Microbial and Functional Diversity within the Phyllosphere of *Espeletia sp.* in Andean High Mountain Ecosystems

Carlos A. Ruiz-Pérez[1,2,3], Silvia Restrepo[2,3] and María Mercedes Zambrano[1,2*]

[1] *Molecular Genetics, Corporación CorpoGen, Bogotá DC, Colombia;* [2] *Colombian Center for Genomics and Bioinformatics of Extreme Environments, Bogotá DC, Colombia;* [3] *Andes University, Bogotá DC, Colombia.*

[*] *Corresponding author*

Abstract

The phyllosphere is a global habitat that harbors diverse communities of microorganisms. Little is known regarding the microbial communities associated with *Espeletia* sp., a plant endemic to Andean high mountain extreme ecosystems. Here, we used 16S rRNA gene amplicon sequencing to assess the structure of the microbial communities and their metabolic potential, coupled with a GeoChip functional microarray to confirm this metabolic capacity. We observed an overall similarity in the structure of microbial communities across different sites of the plant and a versatile and diverse metabolic potential. This characterization is the first step in studying plant phyllosphere in these ecosystems and for future bioprospecting studies in search for enzymes, compounds and even microorganisms relevant to industry or remediation efforts such as the degradation of pesticides or contaminant aromatic compounds.

Introduction

Andean high mountain environments have been reported as hotspots of diversity around the world, mainly because of their endemic species (Myers et al 2000). Colombian Páramos, ecosystems within these Neotropical Andes that consist of isolated, high elevation areas above the tree line, are constantly exposed to

conditions that are considered extreme, such as high incidence of solar radiation and exposure to large shifts in temperatures during the day. These high mountain conditions impose a selective pressure on native plants and their associated microbiota, making them diverse and unique (Hughes and Eastwood 2006), and a model of study due to their ability to survive under conditions that are extreme for other forms of life. In fact, the Páramo ecosystem has been reported to be the world's fastest evolving biodiversity hotspot (Madrinan et al 2013). The National Natural Park los Nevados (PNN los Nevados) in Colombia, includes several interesting high mountain environments in addition to páramos, such as perpetual snows, arid soils, and hot springs associated with volcanic activity (Alfaro 2002, Bohorquez et al 2012, Wyngaarden and Fandiño 2002).

The phyllosphere of endemic plants from Páramos, has not been analyzed to a great extent and represents a unique ecosystem for microbial communities. The phyllosphere refers to all above ground surfaces of any plant, including leaves, stems, buds, flowers and fruits (Whipps et al 2008). It has been reported as probably the largest ecosystem on earth colonized by microorganisms, mainly bacteria and fungi (Lindow and Brandl 2003), with an approximate extension of 10^8 km^2 (Balint-Kurti et al 2010). The phyllosphere, especially the leaves, act as landing stages for spores or other propagules where microorganisms can develop and multiply (Vokou et al 2012). Compared to other ecosystems, the phyllosphere microbiota has had little attention, but interest is growing due to the potential it poses in terms of microbial ecology and bioprospecting. The most representative plant in Colombian Páramos is known as Frailejón, an endemic plant belonging to the genus *Espeletia* (Fagua and Gonzalez 2007, Monasterio and Sarmiento 1991). The genus *Espeletia* is part of the Asteraceae family, they are present between 3,000 and 4,000 meters above sea level (m.a.s.l.), and they have a thick trunk, with succulent hairy leaves disposed in a dense spiral pattern. Based on the developmental stage, these plants can be separated into different "tiers". The upper

tier is composed of young leaves somewhat protected from the environment, the mid-tier is composed of fully mature leaves exposed to environmental conditions, the necromass tier is composed of senescent leaves, and finally the root tier, which is subject to environment conditions that can be very different to those of the phyllosphere. The ecology, evolution, molecular and functional diversity of microbial populations associated with these plants, are key aspects in order to understand how life develops in such extreme habitats and may present insights on how microbial communities interact in harsh environments.

Both environmental conditions and the host must influence the functional ecology of plant microbial communities (Bodenhausen et al 2013), driving their composition and interactions. Microbial communities associated with plants such as *Espeletia* (i.e. the epiphytes and endophytes) should therefore reflect the adaptations to the environmental conditions they are exposed to, and have the metabolic plasticity required for them to thrive. The different plant tiers also represent various microenvironments in which microbial communities should be taxonomically diverse or at least metabolically differentiated. Based on these hypothesis, we analyzed the microbial communities associated with *Espeletia* sp. from Colombian Páramos, which remain mostly unknown with the exception of some studies regarding endophytic fungi and their biocontrol potential (Miles et al 2012). In this work we used culture independent means (i.e. 16S rRNA gene sequencing and GeoChip 5.0 functional microarray) to address community structure, diversity and functional potential using samples from different plant tiers. The description of bacterial communities allowed us to compare microbial structure across the plant and to link certain groups of microorganisms with important metabolic pathways, highlighting microbial contributions to particular geo-biological processes and the potential of these communities in terms of microbial and metabolic plasticity and adaptation.

Materials and methods

Study site and sampling

Sampling was performed at El Coquito hot spring (04°52'27" N; 75°15'51.4" W) in the PNN Los Nevados. Leaves were sampled from the plant *Espeletia* sp., following reported methodologies (Izhaki et al 2013, Redford et al 2010, Vokou et al 2012) with some modifications. Briefly, leaves (50-100g) from three individuals were randomly taken from three different tiers according to the plant morphology (Upper tier, young leaves; mid-tier, mature and fully developed leaves and lower tier, senescent leaves or necromass). Two sets of leaves were taken from each individual, one for the epiphyte community analysis and one for the endophyte community. Roots (1-5g) were taken from two different plants with a sterile scalpel (Figure S1).

Epiphyte, Necromass and Root DNA Extraction

We obtained epiphyte DNA by first releasing bacteria from the surface of leaves, as previously reported (Bodenhausen et al 2013, Izhaki et al 2013, Zhang et al 2010), by submerging 10-20g of healthy plant tissue in 100mL of Release buffer (0.1M Potassium Phosphate, 0.1% Glycerol, 0.15% Tween 80, pH 7.0), and vortexing for 7 minutes. The remaining bacteria were dislodged from the leaves with the help of a sterile swab and the buffer was then filtered through a 0.2µm-pore filter. For the roots and necromass samples, epiphyte and endophyte DNA extractions were combined. Briefly, the tissue was fragmented into 0.5-1cm and transferred to a 50mL tube containing 25mL of Release buffer (0.1M Potassium Phosphate, 0.1% Glycerol, 0.15% Tween 80, pH 7.0). The tissue was then homogenized by vortexing for 10 minutes, the buffer was filtered through a 0.2µm-pore filter and the filters were used for DNA extraction using the PowerSoil DNA Isolation Kit (MOBIO Laboratories, Carlsbad, CA; USA), following the manufacturer's instructions. All DNA extractions were quantified using Qubit® 2.0 Fluorometer (Life Technologies Corporation, Carlsbad, CA; USA). In total, we obtained six DNA extractions for the

epiphyte fraction, corresponding to the upper tier and the mid-tier from all three replicates, three DNA extractions for the necromass tier, one for each of three replicates, and two for the roots.

Endophyte DNA Extraction

For the isolation and DNA extraction of endophytic microorganisms, we followed previously reported methodologies (Araujo et al 2002, Conn and Franco 2004, Miles et al 2012). Briefly, before DNA extraction, the plant tissue was surface-sterilized by first thoroughly washing with sterile H_2O to remove dirt, suspended in NAP buffer (124mM $Na_2HPO_4.H_2O$) and sonicated for 1 min to dislodge epiphytes. Leaves were then shaved to remove the pubescence on their surface in order to help in the subsequent sterilization process (Miles et al 2012). The material was then washed with sterile H_2O, submerged in 90% ethanol (60s), 5.25% sodium hypochlorite solution (6 min), and 70% ethanol (30s), and finally rinsed with sterile distilled water. Sterilization was checked by taking an imprint of the leaf on Malt Extract Medium (MEA) (Miles et al 2012) and incubating at 25ºC. 1g of the previously treated material cut into 0.1-0.5mm sections was placed in a 1.5mL Eppendorf tube containing 1g of sterile 0.1mm-diameter glass beads and 1mL of TE (10mM Tris, 10mM EDTA, pH 8.0), homogenized in a mini-bead beater for 5min (BioSpec Products), and used for DNA extraction using the PowerSoil DNA Isolation Kit (MOBIO Laboratories, Carlsbad, CA; USA), following the manufacturer's instructions. DNA extractions were quantified using Qubit® 2.0 Fluorometer (Life Technologies Corporation, Carlsbad, CA; USA). In total, for the endophyte fraction we obtained six DNA extractions corresponding to the upper tier and the mid-tier from all three replicates.

16S rRNA gene Amplification and Sequencing

The V5-V6 hypervariable region of the 16S rRNA gene of Bacteria and Archaea was amplified with primer 799F (5'-AACMGGATTAGATACCCKG-3'), which minimizes contamination from chloroplast DNA and amplifies a mitochondrial product that is larger and thus easier to separate and differentiate from the microbial amplified products (Chelius and Triplett 2001), and the reverse primer 1050R (5'-AGYTGDCGACRRCCRTGCA-3') (Bohorquez et al 2012). DNA concentration was adjusted to 10ng for the endophyte fractions and 1ng for the epiphyte fraction. Each 25uL PCR reaction contained: 10ng (endophytic fraction) or 1ng (epiphytic fraction) of DNA, 10X AccuBuffer (600mM Tris-HCl, 60mM $(NH_4)_2SO_4$, 100mM KCl, 20mM $MgSO_4$, pH 8.3; Bioline), 10mM dNTP Mix, 0.5uM of each primer and 5 Units of ACCUZYME DNA Polymerase (Bioline USA Inc.; Taunton, MA; USA). Cycling conditions were 94ºC for 2 min, followed by 30 cycles of 94ºC for 30 sec, 55ºC for 30 sec, 72ºC for 45 sec; and a final extension of 72ºC for 10 min. Given the size difference, about 800bp and 300bp for mitochondrial and microbial amplicons, respectively, the PCR products were separated on a 3% agarose gel and purified using the QIAquick gel extraction kit (QIAGEN, Valencia, CA; USA). The remaining samples were purified using the QIAquick PCR Purification Kit (QIAGEN, Valencia, CA; USA). DNA was quantified using Qubit® 2.0 Fluorometer (Life Technologies Corporation, Carlsbad, CA; USA) and amplicons were sequenced on an Illumina MiSeq machine with a Paired-End protocol (PE-250) (Molecular Research MR DNA (www.mrdnalab.com, Shallowater, TX, USA).

Sequence Analysis

In order to increase the length of the reads and therefore our ability to accurately assign taxonomy and function, the forward and reverse reads obtained by Illumina paired-end sequencing were initially assembled using the fastq-join software (Aronesty 2013), which combines paired-end reads that overlap into a single contig. We used a minimum overlap of 150bp given the average length of our reads (~250bp)

and the total length of the V5-V6 region (~300bp). Assembled paired-end reads were then analyzed following a pipeline that includes UPARSE (Edgar 2013) and Qiime 1.8 (Caporaso et al 2010b). The reads were filtered by quality first at a mean quality level >25 and then based on the maximum expected error (<0.5) (Edgar 2013). Reads were then separated according to the barcodes (one barcode for each sample) and those with mismatches in the primer or barcode, as well as sequences containing ambiguous bases were excluded. Operational Taxonomic Units (OTUs) were picked at 97% sequence similarity level using the UPARSE-OTUref algorithm, which includes a *de novo* and reference-based chimera-checking step. Taxonomy assignment was performed using the uclust-based consensus taxonomy classifier as default in Qiime 1.8, using the Greengenes reference database and taxonomy (Version 13_8) (McDonald et al 2012). The phylogenetic overlap at 97% between samples was computed using the core microbiome script in Qiime 1.8. Sequences that were assigned to "chloroplast" or "mitochondria" were removed from the dataset. For tree-based analysis, representative sequences of each OTU were aligned against the Greengenes core dataset (DeSantis et al 2006) using PyNAST (Caporaso et al 2010a). The approximately-maximum-likelihood phylogenetic tree was built using FastTree (Price et al 2010).

Diversity Analysis

Alpha- and Beta-diversity analyses were performed using both Qiime 1.8 (Caporaso et al 2010b) and the R package Phyloseq (McMurdie and Holmes 2013). OTU-accumulation curves, relative abundance plots and alpha richness measurements (Chao1) were estimated for all tiers. For diversity estimates, we computed the Shannon entropy index and the Gini-Simpson index. Diversity indices were transformed to effective number of species for comparison among samples using the formulas provided by (Jost 2007). In order to measure and compare the diversity among plant tiers and microbial populations (i.e. Beta-diversity), we normalized the OTU tables using the variance stabilization function from DESeq2 (Love et

al 2014) wrapped in phyloseq (McMurdie and Holmes 2013), as previously reported (McMurdie and Holmes 2014). Normalized tables were then used to estimate the differential abundance of OTUs between samples, location within the plant and microbial population using DESeq2 (Love et al 2014). Principal Coordinates Analysis (PCoA), Correspondence Analysis (CA) and hierarchical clustering were computed based on UniFrac distances. Alpha and Beta diversity plots were made using Phyloseq and SPSS Statistics 22 (IBM, New York, USA).

Functional Analysis

Functional metabolic pathways were predicted using the recently published software PICRUSt (Langille et al 2013) after performing an additional closed reference OTU picking pipeline using Qiime 1.8 since PICRUSt exclusively uses the GreenGenes database IDs for the analysis. The output of this analysis, consisting of KEGG orthologies, was further processed using HUMAnN (Abubucker et al 2012), which transforms 16S rRNA-based predictions into gene and pathways summaries. This summary was then used as input for the Galaxy web-based application (Blankenberg et al 2010, Giardine et al 2005, Goecks et al 2010) for visualization using GraPhlAn. Total metagenomic DNA was also analyzed using the GeoChip 5.0 functional microarray (He et al 2010) (Glomics (Norman, OK; USA). The normalized hybridization output was organized based on functional categories, singletons were removed (probes that were detected in only one sample of a replicate group), and data were analyzed using the multivariate statistical software package PRIMER-E v6 (Plymouth Marine Laboratory). Nonmetric multidimensional scaling (NMDS) ordinations were used to visualize Bray-Curtis similarities. ANOSIM analysis was used to assess the confidence in the similarities observed. Functional categories in each sample were compared using ANOVA and plotted using SPSS Statistics 22 (IBM, New York, USA).

