthera caracasana* (NP1); *Digitaria sanguinalis* (NP2) and *Dittrichia Viscosa* (NP3), respectively. Within this phylum, main representative classes were *Betaproteobacteria* (7,1%; 40,38%; 23,15% in NP1, NP2 and NP3, respectively), *Alphaproteobacteria* (26,84%; 19,65% ; 22,85% in NP1, NP2 and NP3, respectively) and *Gammaproteobacteria* (28,05%; 7,85%; 15,68% in NP1, NP2 and NP3, respectively). *Deltaproteobacteria* were less abundant. High relative abundances of *Actinobacteria* (22,53%, 15,24%, 22,30% in NP1, NP2 and NP3, respectively) were also found in these native plant samples, however, at lower level, *Actinobacteria* were mainly represented by *Actinomycetales* order (21,81% ; 12,58; 19,17% in NP1, NP2 and NP3, respectively) with wide taxonomic diversity at family and genus level. It is noteworthy that the relative abundance of *Nocardioidaceae* family was significantly reduced in these samples in comparison with the bulk phosphate mine samples (Figure 3). Sequences assigned to *Actinobacteria* were further analyzed in this study. Other major phyla that were found in these samples were *Bacteroidetes* (7,57% ; 4,23% ; 7,63% in NP1, NP2 and NP3, respectively), and *Firmicutes* (5,82%; 1,17% ; 2,83% in NP1, NP2 and NP3, respectively). Broad taxonomic diversity of minor phyla was also found in the native plants samples and included *Acidobacteria*, *Armatimonadetes*, *Chloroflexi*, *Cyanobacteria*, *Gemmatimonadetes* and *Verrucomicrobia* (Figure 3). The wheat rhizosphere samples showed a relatively similar pattern as the bacterial composition of the rhizosphere of native plants. In these samples, the assignment of taxonomic affiliation revealed high relative abundances of *Proteobacteria* (57,42%), *Actinobacteria* (35,24%), *Firmicutes* (5,52%) and *Nitrospirae* (2,32%). Broad taxonomic diversity of minor phyla was also found in the wheat rhizosphere samples. *Actinobacteria* were mainly represented by *Actinomycetales* and *Gaiellales* orders with 15,53% and 9,36%, respectively.

Alpha diversity measures (Observed OTUs, Chao, ACE and Shannon) and rarefaction curves are presented in Figure 4 and supplementary material Figure S1. Rarefaction analysis enabled us to gauge the extent to which total diversity had been recovered at each sample and indicated that the extent of bacterial diversity was captured by the sequencing approach. Figure 4 showed significantly lower diversity and OTUs number in the bulk mine samples. This finding was expected, because of the extreme conditions met in such habitat. Interestingly, the native plants in the phosphate mine showed higher observed OTUs, species richness and species diversity and evenness in comparison with the wheat crop rhizosphere samples, even if the observed differences were statistically not significant (Figure 4). Higher alpha

diversity measures were recorded for the rhizosphere samples of the phosphate mine native plant *Dittrichia Viscosa* (NP3), this increase was statistically significant with other native plants NP1 and NP2 samples (Figure 4). In contrast to other samples, the wheat rhizosphere samples were characterized by variable intra-diversity, which makes the box-plot range larger than in other samples. Some replicates were highly diverse and rich in terms of species, while others were characterized by their lower richness in bacterial communities (Figure 4).

Beta diversity measures included Bray Curtis distance unweighted and weighted UniFrac distances and jaccard dissimilarity. PCoA of the latter was presented in figure 5 and clearly indicated differences in bacterial diversity among at least three kind of samples : The bulk phosphate mine, the rhizosphere of the native wild plants of the phosphate mine and the rhizosphere of the wheat crop. Statistical analyses using PERMANOVA revealed significant differences between the three groups of samples ($p=0,021$). This finding indicated that the plant genotypes had limited influence in the shaping the bacterial diversity in this particular environment.

The distribution of shared and unique OTUs among bacterial communities in the studied samples was presented in figure 6. High number of OTUs was shared between the samples at the phylum level such as *Actinobacteria*, *Proteobacteria*, *Firmicutes* and *Bacteroidetes* (Figure 6a). The shared OTUs between the rhizosphere samples (native plants and wheat crops) included *Armatimonadetes*, *Planctomycetes* and *Verrucomicrobia*, whereas only *Nitrospirae* phylum was shared between the bulk mine samples and the rhizosphere of native plants. Similarly, at lower taxonomic levels e.g., family (Figure 6b) and genus (Figure 6c), high number of OTUs was shared between the samples (28 OTUs among 86 and 33 among 317 at the family and genus levels, respectively). Interestingly, high number of OTUs at the family level (27) were shared between the rhizosphere samples of native plants and wheat crops which included some plant-associated bacteria such as *Rhizobiaceae* and *Phyllobacteriaceae*, and indicated the effect of the plant rhizosphere in the interactions with bacterial diversity. The same statement is true for the shared OTUs at the genus level with high number among the plant samples (figure 6c)

Actinobacteria phylogenetic analysis :

According to taxonomy and representative sequences tables of Qiime2, the number of sequences assigned to *Actinobacteria* varied between samples. Bulk phosphate mine samples accounted for 168 sequences assigned to *Actinobacteria*, whereas 336 sequences were assigned to *Actinobacteria* in the rhizosphere samples of the wild plants in phosphate mine. Figure 7 showed the phylogenetic relatedness and the inferred evolutionary relationships between the *Actinobacteria* sequences in the studied samples and revealed that *Actinobacteria* in Bulk RP were phylogenetically different and separated from those in native plant samples. *Actinobacteria* related sequences in the bulk phosphate mine samples yielded three different clades whereas, those in native plants samples were more diverse with many branches. *Actinobacteria* related Sequences in NP1 and NP3 were the most closely related among all samples with a common node of 72% (Figure 7). *Actinobacteria* in the Bulk phosphate samples were less evolutive

than the sequences in the native plants samples which could be maybe related to adaptation and co-evolution process between the bacteria and the plants. All together these results indicated the selection of the *Actinobacteria* by plants in their rhizosphere and the absence of possible migration of some *Actinobacteria* from the bulk mine samples to the plants rhizosphere in their surrounding environment.

