Enumeration of Citrus endophytic bacterial communities based on illumine metagenomics technique

[1] Sehrish Mushtaq*, [1a] Muhammad Shafiq, [2] Tehseen Ashraf, [1]Muhammad Saleem Haider, Sagheer Atta*,[3]

[1] Faculty of Agricultural Sciences, Department of Plant Pathology, University of the Punjab, Quaid-e-Azam Campus, Lahore.

[1a] Faculty of Agricultural Sciences, Department of Horticulture Sciences, University of the Punjab, Quaid-e-Azam Campus, Lahore.

[2] Department of Horticulture Sciences University of Sargodha, Sargodha, Pakistan.

[3]Department of Plant Protection, Faculty of Agricultural Sciences, Ghazi University Dera Ghazhi Khan, Pakistan

*Corresponding Author e-mail: sherry.a143@gmail.com; satta@gudgk.edu.pk

These authors contributed equally to this work.

[&] These authors also contributed equally to this work.

Running title: Estimation of bacterial diversity from citrus using Illumina Hiseq

Abstract

Citrus is a valuable crop in Pakistan because it is rich in vitamin C and antioxidants. Huanglongbing (HLB) has an influence on citrus production around the world caused by a bacterium "*Candidatus liberibacter asiaticus*" (CLas), africanus and americanus. The structure and diversity of bacterial species in various ecosystems can be quickly examined using NGS. This approach is considerably quicker and more precise than outdated methods. Healthy or citrus greening infected leaf samples of Grapefruit, *Citrus aurantifolia*, and *Citrus reticulata* Blanco was used for diversity analysis. In this study high throughput, NGS technique was used to access the population of both cultivable and non-cultivable bacterial endophytes from citrus leaves, by using PCR amplicons of 16S rDNA sequences (V5–V7 regions) with Illumina Hi seq. As a result, a total number of 68,722 sequences were produced from the test samples. According to the NGS-based diversity classification, the most common genera of exploited bacterial endophytes were Proteobacteria, Firmicutes, Bacteroides, Cyanobacteria, and Actinobacteria. *Citrus aurantifolia* and *Citrus paradisi* showed almost equal diversity, whereas *Citrus reticulata* Blanco had a higher proportion of Proteobacteria and Cyanobacteria in their leaves. To determine alpha diversity (AD), additional data was analyzed using statistical indices such as Shannon, Chao1, and Simpson. According to the inverse Simpson diversity index, the abundance of the microbial population in six different citrus samples was 0.48, 0.567, and 0.163, respectively. The metagenomics of microbiota in plant tissues was successfully recorded by NGS technology, which can help us learn more about the interactions between plants and microbes. This research is the first step toward a better understanding of 16SrRNA-based metagenomics from citrus in Pakistan using Illumina (Hi seq) Technology.

Keywords: Citrus, α-Diversity, bacterial endophytes, Next-generation sequencing, Illumina (Hiseq), 16SrRNA

Introduction

Pakistan is one of the world's largest citrus producers, ranking 13th in total citrus production. Citrus is highly important due to its economic and nutritional benefits. Kinnow is a useful fruit that occupies the first place among all fruits in terms of both area and production [1]. The total area under citrus cultivation during 2014-15 was 192832 hectares with a production of 2395550 (tons) [2]. Punjab is home to nearly all of the world's citrus groves. With more than 75% production of total citrus fruits, 29.55% of the total area is planted in citrus and 60% in kinnow. About 90% of all citrus exports are kinnow. Major Citrus species cultivated in Pakistan are as follows; Grapefruit, Sweet orange, Mandarin, Lemon Lime, Bitter orange [3].

Citrus diseases have emerged as a possible threat to global citrus productivity. HLB, a disease caused by three gram-negative, phloem-limited alphaproteobacteria: "*Candidatus liberibacter asiaticus*" (CLas), africanus, and americanus have a major effect on citrus production worldwide [4]. However, different CLas strains have been recorded from the United States, specifically from Florida [5-8], Iran [9], Mexico [10], Australia [11], and Pakistan [12]. HLB is distinguished by less nutrient transfer, resulting in a variety of distinct effects, including yellow shoots, branch dieback, green fruit remaining, lopsided fruit, reduced size and eventually tree death [13]. The plant microbiome plays a part in different aspects of plant health and disease, including growth rate, vigor, and tolerance, inflammation, and disease resistance [14, 15]. Understanding how the microbiome affects and communicates with the plant would entail the application of several experimental methods, including a meta-analysis of broad Meta datasets with critical variables relevant to plant health, protection, and disease [16].