Results

Changes in phylogenetic composition and structure

In order to assess the phylogenetic composition of microbial communities associated with *Espeletia* sp., we sequenced the V5-V6 16S rRNA region from different tiers of the plant (Figure S1). A total of 17 DNA samples yielded 3,041,094 16S rRNA paired-end sequences. Sequences were assembled into single reads, de-multiplexed and quality-filtered using the UPARSE (Edgar 2013) and QIIME 1.8 (Caporaso et al 2010b) pipelines, yielding a total of 1,762,044 high-quality sequences. These were clustered into OTUs at 0.03 distance cutoff with the UPARSE-OTUref algorithm, which can infer up to the species level (Schloss and Handelsman 2005) and includes a *de novo* and reference-based chimera checking process that removes spurious OTUs and sequencing artifacts (Edgar 2013). Although the forward primer we used (799F) was reported to exclude chloroplast and mitochondrial DNA (Chelius and Triplett 2001), these sequences were found in our data and were therefore removed from the OTU table. After this chimera checking and filtering processes, we obtained a total of 1,548 OTUs represented by sequence reads ranging from 6,744 to 102,266 per sample. Chao1 richness estimates were clearly higher than the number of OTUs observed, with significant differences observed for the root and necromass samples (Table S1).

OTU-accumulation curves (Figure S2) showed that most samples were well characterized with our sampling efforts. This was the case for the epiphyte leaf communities and for the necromass and root samples, the latter two including combined endophyte and epiphyte communities. However, the endophyte leaf communities were less characterized and thus, more sampling and sequencing efforts should be performed. The low number of sequences obtained for the endophyte community could be due to the fact that these samples contained many plant chloroplast and mitochondrial sequences, and that prokaryotes might represent only a small fraction of endophytes (Miles et al 2012). The calculated Shannon and Simpson diversity indices, which take into account OTU abundances (Tuomisto 2010) (Table S1), were

transformed into effective numbers of species (ENS) to compare samples (Jost 2007). Both Shannon and Simpson transformations showed that diversity was significantly higher for the necromass and root fractions (p = 0.000061 and p = 0.000005, respectively) (Figure 1), even when comparing against the combined leaf endophyte and epiphyte fractions (Figure S3).

Taxonomy of the sequences was examined at the phylum level using the uclust-based consensus taxonomy classifier (Figure 2). The most abundant phyla in all plant samples were *Proteobacteria*, *Actinobacteria*, *Bacteroidetes*, *Acidobacteria*, *Firmicutes* and *Crenarchaeota* even when subsampling at the lowest number of reads obtained, 6,700 sequences per sample (data not shown). The phylum *Proteobacteria* was more abundant in endophyte and epiphyte communities of both young and mature leaves (from 71.65% to 85.76%) compared to the necromass fraction (51.94% and 45.40%, respectively). Within this phylum, there were marked differences in relative abundances across samples at the class level, such as the greater abundance of *γ-proteobacteria* in young and mature leaves (both endophytes and epiphytes) when compared to the root or necromass tiers (Figure 2). On the other hand, the necromass and root fractions had a higher relative abundance of the phylum *Acidobacteria* (11.81% and 20.28%, respectively) when compared to the leaf habitat (2.40% to 3.66%), and there were more *Bacteroidetes* and α-*proteobacteria* in the necromass fraction (21.05%) and *Crenarcheaota* in the root fraction (13.93%). *Actinobacteria* showed a relative abundance ranging from 4.07% to 12.02% across samples.

Although we found a total of 1,548 OTUs, there was a marked difference in OTU abundances across samples (Figure S4). The most common genera shared by all samples were OTUs classified as *Pseudomonas* sp., *Klebsiella* sp., *Sphingomonas* sp., *Erwinia* sp., *Baumannia* sp., *Propionibacterium* sp., *Burkholderia* sp., *Acinetobacter* sp., and *Hymenobacter* sp. Not surprisingly, these OTUs belonged to the

core community, which was composed of 174 OTUs shared by all samples (Figure S5). Despite the low percentage of unclassified OTUs (< 4% across samples), there were many OTUs that could not be classified up to species or even genus levels, suggesting that they could not be appropriately classified with the database we used or that they might represent new microorganisms.

The phylogenetic structure of all samples was analyzed by a Principal Coordinates Analysis (PCoA) using weighted UniFrac distances (Figure 3) in order to assess if the samples clustered according to their location on the plant. The root fraction was separated from the leaf phyllosphere (leaf endophyte and epiphyte) and necromass samples, which in turn were also separated from one another. All leaf samples clustered together, indicating that they were similar and differed from the necromass and root fractions as showed using hierarchical clustering (Figure S6). This analysis indicated that the epiphyte and endophyte leaf communities were very similar and, indeed, we only found few OTUs, belonging to *Acidobacteria*, *Bacteroidetes* and *Proteobacteria,* which were more abundant in the epiphyte community compared to the endophyte community (Figure S7). Although the necromass is part of the phyllosphere, we observed clear differences between leaf epiphyte and endophyte communities and the necromass fraction. The necromass fraction had more abundant OTUs from the *Acidobacteria, Actinobacteria, Bacteroidetes, Firmicutes, Proteobacteria,* FBP and TM7 phyla. The differences observed for the root fraction were mainly driven by OTUs from the phyla *Actinobacteria, Acidobacteria, AD3, Armatinomonadetes, Bacteroidetes, Chlamydiae, Chloroflexi, Crenarchaeota, Firmicutes, Proteobacteria, TM6, TM7* and *Verrucomicrobia*. These phyla were significantly more abundant in the root fraction compared to the phyllosphere communities (i.e. leaves and necromass, $p < 0.01$) (Figure S7). Interestingly, *Rothia sp.*, *Corynebacterium sp.* and *Pseudomonas sp,* were more abundant in both the epiphyte and endophyte fractions compared to the necromass and root fractions. In general, the differences between the necromass

and leaf fractions were driven by OTUs that were more abundant in the necromass and belonged mainly to *Proteobacteria, Acidobacteria, Actinobacteria, Bacteroidetes, Firmicutes*, FBP and TM7. The root fraction, on the other hand, showed overall increased abundances in OTUs from several phyla, mentioned above, when compared to the rest of the fractions (Figure S7). This is consistent with the aforementioned higher diversity measurements of the root fraction.

Metabolic profiles vary according to plant microenvironment

To further understand how *Espeletia* sp. plant microhabitats can affect microbial communities, we analyzed the samples using a bioinformatics 16S rRNA-based prediction approach and a functional microarray. We predicted the metabolic potential of each of the communities with the PICRUSt software, which infers the metabolic capabilities by matching 16S rDNA sequences with the closest sequenced genome (Langille et al 2013). Based on this information, we constructed a tree depicting the main metabolic pathways, categorized by the KEGG database (Kanehisa and Goto 2000) using the web-based application Galaxy (Blankenberg et al 2010), combined with a heat map of gene "abundances" for each category in every sample (Figure S8). The functional categories and the abundances varied among samples, showing widespread abundance in genes involved in xenobiotic degradation, carbohydrate, lipid, amino acid and cofactors metabolism, biosynthesis of secondary metabolites and energy metabolism. Interestingly, genes involved in the replication and repair pathways, including base excision repair system, DNA replication, homologous recombination, mismatch repair and nucleotide excision repair systems, were also predicted to be present in the samples. The degradation of xenobiotics and the biosynthesis of secondary metabolites such as antibiotics, alkaloids, polyketides and terpenoids were also a relevant feature of this prediction (Figure S8). Although the 16S-based predictions did not show any clear separation between the communities, we found statistically significant differences between the endophyte

community and the root and necromass communities (ANOSIM, Global R = 0.359, p = 0.008, n = 17, permutation = 9,999; Figure 4A).

To obtain information regarding the actual functional profiles of these microbial communities, we used the GeoChip 5.0 functional microarray, which contains over 167,000 probes covering more than 395,000 coding sequences from approximately 1,500 functional gene families involved in several biogeochemical, cellular and ecological processes. All microbial communities had great functional diversity, showing mainly genes involved in carbon fixation and degradation pathways, nitrogen, phosphorus and sulfur metabolism, organic remediation, secondary metabolism, virulence-related genes and environmental stress responses. There were statistically significant differences between the communities analyzed as reported by the Analysis of Similarity (ANOSIM) (Global R = 0.524, p = 0.002, n = 16, permutation = 9,999). Pairwise comparisons showed that the epiphyte, root and necromass communities shared greater similarity, while the endophyte communities were significantly different from the rest of the communities ($p < 0.05$). Compared to the ordination analysis performed on the 16S-based predictions, which could not separate samples based on the population (endophyte, epiphyte, necromass and root) or plant tier, the GeoChip analysis separated the endophyte communities from the rest (Figure 4). Even though several genes families were detected, in this study we were most interested in functions that could provide insight regarding survival strategies and adaptation to the harsh, high mountain plant environment of *Espeletia* sp.

Nutrient Utilization

The GeoChip analysis identified several genes related to carbon cycling in all communities associated with the plant. Autotrophy capacity was evidenced by the presence of genes involved in six carbon fixation

pathways (Figure 5). These pathways were observed across the plant with an overall lower abundance in the endophyte fractions ($p < 0.05$), indicating the capacity of fixing carbon in all parts of the plant. The Calvin cycle, driven mainly by microorganisms belonging to *Proteobacteria* and *Cyanobacteria*, was the most abundant cycle in all samples. Archaeal pathways were also present in the samples, although at low abundances (the 3-hydroxypropanoate/4-hydroxybutyrate cycle and the dicarboxylate/4-hydroxybutyrate cycle). The remaining cycles, relatively more abundant than the archaeal driven ones, were due mainly to genes present in *Proteobacteria* members, which were highly abundant in our samples. Interestingly, we observed the presence of genes involved in the formation of carboxysomes (bacterial microcompartments that contain enzymes involved in carbon fixation) across all samples with a relative abundance that matched the distribution of the carbon fixation pathways. We also identified C1 metabolism pathways involving methanogenesis and methane oxidation (Figure 5). Methanogenesis was more abundant in the epiphyte and root fractions while methane oxidation performed by *Gamma-* and *Alphaproteobacteria* microorganisms seemed to be less abundant.

Heterotrophic metabolism was also found within our samples. The pathways involved in carbon degradation were highly abundant in all samples, with an overall increased abundance in the epiphyte, necromass and root fractions ($p < 0.05$, Figure S9). Functions ranged from the degradation of labile carbon sources such as starch (which was the most abundant pathway overall) to the metabolism of recalcitrant sources such as lignin and other plant-derived sources such as cellulose, hemicellulose, pectin and terpenes. The ability to degrade carbon sources broadened with the presence of genes involved in remediation of organic compounds (Figure S10). Genes involved in the degradation of aromatic compounds were in fact higher than those involved in “common” carbon metabolism such as cellulose, hemicellulose and pectin in terms of relative abundance. The high abundance of carbon degrading

pathways was due to the presence of genes encoded mainly by *Proteobacteria*, *Actinobacteria* and *Firmicutes*.

Nitrogen cycling pathways were also present in all fractions of the plant, indicating microbial capacity to transform this element (Figure 6). Although we found the same distribution of genes in the samples (lower abundance in the endophyte fraction ($p < 0.05$) in this case, the root fraction had increased abundance in assimilatory nitrate reduction, denitrification, nitrification, and nitrogen fixation. On the other hand, ammonification and anammox pathways were more abundant in the necromass and epiphyte fractions. Most genes in nitrogen cycling pathways were derived from uncultured bacteria (data not shown) with a small presence of *Actinobacteria, Proteobacteria, Firmicutes, Verrucomicrobia¸* among others.

Stress responses and secondary metabolism

The functional analysis showed presence of genes involved in the response to environmental conditions such as acidic and alkaline shock responses, cold and heat shock, glucose, phosphate and nitrogen limitation, osmotic stress, oxidative stress, oxygen limitation and stringent response (Data not shown). Given the high incidence of UV radiation at high Andean mountain locations (Buytaert et al 2006), we looked at radiation resistance mechanisms of these communities (Figure 7). We found genes involved in the production of bacterial pigments that can aid in the protection against the DNA-damaging doses of UV radiation, as well as genes for enzymes involved in the removal of reactive oxygen species (ROS). These mechanisms were associated with several phyla (*Crenarchaeota*, *Euryarchaeota*, *Actinobacteria*, *Firmicutes, Bacteroidetes, Cyanobacteria, Deinococcus-Thermus, Planctomycetes, Proteobacteria* and *Verrucomicrobia* as the most abundant). Finally, we found genes involved in the production of antibiotics mainly in *Actinobacteria* and *Proteobacteria*, across all samples. Moreover, there was a remarkable

abundance of antibiotic resistance genes across several phyla in all samples as well as the presence of plant hormone production genes.

Discussion

Microbial populations that reside in close contact with plants can be found either in the rhizosphere, which is the most commonly studied microhabitat, or in the phyllosphere either as epiphytes on the surface or as endophytes inside plants. The plant phyllosphere, although not considered as extreme as other environments, presents several challenges for microbial colonization and growth, especially in high mountain neo-tropical ecosystems exposed to sudden temperature changes and high incidence of UV radiation. In this work we analyzed for the first time the microbiota associated with *Espeletia* sp., a plant endemic to the Páramo environment of the Andes mountains, and a unique model for studying microbial populations in this harsh and threatened ecosystem. Microbial communities associated with *Espeletia* sp. were tested for both taxonomic composition and functional profiles, thus improving on traditional biodiversity surveys by including novel host-associated microbiomes and their putative functional ecology that can give insight regarding strategies for plant colonization and microbial survival.

The overall composition of the *Espeletia* sp. microbial communities showed abundance of the phyla *Proteobacteria, Actinobacteria* and *Bacteroidetes,* similar to what has been reported for microbiomes associated with plants such as *Arabidopsis thaliana* (Bodenhausen et al 2013), *Thlaspi geosingense* (Idris et al 2004), potato cultivars (Manter et al 2010), and several tree species (Kembel et al 2014, Redford et al 2010). Compared to these tree species where phyla TM7 and *Deinococcus-Thermus* were highly abundant (Redford et al 2010), in *Espeletia* sp. they showed an overall lower abundance. The phylum *Proteobacteria* has also been found to be the most abundant on almond drupes (McGarvey et al 2014).

Acidobacteria and *Crenarchaeota* were especially abundant in the root and necromass fractions, but have been reported to be rare in the phyllosphere and rhizosphere of *A. thaliana* (Bodenhausen et al 2013), suggesting that they may be playing an important functional and ecological role when associated with *Espeletia* sp. The different plant tier communities showed similarities in terms of the taxa identified and they shared several OTUs, indicative of a core community that, in addition, shared several genera with the core community of *A. thaliana* (i.e. *Pseudomonas, Burkholderia, Sphingomonas, Flavobacterium*, among others). However, we observed differences with the most abundant core-community OTUs from several tree species (*Beijerinckia sp., Leptothrix sp., Stenotrophomonas sp.*, and *Niastella sp.*), which were very rare in our samples. This overlap in the most abundant taxa in diverse plant species might indicate that these communities are widely spread across the globe and are adapted to the life in association with plants and the differences observed are probably driven by their relative abundances that reflect adaptations to different plant microbiomes and their interactions.