Prediction of Bacterial functions enriched in the samples :

The Bioinformatic tool Tax4fun was used in this study to predict bacterial functions involved in nitrogen and phosphorus metabolism and stress across the studied samples (Figure 8). Figure 8a showed the enrichment of some nitrogen fixation functions, mainly in the bacteria of the rhizosphere of the phosphate mine native plants *Alternanthera caracasana* (NP1) and *Dittrichia Viscosa* (NP3). The enriched bacterial functions in these samples included nitrogen fixation proteins (mainly NifQ, NifB, NifN, NifT, NifX, NifZ, NifE) and nitrogenase cofactors. The nitrogen fixation protein NifU was mainly enriched in bacteria assessed in wheat rhizosphere and the phosphate mine native plant *Digitaria sanguinalis* (NP2). Nitrogen regulatory proteins PII 1 and PII 2 KEGG were also enriched in bulk phosphate samples and some native plants and are linked to the regulation of carbon metabolism under conditions of low nitrogen availability. Heatmap of the bacterial functions involved in phosphate metabolism (Figure 8b) showed the enrichment of 11 predicted KEGG functions involved in mineralization, solubilization and transport (Phosphatase, acid phosphatases and inorganic phosphate transporter) among bacterial communities in the bulk phosphate mine samples. Interestingly, the assigned bacteria to the rhizosphere of native wild plants showed similar patterns, with the enrichment of many functions involved in the mineralization, solubilization and transport of phosphate. In the wheat rhizosphere samples, phosphate transport functions were mainly enriched with some phosphatase involved the phosphate mineralization processes. Figure 8C showed the enrichment profiles of bacterial functions related to stresses, it indicated that the Na⁺/H⁺ antiporters NhaC and NhaB, GABA permease, anion: H⁺ symporter DASS family, 1-aminocyclopropane-1-acyteltransferase, lipopolysaccharide O-acyteltransferase, polysaccharides transporter PST family, lipopolysaccharides export system permease protein, lipopolysaccharides biosynthesis protein WzzE, acyl homoserine lactone synthase and acyl homoserine lactone efflux proteins, were highly enriched in phosphate rock samples compared to rhizospheric soil of native plants in phosphate mine and the rhizospheric soil of wheat crop. On the other hand, all of these bacterial functions are higher in the rhizospheric soil of native plants compared to the rhizospheric soil of wheat crops. On the other hand, the NhaC family and phosphate:Na⁺ symporter were enriched in the rhizospheric soil of native plants than in bulk phosphate samples.

Discussion

In this work we adopted a cost effective approach to decipher the bacterial communities distributions and functions in soil samples. We used the Ion 16STM metagenomics approach to assess the six hypervariable regions of the 16S rRNA (V2, V3, V4, V6-7, V8 and V9) simultaneously, in the taxonomic assignment, abundance and distribution of the bacterial communities in the studied samples. The

prediction of the bacterial functionalities based on 16S rRNA profiles was performed using Tax4fun software package to highlight any possible interesting bacterial functionalities that open perspectives for in depth future soil microbiome studies in phosphate mine.

With the advancement, the accessibility and the affordability of sequencing technologies, the use of high throughput sequencing platforms in soil microbial ecology has become popular and revolutionized our understanding of the soil – plant – microorganisms interactions. However, Nannipieri et al. [31] discussed and criticized the use of DNA-based tools in the determination of soil microbial diversity and functionalities. Schöler et al. [32] highlighted in their opinion paper biases and pitfalls in the steps of generation, processing and interpretation of 16S rRNA sequencing data in the field of soil and plant microbiome studies. They clearly recommended the sequencing of the entire 16S rRNA gene using some new long read sequencing platforms in comparison to shorter amplicons which are mostly generated by Illumina sequencing.

To our knowledge, the present work is the first study dealing with soil bacterial DNA that used simultaneously processed multiple hypervariable regions of the 16S rRNA gene. Barb et al. [28] used multiple hypervariable regions of the 16S rRNA gene to study the human microbiome and concluded that the use of multiple regions improve the results of identifying bacteria that compose complex communities such as those inhabiting human. The Ion 16S™ Metagenomics kit was also used to evaluate the performance of individual hypervariable regions of the 16S rRNA gene, for capturing human vaginal microbiota [33]. Authors of this study indicated that analyses using the V3 region generally indicated the highest bacterial diversity followed by the V6–V7 and V4 regions, while the V9 region gave the lowest bacterial resolution. With a similar way, we used in our study the Ion 16S™ Metagenomics Kit with both genomic DNA from ZymoBIOMICS mock standard sample and DNA extracted from soil samples originated from different habitat mainly the Moroccan phosphate mine and native plants in the ore. The obtained results in the mock samples presented in figure 1 were in accordance with the findings of Barb et al. [28] and Sirichoat et al. [33], since the hypervariable regions V2, V3, V4 and V6-7 regions gave the best resolution up to both family and genus levels. V9 region gave less resolution and identified only *Proteobacteria* phylum. However, the percentage of taxa resolved by each hypervariable region of the 16S rRNA in the DNA of the studied samples is variable according to the taxonomic level and the sample origin. In this case, V3, V4 and V67 performed better in the taxonomic resolution at different levels. On the other hand, V9 and V8 gave the lesser taxonomic resolution (Figure 2). This finding could also corroborate the suggestion of Barb et al. [28] which stated that the Ion 16S™ Metagenomics Kit does not perform well on the end of the 16S rRNA gene. Fuks et al. [34] proposed a method called Short Multiple Regions Framework (SMURF), to combine sequencing results from different PCR-amplified regions to provide one coherent profiling in order to improve the resolution of bacterial profiling. Recent study benchmarked a new 16S full-length-based synthetic long-read (sFL16S) technology that enables long-read sequencing by using an existing Illumina short-read sequencer and the general 16S rRNA amplicon (V3–V4) sequencing methods [35]. These studies, clearly indicated and recommended the necessity to consider the entire hypervariable regions rather than partial analysis of the 16S rRNA gene in the characterization of the diversity and composition of the microbial communities inhabiting specimen.

In our study, we characterized the bacterial communities structure in the phosphate ore (Bulk_RP) and we compare this bacterial diversity with the rhizosphere samples of some spontaneous wild plants within the phosphate ore (*Alternanthera caracasana* (NP1); *Digitaria sanguinalis* (NP2) and *Dittrichia Viscosa* (NP3)) along with the rhizosphere of wheat in cropping system (WR). We used the consensus analysis of six hypervariable regions of the 16S rRNA gene thanks to the Ion 16S™ metagenomics protocol in the taxonomic units assignments. Interesting findings are provided in this study, the *Actinobacteria* accounted for more than 97% in the ore bulk samples with significantly lower bacterial diversity expressed by the alpha diversity indices (Figure 4), and the enrichment of particular family and genus e.g., *Nocardioides*. *Actinobacteria* are generally ubiquitous and have the ability to survive under extreme conditions. They form mycelia and/or produce spores to reach water and nutrients and to survive in these environments. They can also influence the weathering of mineral elements [36]. Many reports describing the bacterial diversity in extreme habitats e.g., Chilean Deserts, Arabian peninsula desert, alpine pants and contaminated mine tailings are in accordance with our study and found *Actinobacteria* among the dominant phyla in these habitats [21,37-39]. Members of *Nocardioides* genus have also been isolated using cultivation-based approach from some extreme niches in Morocco and the obtained isolates showed plant growth-promoting activities e.g., phosphorus solubilization, potassium solubilization, siderophores and indole-3-acetic acid production [40]. However, Nafis et al. [40] specified that *Nocardioides* genus was rare among the other *Actinobacteria* genera present in these extreme niches which needs to be evaluated in light of our results showing that this genus was highly abundant.