Although we expected great phyllosphere microbial diversity, given the environmental conditions driving evolution and diversity (Li et al 2014), both necromass and roots were more diverse in terms of bacterial composition and functional potential. There are several factors involved in shaping communities and the size of the root microbiome is one of them (Bodenhausen et al 2013). Other factors playing an important role in the survival of microorganisms is the availability of nutrients. The surface of leaves is considered a water- and nutrient-limited environment (Vokou et al 2012), which contrasts with the senescent material of the necromass, containing perhaps more readily-available nutrients, and the more stable and nutrient rich root micro-environment (Bodenhausen et al 2013, Lambers et al 2009). We also expected distinct endophyte and epiphyte communities in the plant, but we observed otherwise. The endophyte and epiphyte communities shared most OTUs and were not differentiated according to the taxonomic analysis.

Compared to *A. thaliana*, where the endophyte leaf community was richer than the epiphyte, we did not observe any differences in richness or diversity between these fractions, suggesting that they are somehow continuously communicated. In addition, the presence of shared taxa among the various plant tiers sampled here also indicates that bacteria can travel from the rhizosphere towards the leaves and vice versa, as has been reported previously (Afzal et al 2014, Beattie and Lindow 1999, Bodenhausen et al 2013). Interestingly, we did not find any bacterial taxa present exclusively in any of the plant communities sampled; differences were due mainly to changes in abundances of bacterial taxa. Some taxa were more abundant is specific samples, such as *Hymenobacter* sp. in the necromass fraction, which has been reported in grassland soils (Kim et al 2008), desert soils (Zhang et al 2009) and has even been isolated from air (Buczolits et al 2002). Representatives of *Crenarchaeota*, which were more abundant in the root fraction, have also been previously reported in tomato roots (Simon et al 2005).

Very little is known about the "global" metabolic capabilities of microorganisms associated with plants, yet a microbial community's capacity to grow under specific environments should be evident in its functional profile. The 16S rRNA-based prediction of the metabolic potential of microbial communities associated with *Espeletia* sp. showed a very diverse metabolism that included metabolic pathways important for microbial survival in this habitat, such as acquisition of nutrients, biosynthesis of secondary metabolites, and DNA repair systems important for repair of radiation damage (Vorholt 2012). This metabolic potential was further analyzed using a GeoChip 5.0 functional microarray, which has been used to assess the metabolic pathways in other ecosystems (Bai et al 2013, Chan et al 2013, He et al 2010, Wang et al 2009, Xie et al 2011). This analysis again indicated a metabolically diverse and versatile plant microbiota and a general agreement with some of the functions identified using bioinformatics alone. Nonetheless, this microarray approach effectively identified specific pathways related to nutrient cycling,

stress responses, and survival mechanisms, among others. In this case, the endophyte communities were separated from the other communities (epiphytes, necromass and root), mainly due to the overall lower abundance of genes in all metabolic pathways for the endophyte fraction.

Nutrient acquisition, in particular the capacity to obtain carbon and nitrogen, is vital for microbial survival, especially in harsh environments such as the phyllosphere of *Espeletia* sp. Carbon acquisition, which is paramount for microbial growth, was evident by the presence of genes for autotrophic and heterotrophic carbon metabolism. Autotrophy was mainly represented by genes involved in the Calvin cycle, driven by *Cyanobacteria* and *Proteobacteria* throughout the plant, which also correlated with the presence of genes for carboxysome formation in several taxa, a bacterial micro compartment which may improve carbon fixation processes. These could be due to *Proteobacteria*, abundant throughout the plant, and *Cyanobacteria*, which were not directly observed in our samples due to the primers used that exclude this taxon (Bodenhausen et al 2013). The remaining carbon fixation pathways were found at relatively low abundance compared to the Calvin cycle, probably due to the anaerobic or microaerophilic conditions they require, which in the plant microenvironment, could only be achieved in biofilm-like structures. Furthermore, C1 metabolism pathways were also present in our samples, specifically methanogenesis whose abundance was comparable to the carbon fixation pathways. C1 pathways are mostly performed by Archaea but we found genes involved in this pathways in other phyla, such as *Proteobacteria*, *Cyanobacteria*, *Chlorobi* and *Firmicutes*. This metabolism is feasible in the plant environment since all precursors are sub-products of cellular metabolism (e.g. acetate from cellulose degradation) and the only limiting factor is the low oxygen concentrations required that, once again, could occur in biofilm structures. Such structures might be provided by microorganisms such as *Pseudomonas sp.* that were highly abundant in the leaf communities compared to the necromass or roots and have been long

recognized as plant colonizers (Bodenhausen et al 2013, Hirano and Upper 2000) and biofilm forming microorganisms (Danhorn and Fuqua 2007). Moreover, methanogenesis has been recently reported to occur in the presence of oxygen in plants by not-yet-identified mechanisms (Keppler et al 2006), raising questions regarding the role of microorganisms in this "aerobic" methane production. On the other hand, methane produced by methanogenic bacteria, besides being used by methanotrophs as carbon source, can be coupled to the nitrogen cycle, as has been reported recently (Ettwig et al 2010, Haroon et al 2013).

Although autotrophic metabolism was detected, the abundance of carbon degrading genes overshadowed the autotrophy-related ones and indicating degradation of plant and insect carbon sources that could be carried out by core community members. These include *Pseudomonas* sp., a common inhabitant of the phyllosphere either as a pathogen or as potential beneficial agent (Enya et al 2007, Hirano and Upper 2000). *Klebsiella* sp. associated with roots (Cakmakci et al 1981), *Propionibacterium* sp., which has been found in potato plants (Rasche et al 2009) and grapevines (Campisano et al 2014), and *Bacillus* sp, which has been reported on lettuce phyllosphere (Rastogi et al 2012), all of which were involved in the degradation of hemicellulose, chitin, starch, pectin and lignin. *Actinobacteria*, which are known for being metabolically versatile microbes that are abundant in soils and can interact with plants (Coombs and Franco 2003, Crawford et al 1993), showed a remarkable abundance of genes involved in carbon degradation. The necromass fraction, where we expected to have the greatest abundance of carbon degrading genes, had significant abundance of *Pedobacter sp.*, *Spirosoma sp.*, *Sphingomonas sp,* and *Methylobacterium sp.*, which indeed showed several genes involved in the degradation of carbon sources. Interestingly, the root community had an increased abundance *Candidatus Koribacter* and *Candidatus Solibacter,* two *Acidobacteria* members that have been isolated from Arctic soils (Rawat et al 2012). These two microorganisms had high abundance of genes related to carbon degradation pathways including the

degradation of starch, hemicellulose, pectin, among others, which correlates with the metabolic versatility previously reported (Rawat et al 2012). Given the chemical relatedness between a recalcitrant carbon source such as lignin and aromatic compounds (Diaz et al 2013), we expected to find genes involved in the degradation of organic aromatic compounds. Genes involved in the degradation of aromatic compounds and other organic contaminants in all samples, coupled with the reported ability of plants to uptake these compounds (Afzal et al 2014) may indicate that this protected natural Páramo environment is starting to be contaminated by human-based activities.

The microbiota of *Espeletia* sp. also exhibited great versatility in terms of nitrogen cycling, another nutrient that is vital for plants and microorganisms alike. The GeoChip results indicated that the microbial communities could perform almost every transformation of this element, and that nitrogen transformations were slightly higher in the root fraction. The different abundances of functional groups throughout the plant samples indicates a possible interaction between microbial communities in the roots and the plant itself, where the first provide nitrogen in a usable form and the latter provide carbon sources as form of nutrients. In the case of leaf communities, *Rothia sp.* and *Corynebacterium sp.* were more abundant compared to the necromass and root fractions. The association of the first two species has been previously reported in different plants such as legumes and herbs (Torche et al 2014, Xiong et al 2013) and their ability to fix nitrogen has been reported (Gtari et al 2012). Although we did not observed genes involved in nitrogen fixation for these species, lateral gene transfer of *nifH* genes has been suggested (Gtari et al 2012).

In addition to nutrient acquisition, microbial communities associated with plants must also survive under the environmental conditions of the particular habitat and, as part of a complex microbial community of

bacteria, fungi, insects and nematodes, they must constantly compete for nutrients and survival. For example, neotropical high Andean plants, and their associated microbiomes, are exposed to high UV radiation. We found several mechanisms for resistance to solar radiation in our samples, such as genes involved in the production of pigments, which may aid in the protection against DNA damage, and genes involved in the protection from Reactive Oxygen Species (ROS). Microbial competition was evident in these microbial communities by the presence of genes involved in antibiotic production, a strategy useful when competing for nutrients and space. Not surprisingly *Actinobacteria* dominated in terms of presence of genes involved in antibiotic production, followed by some members of *Proteobacteria* such as *Pseudomonas* sp. Although these metabolites could be acting as signaling molecules, as suggested (Vaz Jauri et al 2013), the widespread abundance of genes involved in antibiotic resistance across several taxa could also indicate ongoing competition. Finally, the production of plant hormones might suggest an important role for these microorganisms in their association with the plant host. Microbial production of plant hormones has been reported to promote growth and development (Reinhold-Hurek and Hurek 2011) and to prevent the entry of plant pathogens by modulating the plant's immune system (Berendsen et al 2012, Ikeda et al 2010).

The relation between bacteria and host growth and mortality has been previously described (Kembel et al 2014), suggesting that microbial communities may indeed be an important trait in plant development. In this case, the study of these communities in *Espeletia* sp. represents a starting point for understanding plant development and health in ecologically strategic yet threatened ecosystems with little human intervention. Even though leaf phyllosphere communities were similar to one another, they differed from necromass and root microbiota in terms of diversity and relative abundance of taxa, suggestive of niche-selective properties that alter the structure of a community that is nonetheless a continuum of the microorganisms throughout the plant. Given the suggested origin for bacterial communities associated

with plants (air, snow, rain, fog) (Christner 2010, Danhorn and Fuqua 2007, Leveau and Lindow 2001), it is difficult to assess the stability of these communities and perhaps sampling at different months throughout the year could increase the discriminatory capabilities between permanent and temporary residents of the phyllosphere. Despite this difficulty, the fact that the most abundant phyla and bacterial species are shared across samples suggests that there is a widespread "global" core community adapted to the life in the phyllosphere. The functional analysis revealed differences that distinguished endophytic microorganisms and the presence of functional groups relevant for community survival and growth in this particular environment. For example, the presence of genes involved in the degradation of aromatic compounds indicate changing environmental conditions by suggesting that these plants may be more influenced by human intervention than previously thought. Finally, the metabolic potential provides insight regarding plant-microbe interactions and raises the possibility of bioprospecting for microbial processes such as nutrient degradation, remediation and antimicrobial compound production present in the phyllosphere microbiome of high Andean mountain endemic plant species.

Microbial and functional diversity within the phyllosphere of *Espeletia* sp. in Andean High Mountain Ecosystems

Carlos A. Ruiz-Pérez[1,2,3], Silvia Restrepo[2,3] and María Mercedes Zambrano[1,2*]

[1] *Molecular Genetics, Corporación CorpoGen, Bogotá DC, Colombia;* [2] *Colombian Center for Genomics and Bioinformatics of Extreme Environments, Bogotá DC, Colombia;* [3] *Andes University, Bogotá DC, Colombia.*

[*] *Corresponding author*

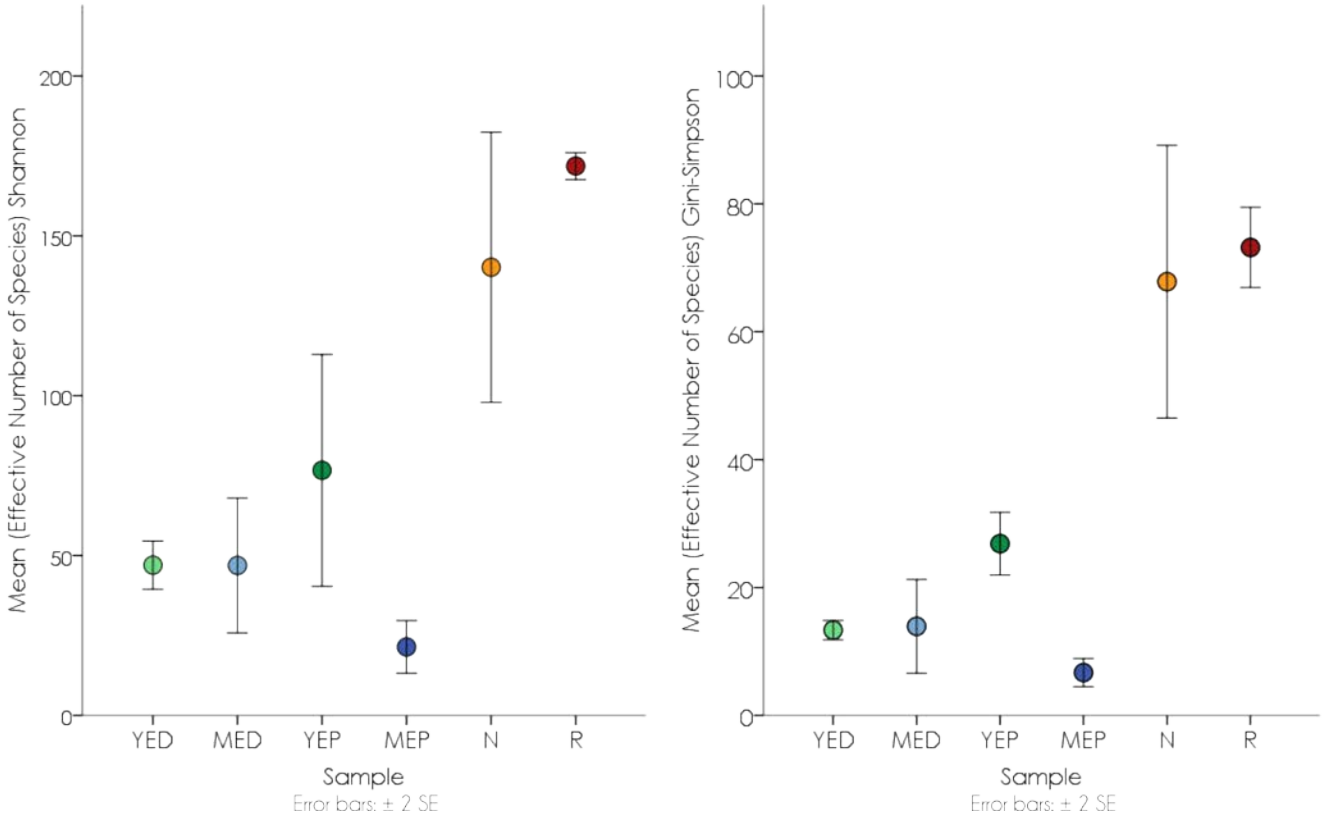

Figure 1. Transformed diversity indices Shannon (A) and Simpson (B). Dots show the mean of the replicates per community (YED – Young Endophyte; MED – Mature Endophyte; YEP - Young Epiphyte; MEP – Mature Epiphyte; N – Necromass; R – Root). Letters represent statistically significant groups calculated using Tukey's *post-hoc* test.

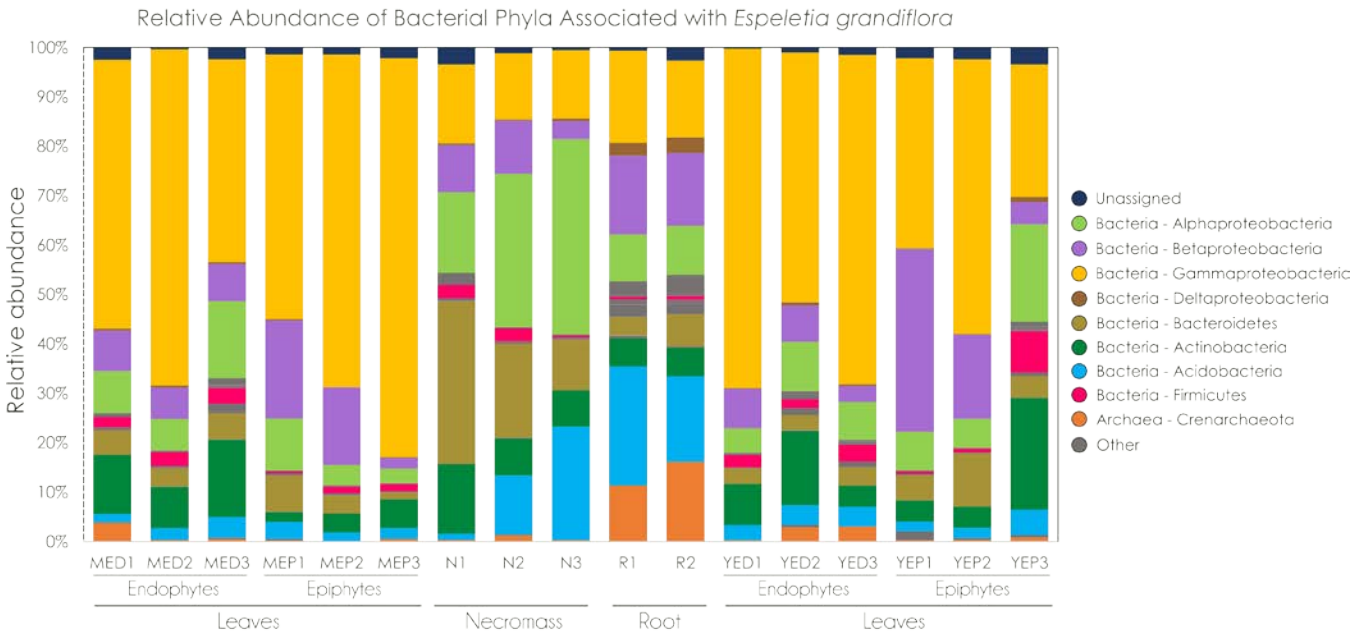

Figure 2. Relative abundance of bacterial phyla associated with *Espeletia* sp. (YED – Young Endophyte; MED – Mature Endophyte; YEP - Young Epiphyte; MEP – Mature Epiphyte; N – Necromass; R – Root). The Proteobacteria phylum has been replaced by the corresponding Classes (Alpha, Beta, Gamma and Delta).

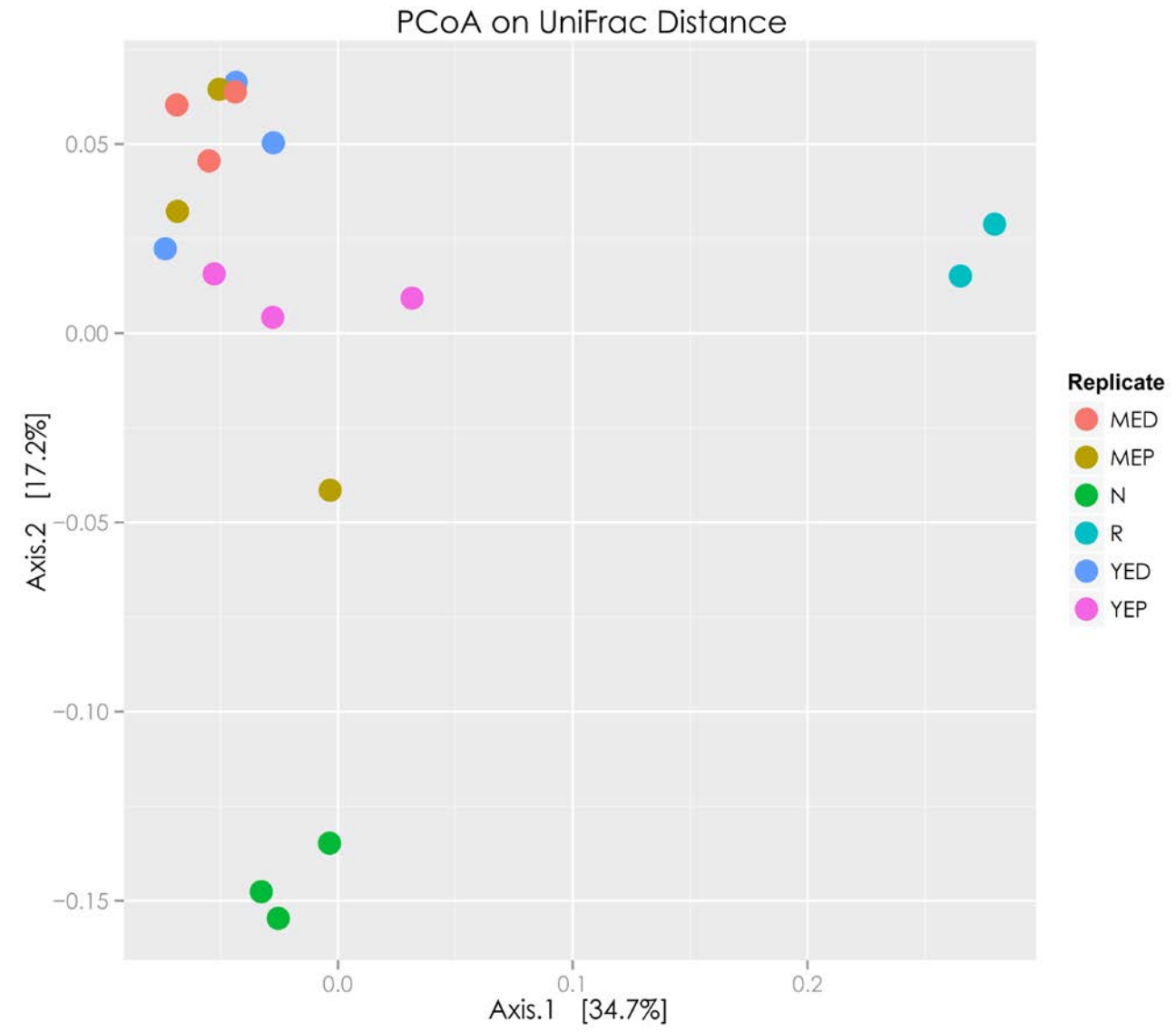

Figure 3. Principal Coordinates Analysis (PCoA) plot based on Weigthed UniFrac distances. The two axes represent 51.9% of the variation in the samples. Individual points represent the replicates for each sample (YED – Young Endophyte; MED – Mature Endophyte; YEP - Young Epiphyte; MEP – Mature Epiphyte; N – Necromass; R – Root).

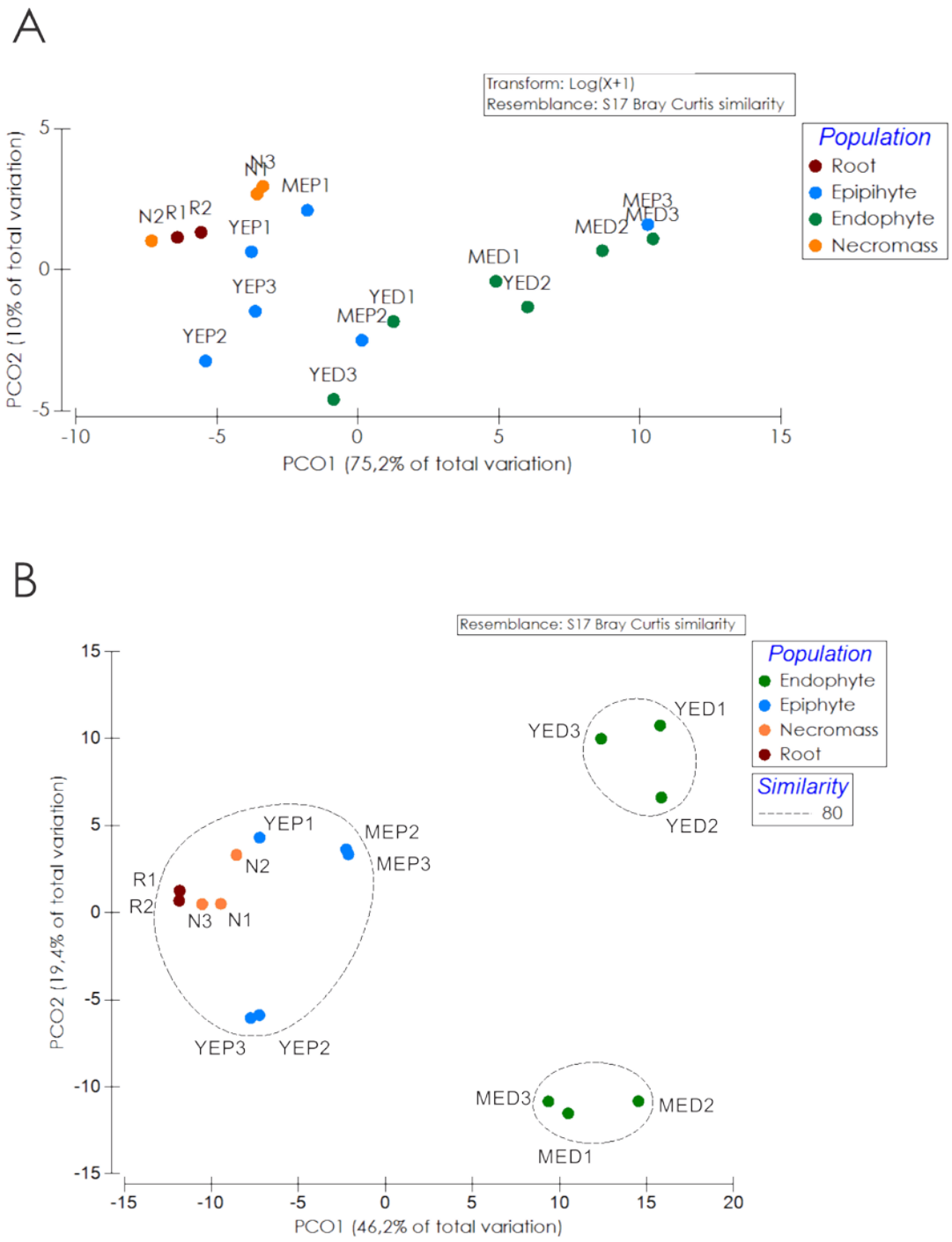

Figure 4. Analysis of microbial metabolic potential. PCoA based on Bray-Curtis similarities were calculated for: A) Metabolic profiles predicted with PICRUSt (the two axes represent 85.5% of the variation), and B) Metabolic profiles derived from the GeoChip analysis (the two axes represent 65.6% of the variation). Individual points represent the replicates for each sample (YED – Young Endophyte; MED – Mature Endophyte; YEP - Young Epiphyte; MEP – Mature Epiphyte; N – Necromass; R – Root).

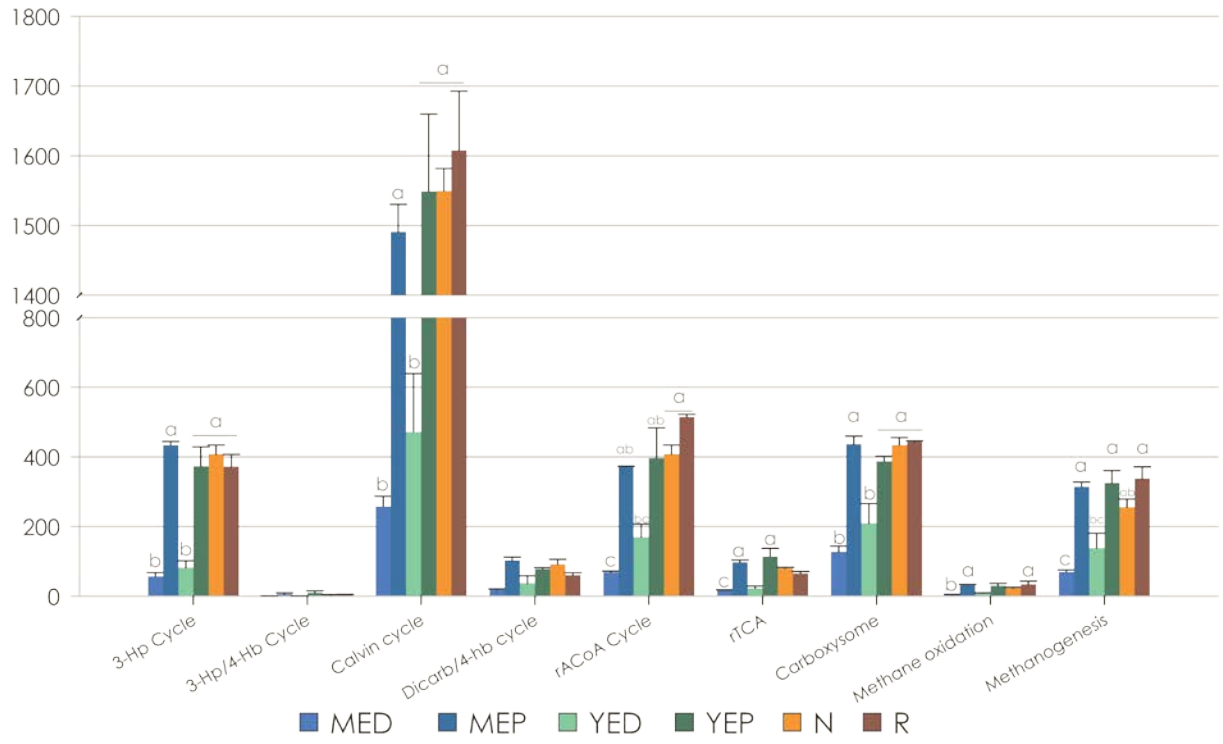

Figure 5. Relative abundance of several carbon-cycling genes clustered by major functional processes: 3-Hydroxypropionate cycle, 3-Hydroxypropionate/4-Hydroxybutyrate cycle, Calvin cycle, Dicarboxylate/4-Hydroxybutyrate cycle, Reductive Acetyl CoA Pathway, Reductive Tricarboxylic Acid cycle, Carboxysome, Methane oxidation and Methanogenesis. The signal intensity was normalized by the mean intensity of the microarray. Mean values of samples were plotted with their respective SE. Variations among communities were tested with one-way ANOVA and letters indicate statistically significant differences ($p < 0.05$) clustered according to the post-hoc Tukey's test.