A shift in the bacterial communities composition was observed between the bulk rock phosphate mine samples and the rhizosphere of native plants grown in this niche (Figure 3). This finding confirmed the roles of root exudates in recruiting and shaping the plant rhizosphere microbiome which is intensively reported [41,42]. This shift was translated by the enrichment of *Betaproteobacteria*, *Alphaproteobacteria*, *Gammaproteobacteria*, *Bacteroidetes* and *Firmicutes* with the appearance of new phyla mainly *Verrucomicrobia* (Supplementary materials Figure S2). *Actinobacteria* were also found with high relative abundance in the native plants rhizosphere however the *Nocardioides* genus was lower than newly enriched taxa within *Actinomycetales* and *Gaiellales* orders (Figure 3). Alpha diversity measures presented in figure 4 indicated low abundance of observed OTUs and diversity in the bulk samples compared to the rhizosphere samples of the natives plants in the phosphate mine. Generally, studies indicated that the bulk soil under “normal conditions” is more abundant and diverse in OTUs than the rhizosphere part and plants recruit best adapted taxa [43-45]. However, Zhang et al. [46] reported similar results in the study of rhizosphere microbial communities diversity of seagrasses *Thalassia hemprichii* and *Enhalus acoroides* in the Xincun Bay in China. One explanation provided by the authors is that bulk sediments are characterized by lower organic matter contents compared to seagrass beds. Lower organic matter and lower humidity are also the case of phosphate mine ore in the studied area compared to the rhizoplane where plants create favorable microenvironment for higher bacterial diversity. Many factors are involved in the plant recruitment of microbial communities such as soil type, edaphic conditions and the host plant genotype [47]. Results of beta diversity presented in figure 5 indicated significantly separated bacterial diversity between the native plants, the bulk mine and the wheat crop

with insignificant differences between the three studied native plants. This result may suggest the influence of the soil conditions rather than the genotype of the host plant in the shaping of bacterial communities. Such suggestion is in accordance with the conclusion of Eida et al. [39] who showed that the host desert plants phylogeny aligns to a lesser degree with the rhizosphere bacterial communities which are largely determined by the soil conditions. In this particular habitat, selective pressures are exerted to plants with adapted genotype and their associated bacterial communities to choose the best adapted phyla to help them thrive in such conditions [45]. Furthermore, it was intriguing to investigate the occurrence of some OTUs in the rhizosphere of native plants through the phylogenetic analysis of *Actinobacteria* related sequences. We found interesting separation of clades according to the sample origins (bulk phosphate vs native plants rhizosphere) (figure 7) which opens hypothesis to explain the origin of *Actinobacteria* diversity in the studied samples mainly in the rhizosphere of the ore native plants. We reported high relative abundance of *Nocardioides* genus in the bulk samples exceeding 90% and which drastically decreased in the rhizosphere samples of the native plants, where we found more diversity within the *Actinomycetales* order. This diversity could be attributed to a plants-driven factors that determine the dominant bacterial populations in their rhizosphere [48]. Another explanation could be the occurrence of distinct ecotypes in the wild plants rhizosphere that have co-evolved. In fact, Lopes et al. [49,50] showed in two different specific studies distinct populations of *Pseudomonas koreensis* and *Pseudomonas putida* between sugarcane rhizosphere or bulk soil and concluded that the divergence of these populations might have resulted from the differential selective pressures posed by the bulk soil versus rhizosphere habitats even in the populations of same species. Furthermore, the effect of microbial seed load on rhizosphere and endosphere bacterial communities was investigated in wheat by Kavamura et al. [51]. In fact, plant seeds harbor diverse microbial communities and their composition is determined by plant genotype, environment, and management practices [52]. The seed borne microbiome transmission is still poorly investigated and links between seed and soil microbiomes are not fully understood. Such transmission could also explain the diversity of bacterial communities in the rhizosphere of the native plants studied in the phosphate ore. In any case, further specific studies are needed to investigate the occurrence of bacterial diversity in different part of plants and the surrounding environment mainly in extreme contexts.

In this study, we predicted the functional profiles of bacterial taxa based on 16S rRNA sequencing data using the Tax4Fun bioinformatic tool (Figure 8). This approach could be adapted in low cost studies based on the 16S metagenomics sequencing rather than the expensive shotgun metagenomics. It could also give insights into some particular bacterial functions under different context and conditions that could be targeted in subsequent studies. The obtained results, presented in figure 8, showed some interesting findings. For example, the enrichment of the nitrogenase functions and the nitrogenase molybdenum-iron (MoFe protein) in the rhizosphere of native plants in the nitrogen-deprived phosphate mines, explained the recruitment by these plants of atmospheric nitrogen-fixing bacteria to ensure its nitrogen requirements. Such functions were not found in rock phosphate bulk samples without plants, which strengthened the recruitment hypothesis. In wheat plants rhizosphere, these functions were less enriched since this environment is rich in nitrogen coming from fertilizers. The nitrogen fixation process is

based on the transfer of electrons from the Fe protein to FeMo-co, in which a nitrogen molecule is converted to two ammonia molecules. The biosynthetic process of all proteins involved in this mechanism such as FeMo-co are extremely vulnerable to environmental and endogenous oxygen. The aerobic and micro-aerobic nitrogen-fixing organisms need to create strict anaerobic microenvironments in the cell to facilitate active nitrogenase functions. To meet this need, special accessory enzymes / proteins are required for the biosynthesis of clusters and mainly FeMo-co such as the genes products of nifH, nifQ, nifB, nifE, and nifN and nitrogen fixation proteins NifX involved in efficient transfer processes of NifB-co [53,54]. In this sense, NifZ is involved in the maturation of the P-cluster, the second cluster of catalyzing the biological nitrogen fixation reaction by nitrogenase. On the other hand, the expression of the nitrogen regulatory proteins PII 1 and PII 2 are linked to the regulation of carbon metabolism under conditions of low nitrogen availability. In bacteria, PII proteins control target proteins in response to cellular ATP/ADP levels and the 2-oxoglutarate status, thereby coordinating the cellular carbon/nitrogen balance. In addition to their involvement in the regulation of carbon metabolism, these proteins are able to control the activity of acetyl-CoA carboxylase and glutamine synthetase (GS) in response to the availability of nitrogen sources [55]. The resulted prediction functions explain the hypothesis of recruitment of nitrogen-fixing bacteria by native plants under nitrogen starvation, while ensuring the control of carbon metabolism and energy for proper functioning of nitrogenase under these stress conditions. Similarly, prediction of bacterial functions related to the solubilization and assimilation of phosphate revealed the enrichment of specific functions according to samples nature. In rock phosphate bulk samples, several functionalities were enriched such as low-affinity inorganic phosphate transporters, acid phosphatase, 3-phytase, glucose-1-phosphatase, alkaline phosphatase, phospholipase, phosphonoacetaldehyde hydrolase, as well as phosphate regulon PhoB (Figure 8b). Enrichment of these functions could be explained by the low availability of inorganic phosphate in the rock phosphate, and the bacterial populations showed these adaptation mechanisms under this harsh environment such as the activation of low-affinity inorganic phosphate transporters. At the same time, the activation of the solubilization potential of phosphate via the expression of the gene products of phosphatases such as acid phosphatase, 3-phytase, glucose-1-phosphatase, alkaline phosphatase, phospholipase, phosphonoacetaldehyde hydrolase. However, due to the dephosphorylating action, acid phosphatase activity could indirectly influence inorganic P solubilization by lowering the medium pH [56]. In the same way, phosphate regulon PhoB which is induced by phosphate deprivation was also enriched in the phosphate rock samples. The Phosphate (Pho) regulon is a bacterial regulatory mechanism used for the conservation and management of inorganic phosphate within the cell [57]. In the other hand, Bacterial high affinity transport systems are involved in active transport of solutes across the cytoplasmic membrane. Most of the bacterial ATP-binding cassette importers are composed of one or two transmembrane permease proteins, one or two nucleotide-binding proteins and a highly specific periplasmic solute-binding protein [58]. This explained the prediction results of the enrichment of functions involved in activation of active transport in the case of wheat crops, such as the phosphate transporters, as well as, the enrichment of The inorganic phosphate transporters (PiT) family which appear to catalyze inorganic phosphate (Pi).