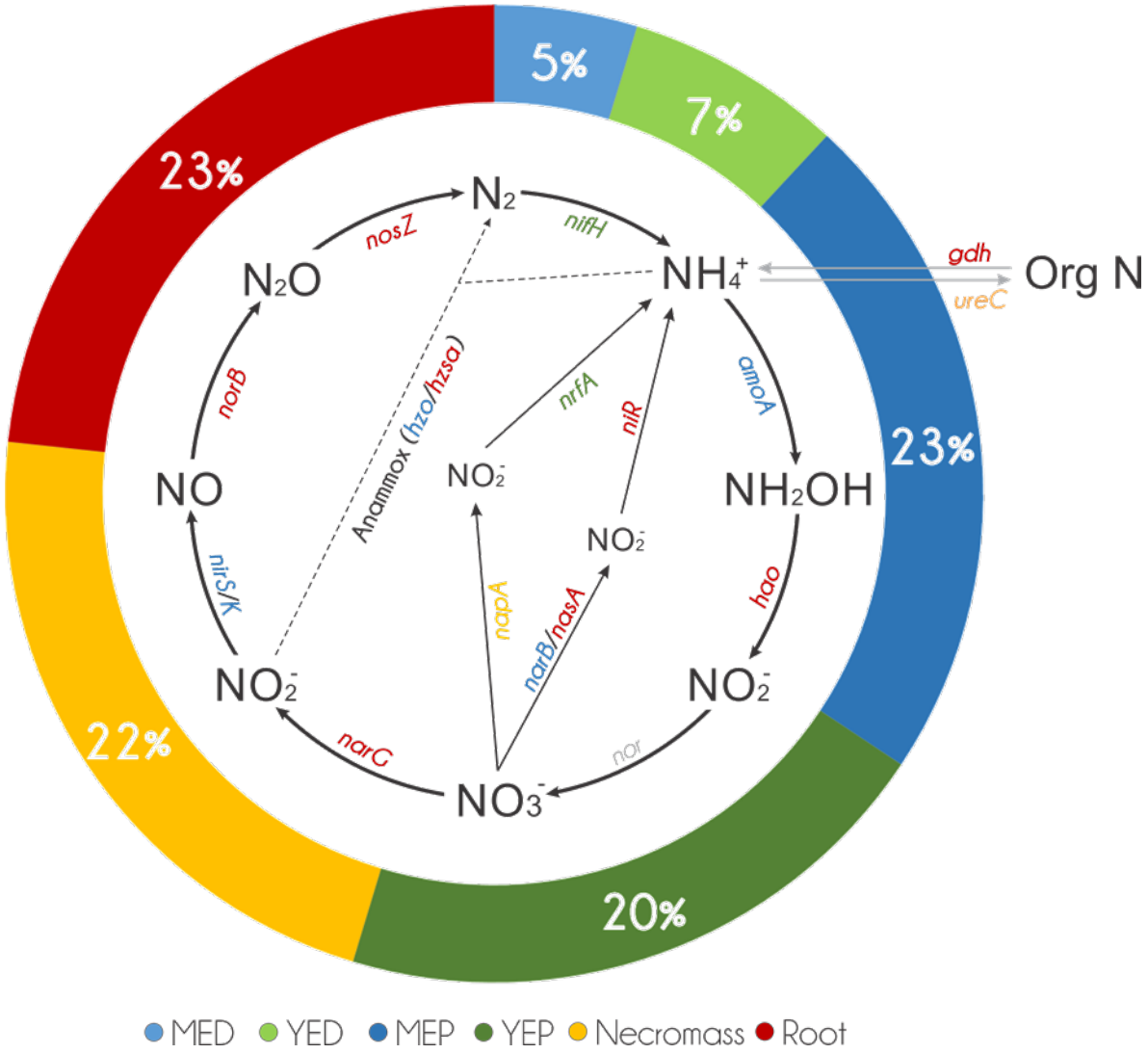

Figure 6. GeoChip predicted nitrogen metabolic transformations present in *Espeletia* sp. microbial communities. Colors in gene names indicate the sample where such gene was relatively more abundant. Outer ring indicates the total relative abundance of genes involved in nitrogen transformations. (YED – Young Endophyte; MED – Mature Endophyte; YEP - Young Epiphyte; MEP – Mature Epiphyte; N – Necromass; R – Root) *Adapted from He *et al*. 2010.

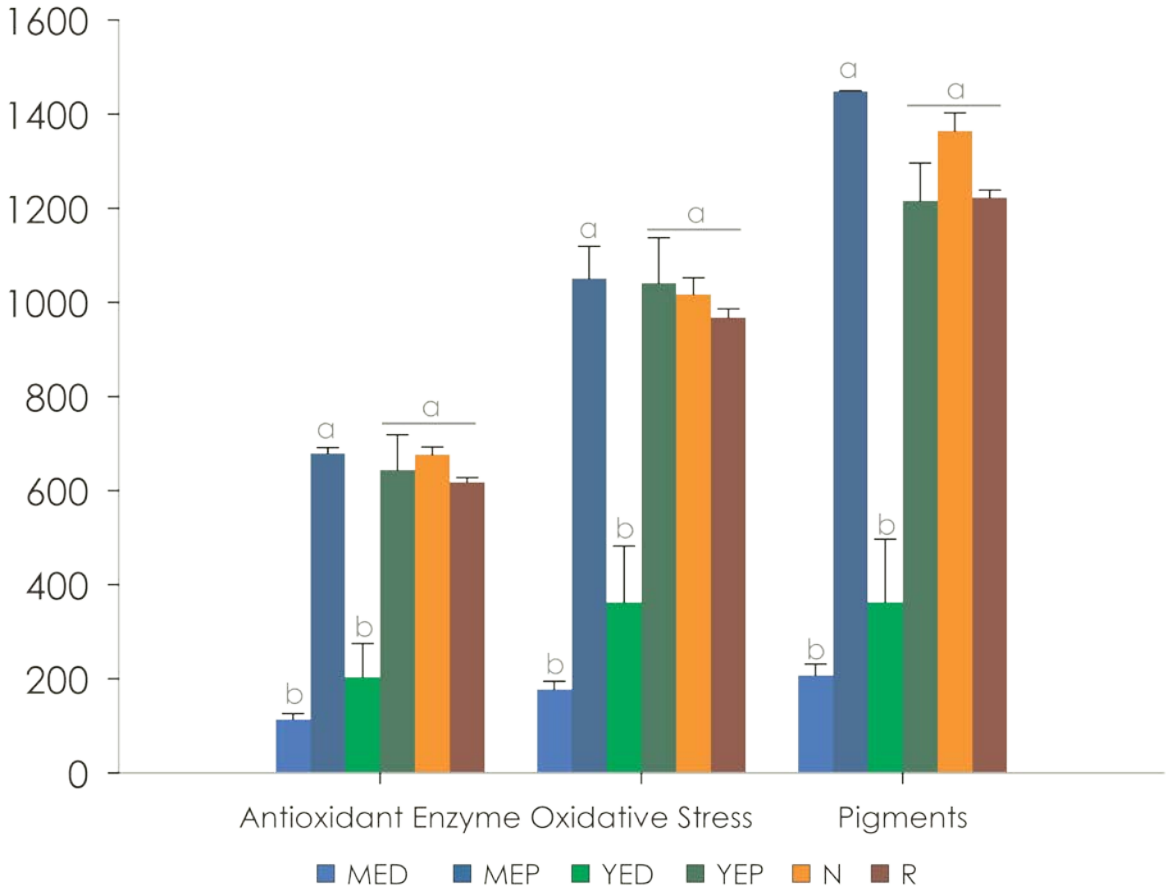

Figure 7. Relative abundance of several UV resitance-related genes clustered by major functions: Antioxidant Enzymes, Oxidate Stress response and Pigment production. The signal intensity was normalized by the mean intensity of the microarray. Mean values of samples were plotted with their respective SE. Variations among communities were tested with one-way ANOVA and letters indicate statistically significant differences ($p < 0.05$) clustered according to the post-hoc Tukey's test. (YED – Young Endophyte; MED – Mature Endophyte; YEP - Young Epiphyte; MEP – Mature Epiphyte; N – Necromass; R – Root).

Supplementary Information

Table S1. Per-sample sequences, observed OTUs, richness estimates, diversity indices and coverage. Samples codes as follows: YED (Young leaves – Endophytes), MED (Mature leaves – Endophytes), YEP (Young leaves – Epiphytes), MEP (Mature leaves – Epiphytes), N (Necromass), R (Roots). Asterisks indicate significant differences ($p < 0.05$).

Sample	Total Sequences	Observed	Richness	Diversity		Coverage
			S_{CHAO}(SE)	Shannon-Weaver Index	Gini-Simpson Index	C_{GOOD}
YED1	25,457	674	910.17(43.03)	3.72	0.92	0.990
YED2	8,472	480	695.58(46.93)	3.99	0.92	0.972
YED3	28,552	693	844.67(29.73)	3.81	0.93	0.993
MED1	14,042	543	722.22(35.27)	3.92	0.95	0.982
MED2	10,997	466	656.94(40.88)	3.30	0.85	0.984
MED3	6,744	485	718.18(46.55)	4.15	0.94	0.966
YEP1	68,833	778	962.20(33.93)	4.03	0.96	0.996
YEP2	58,713	784	995.46(40.16)	4.11	0.96	0.996
YEP3	57,630	948	1057.84(26.04)	4.72	0.97	0.998
MEP1	53,600	729	909.90(34.14)	3.20	0.87	0.996
MEP2	51,394	710	941.75(43.76)	3.28	0.87	0.996
MEP3	23,429	503	757.43(49.53)	2.58	0.78	0.991
N1	56,045	865	1040.11(35.08)*	5.17	0.99	0.997
N2	102,266	954	1113.52(29.84)*	4.96	0.98	0.996
N3	71,260	842	1018.78(33.80)*	4.63	0.98	0.996
R1	71,564	1040	1152.51(24.52)*	5.16	0.99	0.997
R2	66,346	1006	1165.12(33.44)*	5.13	0.99	0.997

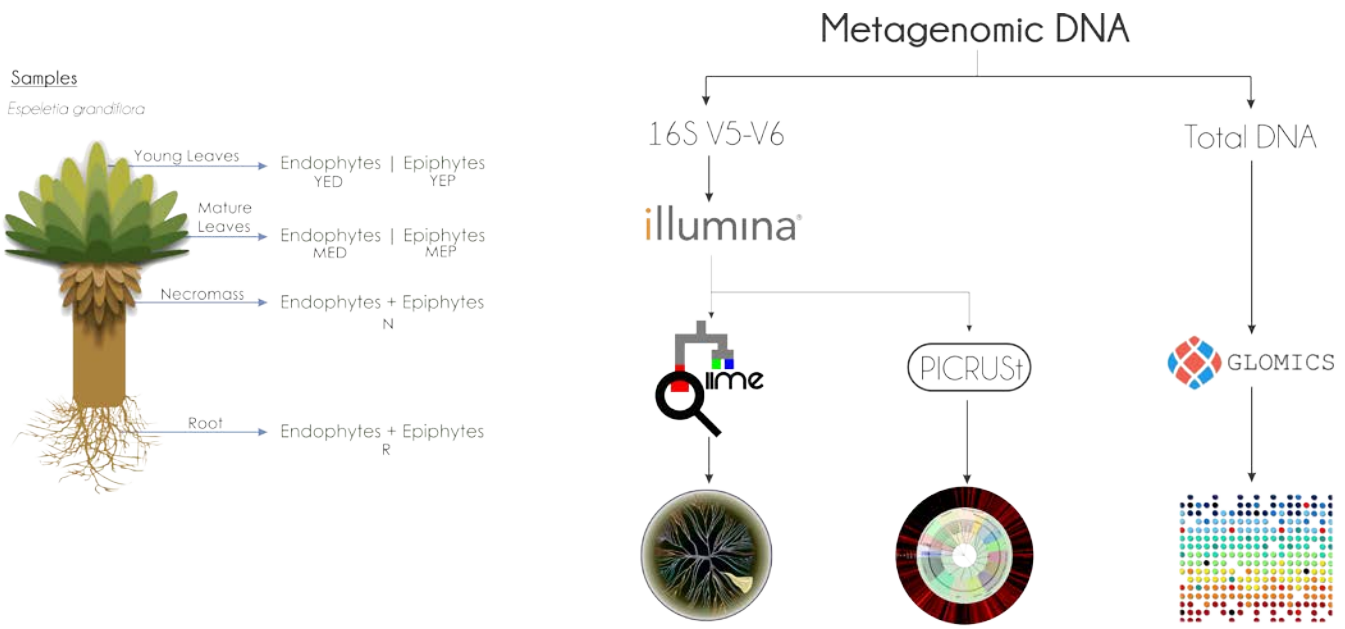

Figure S1. Overview of samples analyzed in the present study and the methods used for the analysis of bacterial communities associated to *Espeletia* sp. Below each community fraction (Endophyte/Epiphyte) the code of the samples used throughout the study (Y: Young leaves, M: Mature leaves, N: Necromass, R: Roots, EP: Epiphyte, ED: Endophyte).

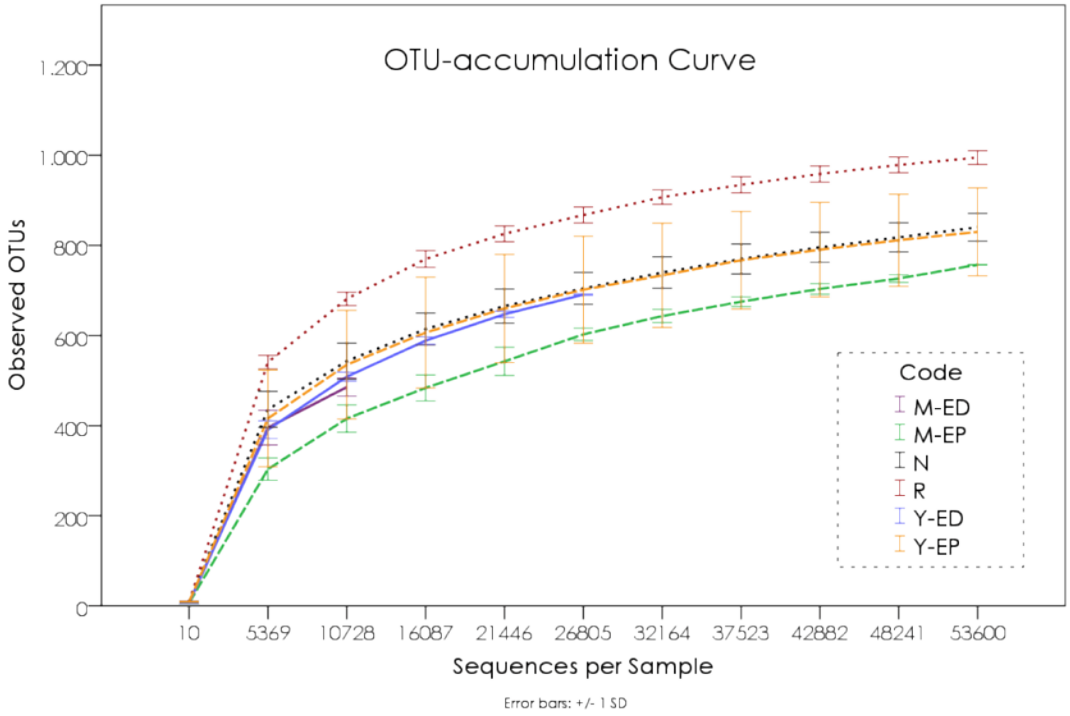

Figure S2. OTU-accumulation curves at 3% distance cutoff. Samples were collected from different tiers of the plant (Y, Young leaves; M, Mature leaves; N, Necromass; R, Roots) and for different microbial communities (EP, Epiphytes; ED, Endophytes). Continuous lines represent the endophytic fraction, dashed lines represent the epiphyte fraction and dotted lines represent both fractions in roots and necromass.

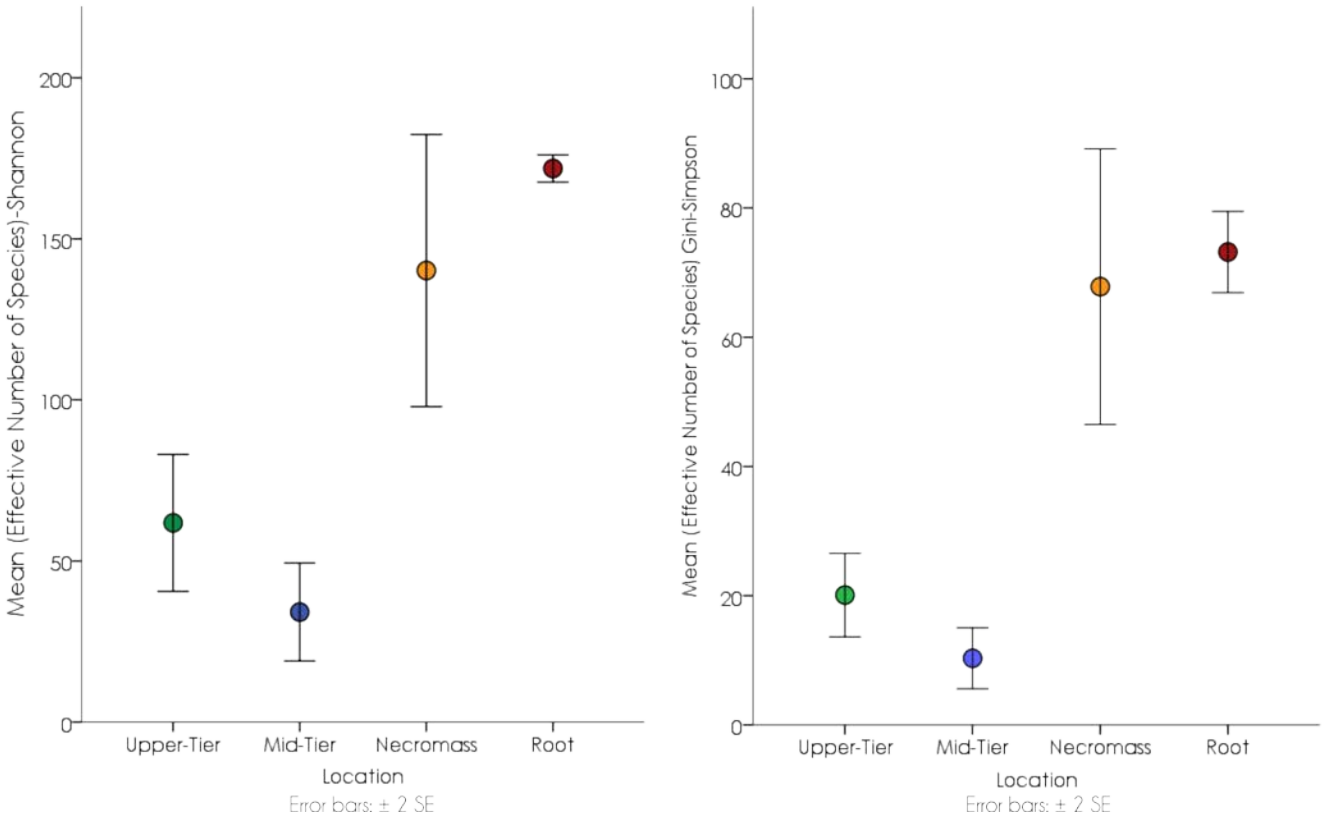

Figure S3. Transformed diversity indices Shannon (A) and Simpson (B) for the different plant tiers (Upper-Tier – YED + YEP; Mid-Tier – MED + MEP). Dots show the mean of the values for each tier. Letters represent statistically significant groups calculated using Tukey's *post-hoc* test.

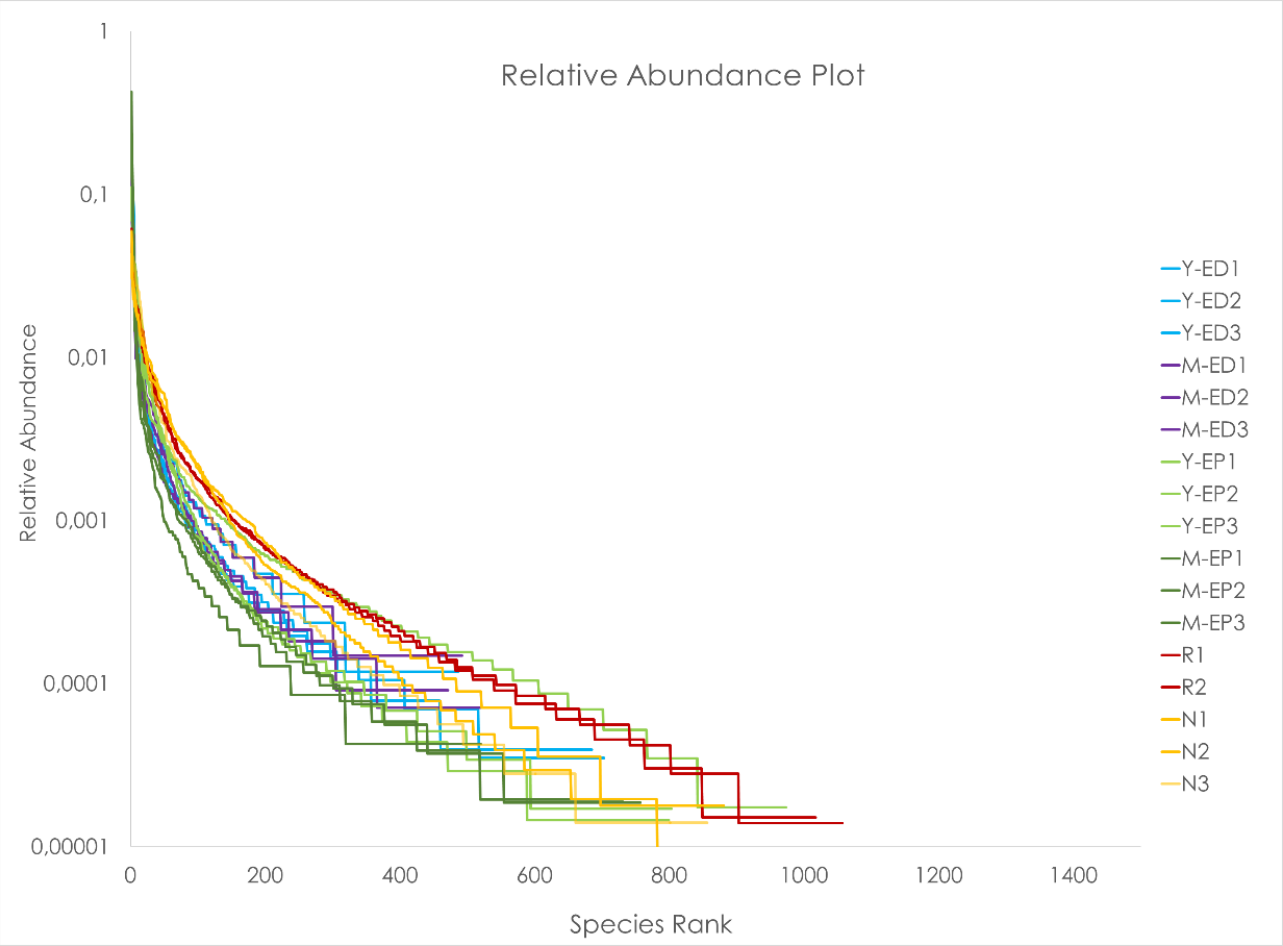

Figure S4. Rank abundance plot of OTUs observed at 97% similarity in datasets of *Espeletia* sp. amplicons, based on sequence abundance for each sample.

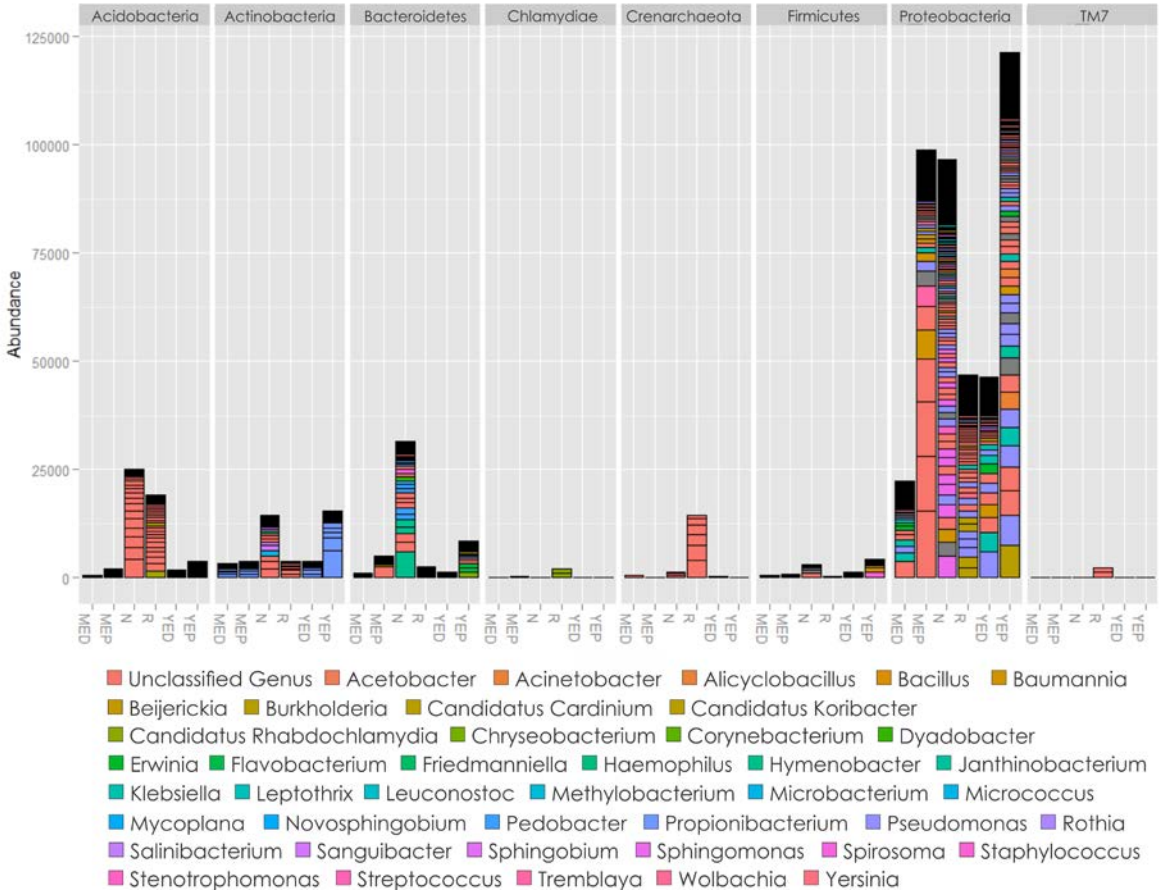

Figure S5. Raw abundance of species belonging to the core community of *Espeletia* sp.. Data is divided into phyla and different colors represent different species. (YED – Young Endophyte; MED – Mature Endophyte; YEP - Young Epiphyte; MEP – Mature Epiphyte; N – Necromass; R – Root).

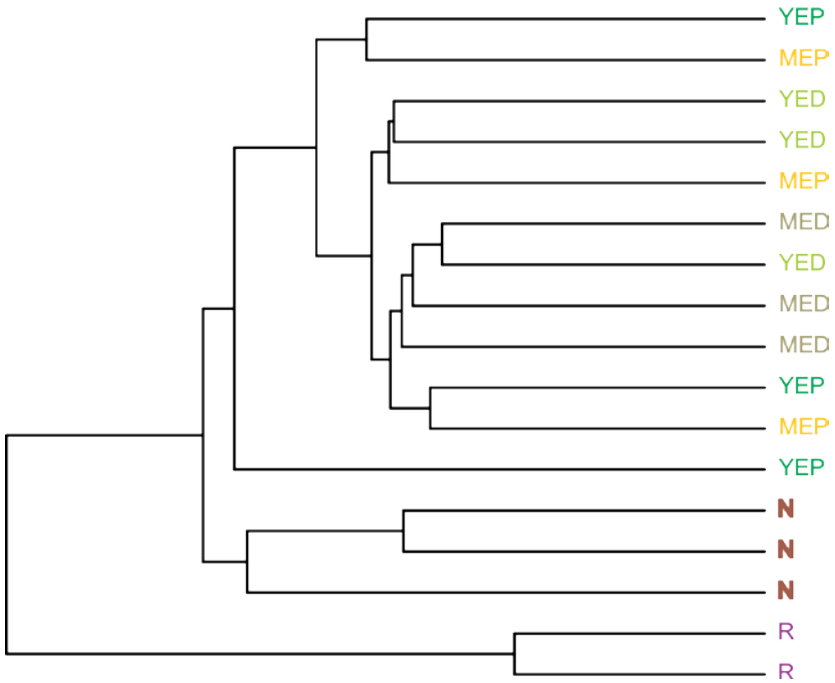

Figure S6. Hierarchical Clustering analysis using Weighted UniFrac distances.