Results of the stress related bacterial functions enrichment showed that they were higher in the bulk phosphate mine samples and in the rhizosphere of native plants compared to wheat crop. Looking at the role of these stress related functions presented in figure 8C, we find that the expression of these functional genes is linked to stress tolerance. Na^+ and H^+ are the most common ions and they play primary roles in cell physiology: both are important in cell bioenergetics. Indeed, when the concentration of these ions becomes too high or too low they turn into potent stressors to cells. Every cell has a very efficient homeostatic mechanism for these ions. Proteins that play a primary role in this homeostatic mechanism are the Na^+/H^+ antiporters. NhaC, the key Na^+/H^+ antiporter, plays a principal role for regulation of alkaline phosphatase biosynthesis, which has the role of solubilization of phosphate in alkaline conditions of rock phosphate. In the case of the rhizospheric soil of native plants existing in the phosphate mine, the low enrichment of this function could be explained by the change in soil conditions thanks to the root exudates rich in organic acid, amino acids and other metabolites and organic carbon. This explains the increase in microbial diversity in this rhizospheric soil of these native plants compared to that of phosphate mine presented in alpha diversity measures. Low-molecular-weight organic compounds in root exudates play a key role in plant–microorganism interactions by influencing the structure and function of soil microbial communities. Under these conditions of nitrogen and carbon stress in the phosphate rock, the GABA permease function was highly enriched. This enzyme is responsible for the release of GABA which plays a very important role in the regulation of nitrogen uptake, and at the same time considered as carbon source under stress conditions [59]. On the other hand, bacterial surface molecules known as lipopolysaccharide (LPS), produced by most Gram-negative bacteria, create a permeability barrier at the cell surface and is a main contributor to the innate resistance that Gram-negative bacteria display against many stresses [60]. This postulate was supported by the high expression of lipopolysaccharide O-acyltransferase, polysaccharides transporter PST family, lipopolysaccharides export system permease protein, lipopolysaccharides biosynthesis protein WzzE in rock phosphate samples considered as a stressful environment. In addition, soil-dwelling bacteria collectively referred to rhizobia synthesize and perceive N-acyl-homoserine lactone (AHL) signals to regulate gene expression in a population density-dependent manner [61]. This signaling mechanism is called Bacterial quorum sensing (QS) involved in many physiological processes in bacteria [62]. According to this analysis, we can say that the functions expressed by the bacterial community in rock phosphate are linked to the tolerance and survival of these microorganisms in this rather extreme environment. The decrease of these functions linked to the increase in the expression of other functions in rhizospheric soil of these extremophilic native plants could be explained by the change in the conditions of this micro-environment by the root exudates rich in organic acid and other metabolites which can influence the pH, the availability of a carbon source that allows a new recruitment of other microbial populations ensuring other functions such as atmospheric nitrogen fixation, phosphate and other nutrients solubilization.

Conclusion

To our knowledge, the present work is the first study dealing with soil bacterial DNA that used simultaneously processed multiple hypervariable regions of the 16S rRNA gene. The proposed amplicon sequencing approach combined to bioinformatic prediction of bacterial functions could be considered as low cost method to study the bacterial communities in soils and plants. The application of such approach to samples from the phosphate mines soil and wild plants in this environment revealed the need to consider multiple hypervariable regions of the 16S rRNA gene. In the taxonomic resolution of bacterial diversity in soil and plant samples we concluded the need to consider the V3 – V4 and V6-7 regions rather than the widely used V3 – V4 primers and the V9 region identified only *Proteobacteria* phylum.

Application of the above mentioned approach to bulk samples from Moroccan phosphate mine and the rhizosphere of some native plants in this environment along with wheat plants in cropping system revealed interesting bacterial diversity. The distribution of different bacterial taxa at different levels was governed by the plant genotypes and to wider extent by the local conditions. The plants in the phosphate mine were able to enrich and/or recruit new phyla not present in the bulk samples. Regarding the origin of the bacterial diversity in the native plants, we hypothesized possible recruitment and co-evolution of these taxa with their hosts and/or the seed borne microbiome transmission to the plants. Such hypothesis need further investigation. Interesting predicted bacterial functionalities were also revealed using bioinformatic tools. The functions were involved in plant growth and development and stress tolerance in extreme environments. These functions were enriched and associated with the bacterial taxa distribution both in the bulk and native plants in phosphate mine samples more than the wheat rhizosphere samples. This opens a paramount opportunities to target and investigate the interesting identified phyla and the predicted bacterial functionalities in the wild plants to develop new generation of selected bacterial inoculants for sustainable agriculture.

Materials And Methods

Sampling and experimental design

In this project, 15 soil samples were collected from the sampling sites that include the bulk soil of the phosphate ore in Ben Guerir city in Morocco (N 32°13'28.81533 ; W 7°50'22.412") (Figure 9A), soil rhizosphere of three native plants in the ore that spontaneously grow in this environment (Figure 9 B, C, D) and the rhizosphere of wheat plants in cropping system in the region of Sidi Allal Tazi, Kénitra city in Morocco (N 34°31'30.5394" ; W 6°19'11.8164"). Three independent samples were taken from the bulk soil of the phosphate mine and coded Bulk_RP1, Bulk_RP2 and Bulk_RP3. Three native plants were also collected and three independent samples were taken from the soil loosely attached to their roots as follow *Alternanthera caracasana* (Figure 9 B) NP1_1, NP1_2 and NP1_3 ; *Digitaria sanguinalis* (Figure 9 C) NP2_1, NP2_2 and NP2_3 and *Dittrichia Viscosa* (Figure 9 D) NP3_1, NP3_2 and NP3_3. Three independent sample were also taken from the durum wheat (*Triticum aestivum L.ssp.durum*) plants rhizosphere and coded WR_1, WR_2 and WR_3. All samples were collected manually under aseptic

conditions at a depth of 10 – 20 cm around the roots, placed into sterile tubes and plastic bags, transported to the laboratory at 4°C and stored at -20°C until the DNA extraction.

Additionally, a mock sample control consisting of the ZymoBIOMICS™ Microbial Community Standard (supplementary materials figure S3), which contains samples from 8 well-known species of bacteria and 2 yeasts was run concurrently along with the soil samples, to assess data quality and the taxonomic resolution of the sequencing strategy.

DNA extraction and metagenomic analysis

Genomic DNA was extracted from the 15 samples using DNeasy® PowerSoil® Kit (Qiagen, 12888), following the manufacturer's procedure. The purity calculated by the absorbance ratio A_{260}/A_{280} and A_{260}/A_{230} and quantity of DNA samples were determined using NanoDrop 2000 UV-Vis spectrophotometer (Thermo scientific, USA) and Qubit™ 4 Fluorometer (Thermo scientific, USA), respectively. The 16S Metagenomic sequencing library preparation protocol (Thermo-Fisher Scientific, USA) was followed for the DNA library preparation. For each sample, the six hypervariable regions of the 16S rRNA gene were amplified using two separate PCR reactions with two specific primer sets V2-4-8 and V3-6,7-9 of the kit Ion 16S Metagenomics (Thermo-Fisher Scientific, Cat. no. A26216). Each 30µl PCR mix contained 15µl 2X Environmental Master Mix, 3µl 16S rRNA Primer set and 12µl DNA sample. Cycling conditions were as follows: denaturation at 95°C for 10min, 25 cycles of 30 s, annealing at 56°C for 30 s and 20 s extension at 72°C. Negative control (water) and positive control (*E. coli* DNA) reactions were used to confirm the amplification of 16S rRNA and equal volumes of the 2 PCR reactions were combined. Final DNA library was made from 10 - 100 ng of combined amplicons using the Ion Plus Fragment Library Kit (Thermo-Fisher Scientific, Cat. no. 4471252) and Ion Xpress Barcodes Adapters, 1–16 (Thermo-Fisher Scientific, (Thermo-Fisher Scientific, Cat. no. 4471252) following recommendations of the manufacturer. Adapter-ligated and nick-repaired DNA was purified using x1.4 volumes of Agencourt AMPure beads (Beckman Coulter, USA) and eluted in Tris-EDTA buffer. The optimal library concentration for template preparation was estimated by qPCR (Applied Biosystems® 7500 Fast) using the Ion Universal Library Quantitation Kit (Thermo-Fisher Scientific, Cat. no. A26217). Each library was adjusted to a concentration of 10 pM and equal volumes of each library were pooled. The Ion Chef™ instrument was used to clonally amplify the pooled library on nanosized ionosphere particles by emulsion PCR. The bead enrichment and the loading of Ion 318™ Chip Kit v2 were also conducted using The Ion Chef™ instrument. The sequencing was run on the Ion PGM™ System using the Ion PGM™ Sequencing 400 Kit according to the manufacturer protocol.

Basecalling and Bioinformatics analysis of sequencing data

Basecalling and demultiplexing were performed with the Torrent Suite™ Software version 5.0.5 (Life Technologies) using recommended parameter settings. DNA sequencing data were collected in FASTQ format, analyzed using QIIME 2 v2020.8 [62]. Briefly, Raw torrent ion sequence reads were imported from QIIME2 using the "Sample Data" (Sequences with Quality). Then denoised and filtered with the dada2

pipeline to eliminate noisy and chimeric sequences, demultiplexed, and construction of noised paired-end sequences. Taxonomic identification was performed using Curated MicroSEQ(R) 16S Reference Library v2013.1, Curated Greengenes v13.5Greengenes v.13.5 and Silva databases, threshold value for percentage identity for genus and species ID was 97%.

All the basic measures used in the analysis of alpha and beta diversity were calculated on the basis of the rooted phylogenetic tree. In addition to observed OTU richness via alpha diversity, we used the phylogenetic diversity of Faith'DP (a phylogenetic measure of diversity based on total branch length of the bacterial 16S rRNA gene phylogeny captured by each sample) [63].The sampling depth of 100 readings, a reason to normalize the variance, which excluded four samples. Among these, the Kruskal-Wallis test (in pairs, p-value 0,05) was used to assess the statistical significance of alpha diversity. For Beta diversity analysis, following QIIME2 diversity plugin, OTUs were identified with qualitative (Jaccard and unweighted UniFrac) and quantitative (Bray-Curtis and weighted UniFrac) distance measurements at a sample depth of 100 readings. The statistical significance between the different groups was assessed by the permutation-based ANOVA (PerMANOVA) test [64] with 999 permutations (beta group significance command in the diversity plugin). Thereafter, Principal coordinate analysis (PCoA) plots were generated by the QIIME2 Emperor tool to explore the structure of the bacterial community. The bar graphs showing the taxonomy levels were generated by the Microsoft excel tool using the OTUs tables generated by Ion Reporter™ Software version 5.18. The package phyloseq was used to perform post-analysis of Alpha and Beta diversity. The Figures were generated by ggplot in R.

To separate reads into different variable regions from the 16S Metagenomics Kit we followed two strategies. The first using Mothur as it is described in Barb et al. [28] and the second by Qiime2 using specific primers for each region. For the Mothur method, briefly, we aligned the full set of reads with the Mothur script align.seqs using Silva as the reference database. The aligned report file was then submitted to an R script that counts the number of reads with the same start and stop alignment position along the reference database. The forward and reverse reads were then visualized along the translated 16S rRNA gene coordinates and reads were grouped into their corresponding variable regions based on where they aligned.

However for Qiime2 method, The regions V2; V3; V4, V6-7 and V8 were analyzed by the same Qiime2 pipeline described above. in addition, analysis as well as for comparison between the regions were carried out in the R environment (<http://www.r-project.org/>). Also, the Venn diagram was generated by R package according to different taxonomic levels and the graphical output was in the form of a Venn / Euler diagram.

Table 1. Information on the selected primer pairs used for extracting the different hypervariable regions of the 16S rRNA gene for the mothur analysis pipeline described by Barb et al. (2016)

S.No.	Forward primer	Sequence of the forward primer	Reverse primer	Sequence of the reverse primer	Region
1	119	AGYGGCGNACGGGTGAGTAA	338	TGCTGCCTCCCGTAGGAGT	V2
2	357	CCTACGGGAGGCAGCAG	518	ATTACCGCGGCTGCTGG	V3
3	577	AYTGGGYDTAAAGNG	785	TACNVGGGTATCTAATCC	V4
4	785	AGGATTAGATACCCT	907	CCGTCAATTCCTTTGAGTTT	V5
5	978	TCGAtGCAACGCGAAGAA	1062	ACATtTCACaACACGAGCTGACGA	V6
6	1114	GYAACGAGCGCAACCC	1220	GTAGCRCGTGTGTMGCCC	V7
7	1070	ATGGCTGTCGTCAGCT	1385	ACGGGCGGTGTGTAC	V8
8	119	AGYGGCGNACGGGTGAGTAA	518	ATTACCGCGGCTGCTGG	V23
9	357	CCTACGGGRSGCAGCAG	798	GGGGTATCTAATCCC	V34
10	357	CCTACGGGAGGCAGCAG	907	CCGTCAATTCCTTTGAGTTT	V35
11	577	AYTGGGYDTAAAGNG	907	CCGTCAATTYYTTTTRAGTTT	V45
12	805	GGATTAGATACCCTGGTAGTC	1062	ACAGCCATGCAGCACCT	V56
13	985	CAACGCGAAGAACCCTTACC	1220	GTAGCRCGTGTGTMGCCC	V67
14	1065	AGGTGCTGCATGGCTGT	1391	GACGGGCGGTGWGTRCA	V78

Phylogenetic analysis of Actinobacteria related sequences :

To analyze the Actinobacteria sequences obtained in the Meta-taxonomics output, the *Actinobacteria* assigned sequences were exported from Qiime2 via the "Qiime taxa filter-seqs" command and then transformed to a fasta format to perform the phylogenetic analyses. Subsequently, the sequences of plant sequences were coded by plant name or Bulk via the command "sed 's/>NP1_ */' foo.in >bar.out". The sequences were then combined in a single queue and aligned by MAFFT. *E. coli* "NR_024570.1" sequence was set as a reference for the alignment. The output of MAFFT is imported into MEGAX [65] to construct the phylogenetic tree by the Maximum likelihood method with a bootstrap of 1000 times. We collected also the matrix of phylogenetic distance and the diversity parameters calculated by MEGAX. The phylogenetic signal with K and λ were calculated by R via package phytools.