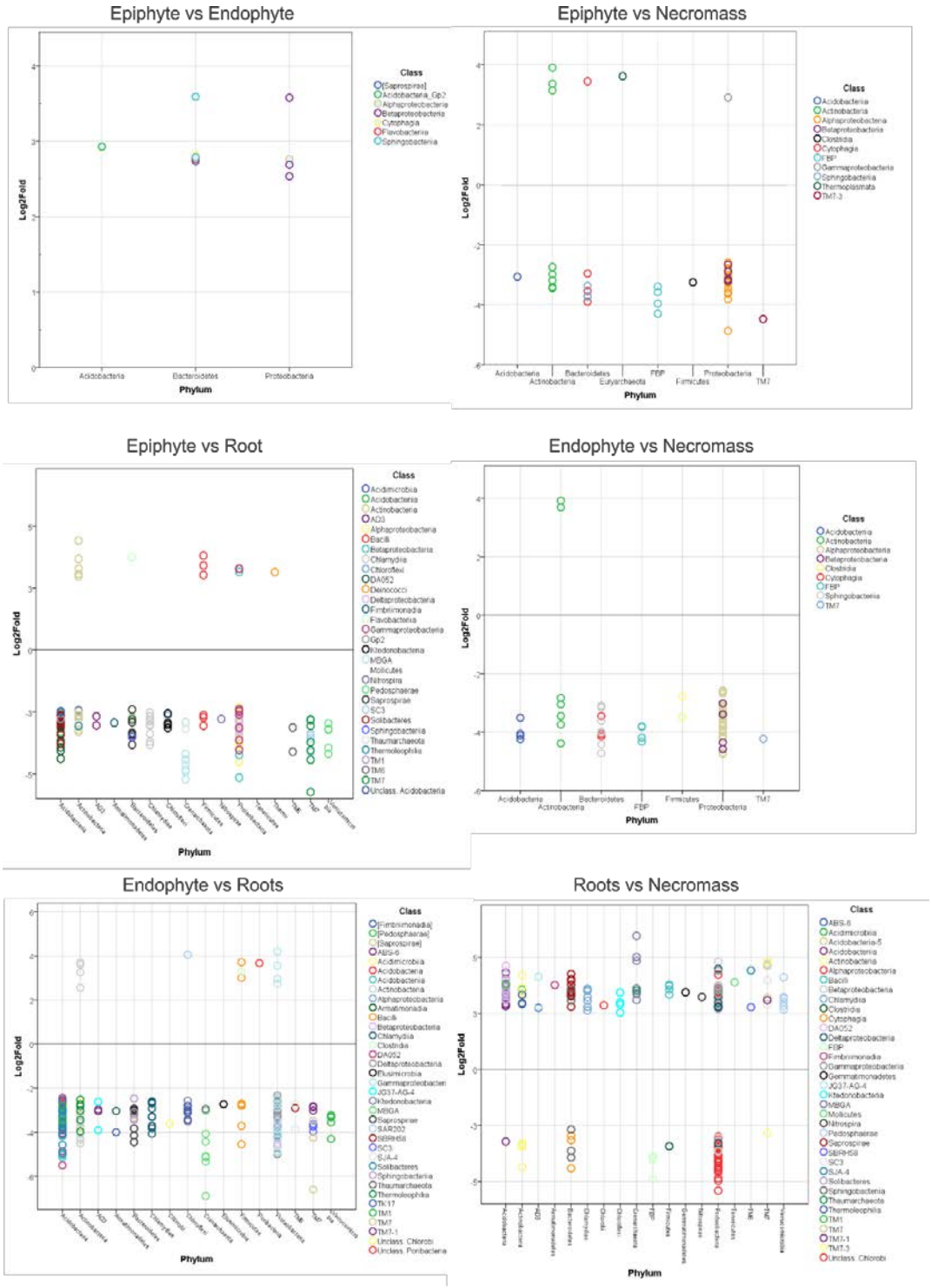

Figure S7. Pairwise community comparisons based on differentially abundant OTUs. *X* axis represents the phylum of each OTU and the *Y* axis the Log2 fold change of the first community vs the second. OTUs are colored according to their respective class. All differences are statistically significant ($p < 0.01$).

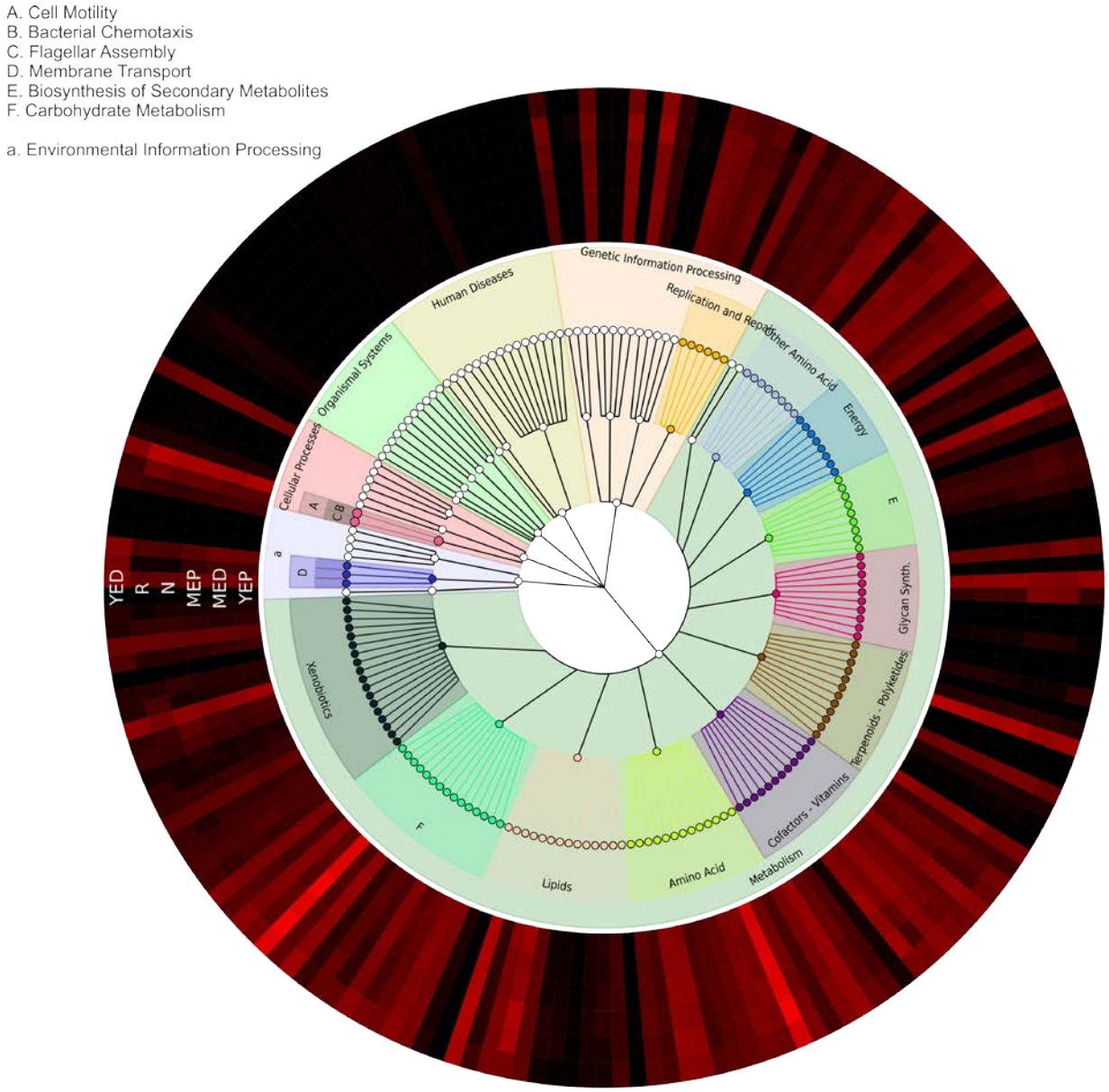

Figure S8. Potential metabolic pathways present in the microbiome associated with *Espeletia* sp. predicted using PICRUSt. The inner circles represent the metabolic categories as classified in the KEGG database. The heat map represents the predicted "abundances" for determined metabolic pathways inside the categories. (YED – Young Endophyte; MED – Mature Endophyte; YEP - Young Epiphyte; MEP – Mature Epiphyte; N – Necromass; R – Root).

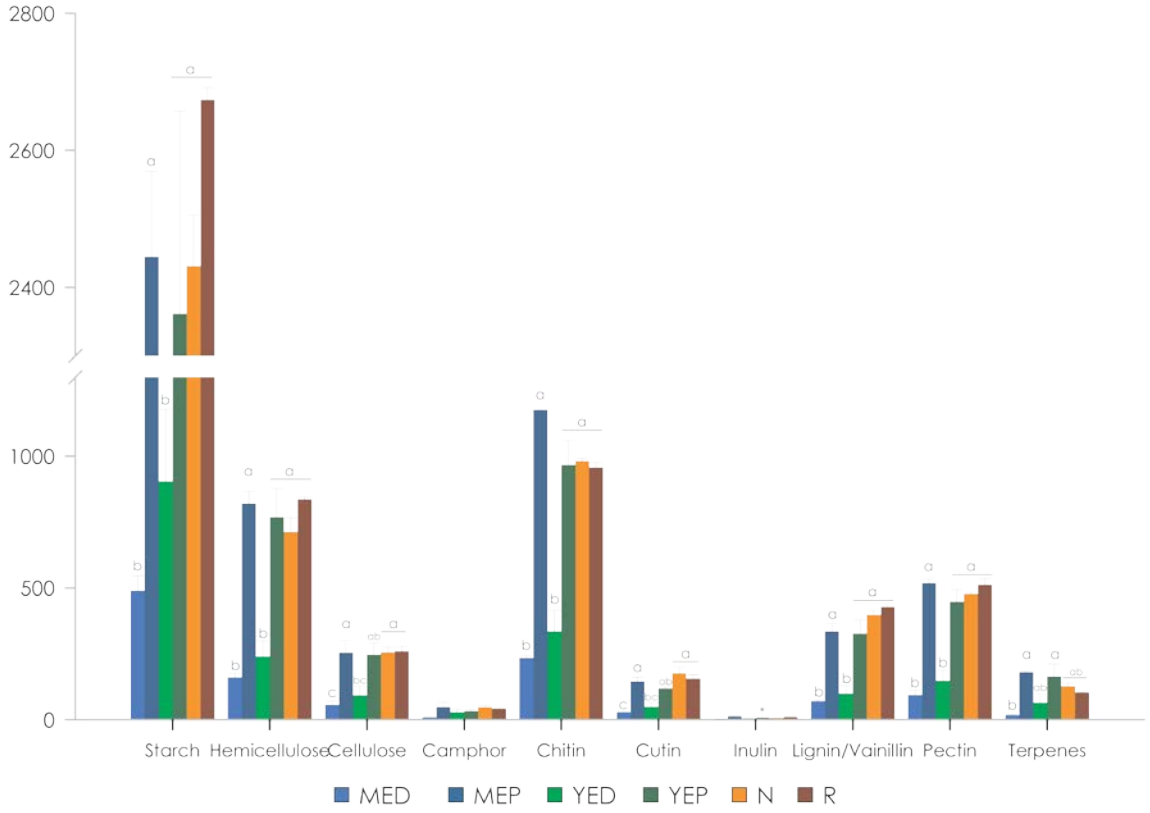

Figure S9. Relative abundance of several carbon degrading genes clustered by major carbon sources: Starch, hemicellulose, cellulose, camphor, chitin, cutin, inulin, lignin, pectin, terpenes. The signal intensity was normalized by the mean intensity of the microarray. Mean values of samples were plotted with their respective SE. Variations among communities were tested with one-way ANOVA and letters indicate statistically significant differences ($p < 0.05$) clustered according to the post-hoc Tukey's test.

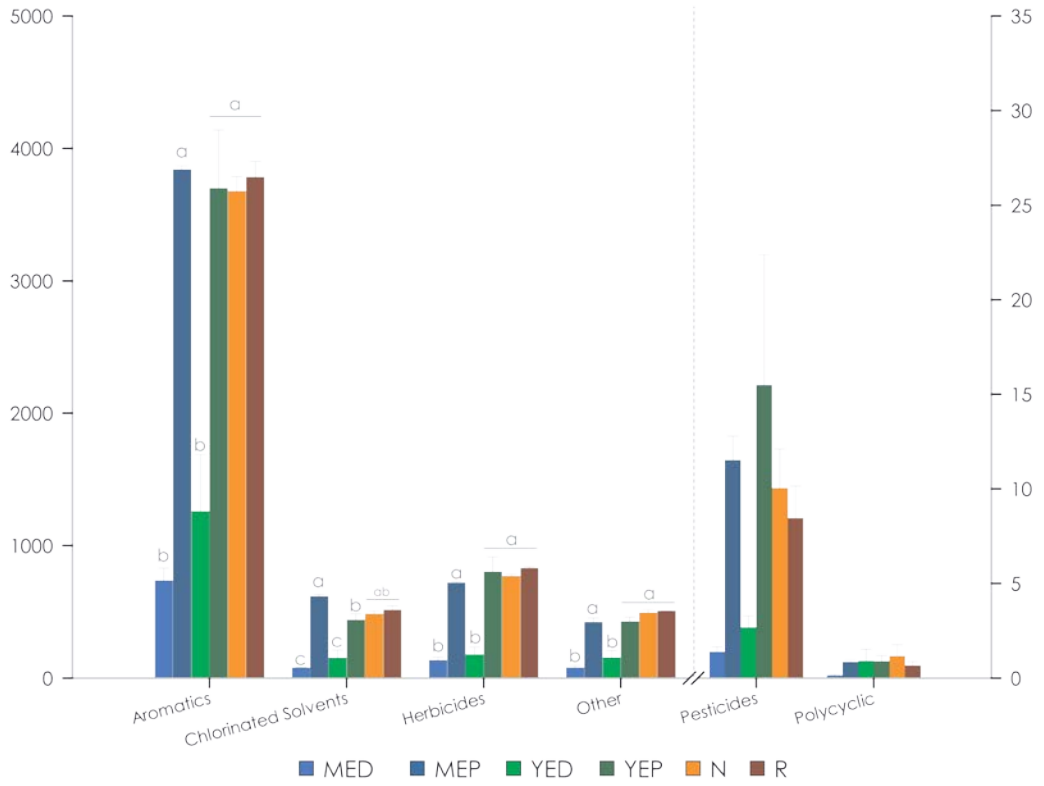

Figure S10. Relative abundance of several genes involved in remediation of organic compounds clustered by major compounds: Aromatics, chlorinated solvents, herbicides, other organic compounds, pesticides, polycyclic aromatics. The signal intensity was normalized by the mean intensity of the microarray. Mean values of samples were plotted with their respective SE. Variations among communities were tested with one-way ANOVA and letters indicate statistically significant differences ($p < 0.05$) clustered according to the post-hoc Tukey's test.