Predicting functional profiles from 16S rRNA data using Tax4Fun

The prediction of the functional profiles of bacterial taxa based on 16S rRNA sequencing data was performed with Tax4Fun which an open-source R package using the OTUs Table containing taxonomic data according to Silva's database [30]. This tool was used to predict functional community profiles based on 16S rRNA data to outline bacterial metabolism in the studied samples in the phosphate mines in comparison to the wheat crop. A list of 82 KEGG functions involved in Biological nitrogen fixation (16 KEGG), phosphate metabolism (41 KEGG) and stress related functions (25 KEGG) were assessed for differentially enrichment (KeGG Ontology, KO) using non parametric T-test performed on QIIME. Results were presented as heatmap using R.

Declarations

Availability of Data and materials

The datasets generated during the current study are available in the NCBI repository, under the Bioproject number PRJNA729805 available through the web link <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA729805>. KeGG Ontology tables and additional data are available upon request.

Acknowledgements

Authors are sincerely thankful for the financial support of the OCP Group and for the access to phosphate mine. Special Acknowledgements to members of laboratories involved in this work for their support during the accomplishment of experiments. A huge thanks to Rachid GHANI and Dr. Zakaria HAZZOUNI for their precious help in sampling and plants identification.

Funding information

This work was financially supported by OCP group and is part of a research program conducted in the Moroccan Foundation for Advanced Science, Innovation and Research (MAScIR) in collaboration with Mohammed VI Polytechnic University.

References

1. Berg G, Rybakova D, Fischer D, Cernava T, Vergès M-CC, Charles T, et al. Microbiome definition revisited: old concepts and new challenges. *Microbiome*. 2020;8:103. doi:10.1186/s40168-020-00875-0.
2. Trivedi P, Leach JE, Tringe SG, Sa T, Singh BK. Plant-microbiome interactions: from community assembly to plant health. *Nat Rev Microbiol*. 2020;18:607–21.
3. Fiorilli V, Vannini C, Ortolani F, Garcia-Seco D, Chiapello M, Novero M, et al. Omics approaches revealed how arbuscular mycorrhizal symbiosis enhances yield and resistance to leaf pathogen in wheat. *Sci Rep*. 2018;8:9625. doi:10.1038/s41598-018-27622-8.
4. Mousa WK, Shearer C, Limay-Rios V, Ettinger CL, Eisen JA, Raizada MN. Root-hair endophyte stacking in finger millet creates a physicochemical barrier to trap the fungal pathogen *Fusarium graminearum*. *Nat Microbiol*. 2016;1:16167.
5. Bergelson J, Mittelstrass J, Horton MW. Characterizing both bacteria and fungi improves understanding of the *Arabidopsis* root microbiome. *Sci Rep*. 2019;9:24. doi:10.1038/s41598-018-37208-z.
6. Rodriguez RJ, Henson J, Van Volkenburgh E, Hoy M, Wright L, Beckwith F, et al. Stress tolerance in plants via habitat-adapted symbiosis. *ISME J*. 2008;2:404–16. doi:10.1038/ismej.2007.106.
7. Coleman-Derr D, Tringe SG. Building the crops of tomorrow: advantages of symbiont-based approaches to improving abiotic stress tolerance. *Front Microbiol*. 2014;5:283. doi:10.3389/fmicb.2014.00283.

8. Yuan Y, Brunel C, van Kleunen M, Li J, Jin Z. Salinity-induced changes in the rhizosphere microbiome improve salt tolerance of *Hibiscus hamabo*. *Plant Soil*. 2019;443:525–37. doi:10.1007/s11104-019-04258-9.
9. National Academies of. Sciences EM, Education DBSS, Society BEC, Division HM, Board FN, Studies DEL, et al. Science Breakthroughs to Advance Food and Agricultural Research by 2030. National Academies Press; 2019. <https://books.google.co.ma/books?id=hjiPDwAAQBAJ>.
10. Berg G. Plant-microbe interactions promoting plant growth and health: perspectives for controlled use of microorganisms in agriculture. *Appl Microbiol Biotechnol*. 2009;84:11–8.
11. Mueller UG, Sachs JL. Engineering Microbiomes to Improve Plant and Animal Health. *Trends Microbiol*. 2015;23:606–17. doi:<https://doi.org/10.1016/j.tim.2015.07.009>.
12. Song C, Zhu F, Carrión VJ, Cordovez V. Beyond Plant Microbiome Composition: Exploiting Microbial Functions and Plant Traits via Integrated Approaches. *Front Bioeng Biotechnol*. 2020;8:896. doi:10.3389/fbioe.2020.00896.
13. Amann RI, Ludwig W, Schleifer KH. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev*. 1995;59:143–69. doi:10.1128/mr.59.1.143-169.1995.
14. Schleper C, Jurgens G, Jonuscheit M. Genomic studies of uncultivated archaea. *Nat Rev Microbiol*. 2005;3:479–88.
15. Fadiji AE, Babalola OO. Metagenomics methods for the study of plant-associated microbial communities: A review. *J Microbiol Methods*. 2020;170:105860. doi:<https://doi.org/10.1016/j.mimet.2020.105860>.
16. Fricker AM, Podlesny D, Fricke WF. What is new and relevant for sequencing-based microbiome research? A mini-review. *J Adv Res*. 2019;19:105–12. doi:<https://doi.org/10.1016/j.jare.2019.03.006>.
17. Johnston-Monje D, Lopez Mejia J. Botanical microbiomes on the cheap: Inexpensive molecular fingerprinting methods to study plant-associated communities of bacteria and fungi. *Appl Plant Sci*. 2020;8:e11334–4. doi:10.1002/aps3.11334.
18. D'Amore R, Ijaz UZ, Schirmer M, Kenny JG, Gregory R, Darby AC, et al. A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. *BMC Genom*. 2016;17:55. doi:10.1186/s12864-015-2194-9.
19. Luo C, Tsementzi D, Kyrpides N, Read T, Konstantinidis KT. Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. *PLoS One*. 2012;7:e30087.
20. IonTorrent. Ion 16S Metagenomics Kit and Ion Reporter metagenomics workflow solution. 2016;;2. <https://tools.thermofisher.com/content/sfs/brochures/ion-16s-reporter-metagenomics-workflow-flyer.pdf>.
21. Jorquera MA, Maruyama F, Ogram AV, Navarrete OU, Lagos LM, Inostroza NG, et al. Rhizobacterial Community Structures Associated with Native Plants Grown in Chilean Extreme Environments. *Microb Ecol*. 2016;72:633–46. doi:10.1007/s00248-016-0813-x.