References

Abubucker S, Segata N, Goll J, Schubert AM, Izard J, Cantarel BL *et al* (2012). Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS computational biology* **8:** e1002358.

Afzal M, Khan QM, Sessitsch A (2014). Endophytic bacteria: Prospects and applications for the phytoremediation of organic pollutants. *Chemosphere* **117C:** 232-242.

Alfaro C (2002). *Geoquímica del Sistema Geotérmico de Paipa.* . INGEOMINAS – OIEA: Bogotá. Colombia. .

Araujo WL, Marcon J, Maccheroni W, Jr., Van Elsas JD, Van Vuurde JW, Azevedo JL (2002). Diversity of endophytic bacterial populations and their interaction with Xylella fastidiosa in citrus plants. *Applied and environmental microbiology* **68:** 4906-4914.

Aronesty E (2013). Comparison of Sequencing Utility Programs. *The Open Bioinformatics Journal* **7:** 1-8.

Bai S, Li J, He Z, Van Nostrand JD, Tian Y, Lin G *et al* (2013). GeoChip-based analysis of the functional gene diversity and metabolic potential of soil microbial communities of mangroves. *Applied microbiology and biotechnology* **97:** 7035-7048.

Balint-Kurti P, Simmons SJ, Blum JE, Ballare CL, Stapleton AE (2010). Maize leaf epiphytic bacteria diversity patterns are genetically correlated with resistance to fungal pathogen infection. *Molecular plant-microbe interactions : MPMI* **23:** 473-484.

Beattie GA, Lindow SE (1999). Bacterial colonization of leaves: a spectrum of strategies. *Phytopathology* **89:** 353-359.

Berendsen RL, Pieterse CM, Bakker PA (2012). The rhizosphere microbiome and plant health. *Trends in plant science* **17:** 478-486.

Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M *et al* (2010). Galaxy: a web-based genome analysis tool for experimentalists. *Current protocols in molecular biology / edited by Frederick M Ausubel [et al]* **Chapter 19:** Unit 19 10 11-21.

Bodenhausen N, Horton MW, Bergelson J (2013). Bacterial communities associated with the leaves and the roots of Arabidopsis thaliana. *PloS one* **8:** e56329.

Bohorquez LC, Delgado-Serrano L, Lopez G, Osorio-Forero C, Klepac-Ceraj V, Kolter R *et al* (2012). In-depth characterization via complementing culture-independent approaches of the microbial community in an acidic hot spring of the Colombian Andes. *Microbial ecology* **63:** 103-115.

Buczolits S, Denner EB, Vybiral D, Wieser M, Kampfer P, Busse HJ (2002). Classification of three airborne bacteria and proposal of Hymenobacter aerophilus sp. nov. *International journal of systematic and evolutionary microbiology* **52:** 445-456.

Buytaert W, Célleri R, De Bièvre B, Cisneros F, Wyseure G, Deckers J *et al* (2006). Human impact on the hydrology of the Andean páramos. *Earth-Science Reviews* **79:** 53-72.

Cakmakci ML, Evans HJ, Seidler RJ (1981). Characteristics of nitrogen-fixingKlebsiella oxytoca isolated from wheat roots. *Plant Soil* **61:** 53-63.

Campisano A, Ometto L, Compant S, Pancher M, Antonielli L, Yousaf S *et al* (2014). Interkingdom transfer of the acne-causing agent, Propionibacterium acnes, from human to grapevine. *Molecular biology and evolution* **31:** 1059-1065.

Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R (2010a). PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26:** 266-267.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al* (2010b). QIIME allows analysis of high-throughput community sequencing data. *Nature methods* **7:** 335-336.

Conn VM, Franco CM (2004). Analysis of the endophytic actinobacterial population in the roots of wheat (Triticum aestivum L.) by terminal restriction fragment length polymorphism and sequencing of 16S rRNA clones. *Applied and environmental microbiology* **70:** 1787-1794.

Coombs JT, Franco CM (2003). Isolation and identification of actinobacteria from surface-sterilized wheat roots. *Applied and environmental microbiology* **69:** 5603-5608.

Crawford DL, Lynch JM, Whipps JM, Ousley MA (1993). Isolation and characterization of actinomycete antagonists of a fungal root pathogen. *Applied and environmental microbiology* **59:** 3899-3905.

Chan Y, Van Nostrand JD, Zhou J, Pointing SB, Farrell RL (2013). Functional ecology of an Antarctic Dry Valley. *Proceedings of the National Academy of Sciences of the United States of America* **110:** 8990-8995.

Chelius MK, Triplett EW (2001). The Diversity of Archaea and Bacteria in Association with the Roots of Zea mays L. *Microbial ecology* **41:** 252-263.

Christner BC (2010). Bioprospecting for microbial products that affect ice crystal formation and growth. *Applied microbiology and biotechnology* **85:** 481-489.

Danhorn T, Fuqua C (2007). Biofilm formation by plant-associated bacteria. *Annual review of microbiology* **61:** 401-422.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K *et al* (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and environmental microbiology* **72:** 5069-5072.

Diaz E, Jimenez JI, Nogales J (2013). Aerobic degradation of aromatic compounds. *Current opinion in biotechnology* **24:** 431-442.

Edgar RC (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature methods* **10:** 996-998.

Enya J, Shinohara H, Yoshida S, Tsukiboshi T, Negishi H, Suyama K *et al* (2007). Culturable leaf-associated bacteria on tomato plants and their potential as biological control agents. *Microbial ecology* **53:** 524-536.

Ettwig KF, Butler MK, Le Paslier D, Pelletier E, Mangenot S, Kuypers MM *et al* (2010). Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. *Nature* **464:** 543-548.

Fagua JC, Gonzalez VH (2007). Growth rates, reproductive phenology, and pollination ecology of Espeletia grandiflora (Asteraceae), a giant Andean caulescent rosette. *Plant Biol (Stuttg)* **9:** 127-135.

Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P *et al* (2005). Galaxy: a platform for interactive large-scale genome analysis. *Genome research* **15:** 1451-1455.

Goecks J, Nekrutenko A, Taylor J, Galaxy T (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome biology* **11:** R86.

Gtari M, Ghodhbane-Gtari F, Nouioui I, Beauchemin N, Tisa LS (2012). Phylogenetic perspectives of nitrogen-fixing actinobacteria. *Archives of microbiology* **194:** 3-11.

Haroon MF, Hu S, Shi Y, Imelfort M, Keller J, Hugenholtz P *et al* (2013). Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage. *Nature* **500:** 567-570.

He Z, Deng Y, Van Nostrand JD, Tu Q, Xu M, Hemme CL *et al* (2010). GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity. *The ISME journal* **4:** 1167-1179.

Hirano SS, Upper CD (2000). Bacteria in the leaf ecosystem with emphasis on Pseudomonas syringae-a pathogen, ice nucleus, and epiphyte. *Microbiology and molecular biology reviews : MMBR* **64:** 624-653.

Hughes C, Eastwood R (2006). Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the Andes. *Proceedings of the National Academy of Sciences of the United States of America* **103:** 10334-10339.

Idris R, Trifonova R, Puschenreiter M, Wenzel WW, Sessitsch A (2004). Bacterial communities associated with flowering plants of the Ni hyperaccumulator Thlaspi goesingense. *Applied and environmental microbiology* **70:** 2667-2677.

Ikeda S, Okubo T, Anda M, Nakashita H, Yasuda M, Sato S *et al* (2010). Community- and genome-based views of plant-associated bacteria: plant-bacterial interactions in soybean and rice. *Plant & cell physiology* **51:** 1398-1410.

Izhaki I, Fridman S, Gerchman Y, Halpern M (2013). Variability of bacterial community composition on leaves between and within plant species. *Current microbiology* **66:** 227-235.

Jost L (2007). Partitioning diversity into independent alpha and beta components. *Ecology* **88:** 2427-2439.

Kanehisa M, Goto S (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* **28:** 27-30.

Kembel SW, O'Connor TK, Arnold HK, Hubbell SP, Wright SJ, Green JL (2014). Relationships between phyllosphere bacterial communities and plant functional traits in a neotropical forest. *Proceedings of the National Academy of Sciences of the United States of America* **111:** 13715-13720.

Keppler F, Hamilton JT, Brass M, Rockmann T (2006). Methane emissions from terrestrial plants under aerobic conditions. *Nature* **439:** 187-191.

Kim KH, Im WT, Lee ST (2008). Hymenobacter soli sp. nov., isolated from grass soil. *International journal of systematic and evolutionary microbiology* **58:** 941-945.

Lambers H, Mougel C, Jaillard B, Hinsinger P (2009). Plant-microbe-soil interactions in the rhizosphere: an evolutionary perspective. *Plant Soil* **321:** 83-115.

Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA *et al* (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature biotechnology* **31:** 814-821.

Leveau JH, Lindow SE (2001). Appetite of an epiphyte: quantitative monitoring of bacterial sugar consumption in the phyllosphere. *Proceedings of the National Academy of Sciences of the United States of America* **98:** 3446-3453.

Li SJ, Hua ZS, Huang LN, Li J, Shi SH, Chen LX *et al* (2014). Microbial communities evolve faster in extreme environments. *Scientific reports* **4:** 6205.

Lindow SE, Brandl MT (2003). Microbiology of the phyllosphere. *Applied and environmental microbiology* **69:** 1875-1883.

Love MI, Huber W, Anders S (2014). *Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2*.

Madrinan S, Cortes AJ, Richardson JE (2013). Paramo is the world's fastest evolving and coolest biodiversity hotspot. *Frontiers in genetics* **4:** 192.

Manter DK, Delgado JA, Holm DG, Stong RA (2010). Pyrosequencing reveals a highly diverse and cultivar-specific bacterial endophyte community in potato roots. *Microbial ecology* **60:** 157-166.

McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A *et al* (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME journal* **6:** 610-618.

McGarvey JA, Connell JH, Stanker LH, Hnasko R (2014). Bacterial population structure and dynamics during the development of almond drupes. *Journal of applied microbiology*.

McMurdie PJ, Holmes S (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS one* **8:** e61217.

McMurdie PJ, Holmes S (2014). Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS computational biology* **10:** e1003531.

Miles LA, Lopera CA, González S, García MCC, Franco AE, Restrepo S (2012). Exploring the biocontrol potential of fungal endophytes from an Andean Colombian Paramo ecosystem. *BioControl* **57:** 697-710.

Monasterio M, Sarmiento L (1991). Adaptive radiation of Espeletia in the cold andean tropics. *Trends in Ecology & Evolution* **6:** 387-391.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GA, Kent J (2000). Biodiversity hotspots for conservation priorities. *Nature* **403:** 853-858.

Price MN, Dehal PS, Arkin AP (2010). FastTree 2--approximately maximum-likelihood trees for large alignments. *PloS one* **5:** e9490.

Rasche F, Lueders T, Schloter M, Schaefer S, Buegger F, Gattinger A *et al* (2009). DNA-based stable isotope probing enables the identification of active bacterial endophytes in potatoes. *The New phytologist* **181:** 802-807.

Rastogi G, Sbodio A, Tech JJ, Suslow TV, Coaker GL, Leveau JH (2012). Leaf microbiota in an agroecosystem: spatiotemporal variation in bacterial community composition on field-grown lettuce. *The ISME journal* **6:** 1812-1822.

Rawat SR, Mannisto MK, Bromberg Y, Haggblom MM (2012). Comparative genomic and physiological analysis provides insights into the role of Acidobacteria in organic carbon utilization in Arctic tundra soils. *FEMS microbiology ecology* **82:** 341-355.

Redford AJ, Bowers RM, Knight R, Linhart Y, Fierer N (2010). The ecology of the phyllosphere: geographic and phylogenetic variability in the distribution of bacteria on tree leaves. *Environmental microbiology* **12:** 2885-2893.

Reinhold-Hurek B, Hurek T (2011). Living inside plants: bacterial endophytes. *Current opinion in plant biology* **14:** 435-443.

Schloss PD, Handelsman J (2005). Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Applied and environmental microbiology* **71:** 1501-1506.

Simon HM, Jahn CE, Bergerud LT, Sliwinski MK, Weimer PJ, Willis DK *et al* (2005). Cultivation of mesophilic soil crenarchaeotes in enrichment cultures from plant roots. *Applied and environmental microbiology* **71:** 4751-4760.

Torche A, Benhizia H, Rosselli R, Romoli O, Zanardo M, Baldan E *et al* (2014). Characterization of bacteria associated with nodules of two endemic legumes of Algeria, Hedysarum naudinianum and H. perrauderianum. *Ann Microbiol* **64:** 1065-1071.

Tuomisto H (2010). A consistent terminology for quantifying species diversity? Yes, it does exist. *Oecologia* **164:** 853-860.

Vaz Jauri P, Bakker MG, Salomon CE, Kinkel LL (2013). Subinhibitory antibiotic concentrations mediate nutrient use and competition among soil streptomyces. *PloS one* **8:** e81064.

Vokou D, Vareli K, Zarali E, Karamanoli K, Constantinidou HI, Monokrousos N *et al* (2012). Exploring biodiversity in the bacterial community of the Mediterranean phyllosphere and its relationship with airborne bacteria. *Microbial ecology* **64:** 714-724.

Vorholt JA (2012). Microbial life in the phyllosphere. *Nature reviews Microbiology* **10:** 828-840.

Wang F, Zhou H, Meng J, Peng X, Jiang L, Sun P *et al* (2009). GeoChip-based analysis of metabolic diversity of microbial communities at the Juan de Fuca Ridge hydrothermal vent. *Proceedings of the National Academy of Sciences of the United States of America* **106:** 4840-4845.

Whipps JM, Hand P, Pink D, Bending GD (2008). Phyllosphere microbiology with special reference to diversity and plant genotype. *Journal of applied microbiology* **105:** 1744-1755.

Wyngaarden V, Fandiño M (2002). *Parque Nacional los Nevados. Un caso de selección y zonificación de áreas de conservación biológica.*: Bogotá – Colombia: Pontificia Universidad Javeriana.

Xie J, He Z, Liu X, Liu X, Van Nostrand JD, Deng Y *et al* (2011). GeoChip-based analysis of the functional gene diversity and metabolic potential of microbial communities in acid mine drainage. *Applied and environmental microbiology* **77:** 991-999.

Xiong ZJ, Zhang JL, Zhang DF, Zhou ZL, Liu MJ, Zhu WY *et al* (2013). Rothia endophytica sp. nov., an actinobacterium isolated from Dysophylla stellata (Lour.) Benth. *International journal of systematic and evolutionary microbiology* **63:** 3964-3969.

Zhang B, Bai Z, Hoefel D, Wang X, Zhang L, Li Z (2010). *Microbial diversity within the phyllosphere of different vegetable species*, vol. 2. FORMATEX: Spain.

Zhang L, Dai J, Tang Y, Luo X, Wang Y, An H *et al* (2009). Hymenobacter deserti sp. nov., isolated from the desert of Xinjiang, China. *International journal of systematic and evolutionary microbiology* **59:** 77-82.