22. Zhang Q, Acuña JJ, Inostroza NG, Mora ML, Radic S, Sadowsky MJ, et al. Endophytic Bacterial Communities Associated with Roots and Leaves of Plants Growing in Chilean Extreme Environments. *Sci Rep*. 2019;9:4950. doi:10.1038/s41598-019-41160-x.
23. Pérez-Jaramillo JE, Mendes R, Raaijmakers JM. Impact of plant domestication on rhizosphere microbiome assembly and functions. *Plant Mol Biol*. 2016;90:635–44. doi:10.1007/s11103-015-0337-7.
24. Acuña JJ, Campos M, de la Luz Mora M, Jaisi DP, Jorquera MA. ACCD-producing rhizobacteria from an Andean Altiplano native plant (*Parastrephia quadrangularis*) and their potential to alleviate salt stress in wheat seedlings. *Appl Soil Ecol*. 2019;136:184–90. doi:https://doi.org/10.1016/j.apsoil.2019.01.005.
25. Azaroual SE, Hazzoumi Z, Mernissi N, El, Aasfar A, Meftah Kadmiri I, Bouizgarne B. Role of Inorganic Phosphate Solubilizing Bacilli Isolated from Moroccan Phosphate Rock Mine and Rhizosphere Soils in Wheat (*Triticum aestivum* L) Phosphorus Uptake. *Curr Microbiol*. 2020;77:2391–404. doi:10.1007/s00284-020-02046-8.
26. Hamdali H, Bouizgarne B, Hafidi M, Lebrihi A, Virolle MJ, Ouhdouch Y. Screening for rock phosphate solubilizing Actinomycetes from Moroccan phosphate mines. *Appl Soil Ecol*. 2008;38:12–9. doi:https://doi.org/10.1016/j.apsoil.2007.08.007.
27. Hamdali H, Hafidi M, Virolle MJ, Ouhdouch Y. Growth promotion and protection against damping-off of wheat by two rock phosphate solubilizing actinomycetes in a P-deficient soil under greenhouse conditions. *Appl Soil Ecol*. 2008;40:510–7. doi:https://doi.org/10.1016/j.apsoil.2008.08.001.
28. Barb JJ, Oler AJ, Kim H-S, Chalmers N, Wallen GR, Cashion A, et al. Development of an Analysis Pipeline Characterizing Multiple Hypervariable Regions of 16S rRNA Using Mock Samples. *PLoS One*. 2016;11:1–18. doi:10.1371/journal.pone.0148047.
29. Tamošiūnė I, Gelvonauskienė D, Haimi P, Mildažienė V, Koga K, Shiratani M, et al. Cold Plasma Treatment of Sunflower Seeds Modulates Plant-Associated Microbiome and Stimulates Root and Lateral Organ Growth. *Front Plant Sci*. 2020;11:1347. doi:10.3389/fpls.2020.568924.
30. Aßhauer KP, Wemheuer B, Daniel R, Meinicke P. Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. *Bioinformatics*. 2015;31:2882–4.
31. NANNIPIERI P, ASCHER-JENULL J, CECCHERINI MT, PIETRAMELLARA G, RENELLA G, SCHLOTTER M. Beyond microbial diversity for predicting soil functions: A mini review. *Pedosphere*. 2020;30:5–17. doi:https://doi.org/10.1016/S1002-0160(19)60824-6.
32. Schöler A, Jacquioud S, Vestergaard G, Schulz S, Schlöter M. Analysis of soil microbial communities based on amplicon sequencing of marker genes. *Biol Fertil Soils*. 2017;53:485–9. doi:10.1007/s00374-017-1205-1.
33. Sirichoat A, Sankuntaw N, Engchanil C, Buppasiri P, Faksri K, Namwat W, et al. Comparison of different hypervariable regions of 16S rRNA for taxonomic profiling of vaginal microbiota using next-generation sequencing. *Arch Microbiol*. 2021;203:1159–66.

34. Fuks G, Elgart M, Amir A, Zeisel A, Turnbaugh PJ, Soen Y, et al. Combining 16S rRNA gene variable regions enables high-resolution microbial community profiling. *Microbiome*. 2018;6:17. doi:10.1186/s40168-017-0396-x.
35. Jeong J, Yun K, Mun S, Chung W-H, Choi S-Y, Nam Y, et al. The effect of taxonomic classification by full-length 16S rRNA sequencing with a synthetic long-read technology. *Sci Rep*. 2021;11:1727. doi:10.1038/s41598-020-80826-9.
36. Řeháková K, Chroňáková A, Křišťůfek V, Kuchtová B, Čapková K, Scharfen J, et al. Bacterial community of cushion plant *Thylacospermum ceaspiosum* on elevational gradient in the Himalayan cold desert. *Front Microbiol*. 2015;6:304. doi:10.3389/fmicb.2015.00304.
37. Ghosh S, Das AP. Metagenomic insights into the microbial diversity in manganese-contaminated mine tailings and their role in biogeochemical cycling of manganese. *Sci Rep*. 2018;8:8257. doi:10.1038/s41598-018-26311-w.
38. Wassermann B, Cernava T, Müller H, Berg C, Berg G. Seeds of native alpine plants host unique microbial communities embedded in cross-kingdom networks. *Microbiome*. 2019;7:108. doi:10.1186/s40168-019-0723-5.
39. Eida AA, Ziegler M, Lafi FF, Michell CT, Voolstra CR, Hirt H, et al. Desert plant bacteria reveal host influence and beneficial plant growth properties. *PLoS One*. 2018;13:1–20. doi:10.1371/journal.pone.0208223.
40. Nafis A, Raklami A, Bechtaoui N, El Khalloufi F, El Alaoui A, Glick BR, et al. Actinobacteria from Extreme Niches in Morocco and Their Plant Growth-Promoting Potentials. *Diversity*. 2019;11. doi:10.3390/d11080139.
41. Philippot L, Raaijmakers JM, Lemanceau P, Putten WH, Van Der. Going back to the roots: the microbial ecology of the rhizosphere. *Nat Publ Gr*. 2013;11:789–99. doi:10.1038/nrmicro3109.
42. Haichar F el, Marol Z, Berge C, Rangel-Castro O, Prosser JI, Balesdent JI. J, et al. Plant host habitat and root exudates shape soil bacterial community structure. *ISME J*. 2008;2:1221–30. doi:10.1038/ismej.2008.80.
43. Praeg N, Pauli H, Illmer P. Microbial Diversity in Bulk and Rhizosphere Soil of *Ranunculus glacialis* Along a High-Alpine Altitudinal Gradient. *Front Microbiol*. 2019;10:1429. doi:10.3389/fmicb.2019.01429.
44. Araya JP, González M, Cardinale M, Schnell S, Stoll A. Microbiome Dynamics Associated With the Atacama Flowering Desert. *Front Microbiol*. 2020;10:3160. doi:10.3389/fmicb.2019.03160.
45. Alsharif W, Saad MM, Hirt H. Desert Microbes for Boosting Sustainable Agriculture in Extreme Environments. *Front Microbiol*. 2020;11:1666. doi:10.3389/fmicb.2020.01666.
46. Zhang X, Zhao C, Yu S, Jiang Z, Liu S, Wu Y, et al. Rhizosphere Microbial Community Structure Is Selected by Habitat but Not Plant Species in Two Tropical Seagrass Beds. *Front Microbiol*. 2020;11:161. doi:10.3389/fmicb.2020.00161.
47. Berg G, Smalla K. Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. *FEMS Microbiol Ecol*. 2009;68:1–13. doi:10.1111/j.1574-

6941.2009.00654.x.

48. García-Salamanca A, Molina-Henares MA, van Dillewijn P, Solano J, Pizarro-Tobías P, Roca A, et al. Bacterial diversity in the rhizosphere of maize and the surrounding carbonate-rich bulk soil. *Microb Biotechnol*. 2013;6:36–44.
49. Lopes LD, Pereira e Silva M, de C, Weisberg, Davis AJ, Yan IIEW, Q, Varize C de. S, et al. Genome variations between rhizosphere and bulk soil ecotypes of a *Pseudomonas koreensis* population. *Environ Microbiol*. 2018;20:4401–14. doi:<https://doi.org/10.1111/1462-2920.14363>.
50. Lopes LD, Weisberg AJ, Davis EW 2nd, Varize C, de S, Pereira, Silva E, M de C, Chang. JH, et al. Genomic and metabolic differences between *Pseudomonas putida* populations inhabiting sugarcane rhizosphere or bulk soil. *PLoS One*. 2019;14:e0223269.
51. Kavamura VN, Robinson RJ, Hayat R, Clark IM, Hughes D, Rossmann M, et al. Land Management and Microbial Seed Load Effect on Rhizosphere and Endosphere Bacterial Community Assembly in Wheat. *Front Microbiol*. 2019;10:2625. doi:10.3389/fmicb.2019.02625.
52. Berg G, Raaijmakers JM. Saving seed microbiomes. *ISME J*. 2018;12:1167–70. doi:10.1038/s41396-017-0028-2.
53. Imperial J, Ugalde RA, Shah VK, Brill WJ. Role of the *nifQ* gene product in the incorporation of molybdenum into nitrogenase in *Klebsiella pneumoniae*. *J Bacteriol*. 1984;158:187–94.
54. Shah VK, Hoover TR, Imperial J, Paustian TD, Roberts GPLP. Role of *nif* gene products and homocitrate in the biosynthesis of iron-molybdenum cofactor. *Bothe H, deBruijn FJ, Newt WE Nitrogen Fixat hundred years after*. 1988;:pp115-120.
55. Durand A, Merrick M. In vitro analysis of the *Escherichia coli* AmtB-GlnK complex reveals a stoichiometric interaction and sensitivity to ATP and 2-oxoglutarate. *J Biol Chem*. 2006;281:29558–67.
56. Achal V, Savant VV, Reddy MS. Phosphate solubilization by a wild type strain and UV-induced mutants of *Aspergillus tubingensis*. *Soil Biol Biochem*. 2007;39:695–9. doi:<https://doi.org/10.1016/j.soilbio.2006.09.003>.
57. Santos-Beneit F. The Pho regulon: a huge regulatory network in bacteria. *Front Microbiol*. 2015;6:402. doi:10.3389/fmicb.2015.00402.
58. Saurin W, Dassa E. Sequence relationships between integral inner membrane proteins of binding protein-dependent transport systems: evolution by recurrent gene duplications. *Protein Sci*. 1994;3:325–44.
59. Batushansky A, Kirma M, Grillich N, Pham PA, Rentsch D, Galili G, et al. The transporter GAT1 plays an important role in GABA-mediated carbon-nitrogen interactions in *Arabidopsis*. *Front Plant Sci*. 2015;6:785. doi:10.3389/fpls.2015.00785.
60. Bertani B, Ruiz N. Function and Biogenesis of Lipopolysaccharides. *EcoSal Plus*. 2018;8.
61. Calatrava-Morales N, McIntosh M, Soto MJ. Regulation Mediated by N-Acyl Homoserine Lactone Quorum Sensing Signals in the *Rhizobium-Legume* Symbiosis. *Genes (Basel)*. 2018;9.

62. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol.* 2019;37:852–7. doi:10.1038/s41587-019-0209-9.
63. Faith DP. Conservation evaluation and phylogenetic diversity. *Biol Conserv.* 1992;61:1–10. doi:https://doi.org/10.1016/0006-3207(92)91201-3.
64. Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* 2001;26:32–46. doi:10.1111/j.1442-9993.2001.01070.pp.x.
65. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol.* 2018;35:1547–9. doi:10.1093/molbev/msy096.

Competing Interests

The authors declare no competing interests.

Figures

Figure 1

Percentages resolution for Phylum, Family, Genus and species present in the ZymoBIOMICS standard according to different hypervariable regions of the 16S rRNA (V2, V3, V4, V67, V8 and V9) analyzed by the 16S ion Metagenomics sequencing protocol.

Figure 2

Percentages resolution presented by heatmap plots for Phylum, Family, Genus and species estimated in Bulk phosphate mines (Bulk RP) ; Alternanthera caracasana (NP1) ; Digitaria sanguinalis (NP2); Dittrichia Viscosa (NP3) and wheat crop (WR) samples according to different hypervariable regions of the 16S rRNA (V2, V3, V4, V67, V8 and V9) analyzed by the 16S ion Metagenomics sequencing protocol.

Figure 3

Relative abundances of Bacterial composition at phylum (a) family (b) and genus (c) level in Bulk rock phosphate samples (Bulk_RP) and rhizosphere samples of phosphate native plants *Alternanthera caracasana* (NP1), *Digitaria sanguinalis* (NP2), *Dittrichia Viscosa* (NP3) and wheat crop (WR).

Figure 4

Box plot representation of Alpha diversity measures (Observed OTUs, Chao, ACE, Shannon) of bacterial communities in Bulk phosphate mines (Bul), native wild plants of the phosphate mine (NP1, NP2 and NP3) and the wheat Rhizosphere (WR) samples (n=3 for all samples)

Figure 5

Principal coordinate analysis (PCoA) of bacterial communities from Bulk phosphate mine samples (from Bulk_RP1 to Bulk_RP3), wild phosphate mine plants *Alternanthera caracasana* (from NP1_1 to NP1_3), *Digitaria sanguinalis* (from NP2_1 to NP2_3), *Dittrichia Viscosa* from NP3_1 to NP3_3) and wheat crop (from WR_1 to WR_3). Analyses were done by using Jaccard distance matrix and visualized with emperor

Figure 6

Shared OTUs at the phylum (a) , family (b), genus (c) and species (d) level among bacterial communities in Bulk phosphate mine samples (Bulk), rhizosphere samples from wild phosphate mine plants *Alternanthera caracasana* (NP1), *Digitaria sanguinalis* (NP2), *Dittrichia Viscosa* (NP3) and wheat crop (WR).

Figure 7

phylogenetic tree of Actinobacteria sequences in three native plants samples represented by light colors (pink, sky blue and gray) and bulk phosphate mines samples represented by dark colors (purple, blue and green). The tree was generated by multiple pairwise alignment of sequences assigned to Actinobacteria derived from Qiime2 output using MEGAX and MAFFT Software. The phylogenetic tree was constructed by the Maximum likelihood program and a bootstrap analysis was performed with 1000 trials. *E. coli* "NR_024570.1" sequence was used as a reference.

Figure 9

Geographic location (google earth images) and Samples of Moroccan phosphate mine (A) rock phosphate mining site – Ben Guerir – Morocco ((B)NP1: *Alternanthera bettzickiana*; (C)NP2: *Digitaria sanguinalis*; (D)NP3: *Dittrichia Viscosa*).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Azaroualetal.MicrobiomeBMCSupplementarymaterials.docx](#)