NGS is a culture-independent method that is useful for the study of the entire microbial population within a sample. High-throughput sequencing technologies [17] refer to a group of tools that can be used to sequence DNA of various base pairs faster and cheaper than previous methods. NGS sequence of DNA fragment (16S rRNA) in the form of reading (short DNA fragment) as compared to reference sequences from databases in lesser time to identify the related bacterium with this fragment [18, 19]. There are various studies of 16SrRNA gene base sequencing for targeted amplification of bacterial communities [20]. Although, in this era of science researchers are using the most effective variable (V) region of the 16SrRNA gene for sequencing, with many studies selecting to examine more than one region as no single region has been shown to optimally differentiate among bacteria [21, 22]. All nine Variable regions of 16S rRNA displayed bacterial diversity and the most important step is determining which variable region to sequence, since classification bias variable region has been found previously [23]. The use of PCR-based molecular techniques (polymerase chain reaction) has made it possible to research the total diversity of microbes in the natural environment without the cultivation of microbes [24]. These new advanced techniques are valuable in increasing our understanding of the microbial communities regardless of some amplification biases demonstrated due to the selection of suitable primers, the concentration of template, and the number of amplification cycles [25, 26].

NGS-based microbial community research has paved the way for the development of novel culture-independent bacterial strains capable of identifying biological control agents against the HLB pathogen (*Candidatus liberibacter asiaticus*). The study of biological control organisms' natural microbial niches, which are close to those of pathogens, could lead to more successful disease control. Microbial diversity associated with citrus leaf (phloem) can be identified by either cultivation-dependent or cultivation-independent methods. On the other hand, the fraction of bacterial

diversity measured using previous culture techniques accounts for just 0.1 to 10% of the overall estimated diversity [27, 28] , suggesting that laboratory culture techniques are substantially biased. However, it is a fact that the majority of dominant bacteria present in environmental samples are uncultivable [29-32]. 16S rDNA-based phylogenetic analysis has been widely used to classify microbial diversity in different environmental niches, such as soil [30], plants [33, 34], subsurface sediments, and rocks [35]. The primary aim of this research was to determine whether bacteria other than *Ca. Liberibacter spp*. is associated with the citrus greening disease.

Microbial diversity research is important for recognizing the microbial flora that exists on plants in their natural environment. The diversity of bacterial endophytes from citrus in Pakistan is the focus of this report, which is based on preliminary research. The uncultivable and cultivable fraction of bacteria is first time exploited from citrus leaves through the Illumina metagenomics technique (Hi seq) in Pakistan. There has been an increased recognition that it is necessary to pay more attention to this area. NGS (next-generation sequencing) is an incredibly valuable technique to access the uncultivable fraction of bacterial endophytes in plant tissues. This could help us better understand the microbes that live on plant surfaces in natural conditions and how they interact. To the best of our understanding, this is Pakistan's initial 16SrRNA-based metagenomics study from citrus leaves using Illumina (Hi seq). The main objectives of this research were to investigate the microbial species associated with the leaf midribs of HLB symptomatic and asymptomatic citrus (*Citrus aurantifolia*, *Citrus paradisi*, *Citrus reticulata* Blanco) trees and also to know their relative abundance, and phylogenetic diversity by using high-throughput 16S rDNA (V5-V7) next-generation sequencing through Illumina (Hi-seq).

Materials and Methods

Samples collection and DNA isolation:

Leaf samples (healthy/infected) of grapefruit, *Citrus aurantifolia***,** and *Citrus reticulata* Blanco were obtained from IAGS, Pu, Lahore backfields and preserved at -80^{o}C. Citrus plants that were six years old were used for this experiment and five leaves per plant were taken as a sample and stored at -80^{o}C. To extract soil particles, every plant leaf was washed and cleaned under running tap water. The leaves were washed in autoclaved water with a few drops of Tween-20 and set aside to drain for 10–15 minutes. Then they were cut into 4–5 bits, each measuring 2–3cm in length. Surface sterilization was carried out using the methods defined by [36], with some variations in the Ethanol conc. and sterilization time. Soft tissue were submerged in ninety percent ethanol soln. for 5 minutes, then in a 3 percent sodium hypochlorite solution for 2 minutes, and finally in 75% ethanol (3 min). The disinfected leaves were drained in a laminar flow hood after being rinsed three times with autoclaved distilled water. The surface-sterilized tissues (control) and the last rinsing water were inoculated onto nutrient agar plates to confirm the efficacy of the surface sterilization procedure. Any bacteria growth in the control agar plates within 24 hours of incubation (30°C±2°C) indicates ineffective surface sterilization. The complete genome of DNA was extracted using the CTAB method (cetyl trimethyl ammonium bromide), as defined by [37]. At A260/280 nm (1.9–2.0), the isolated DNA was quantified and tested for purity (Nanodrop at School of Biological Sciences Pu, Lahore) and stored at -20°C before being processed. For NGS (Illumina Hi seq), these quantified DNA samples were sent to the Novo gene (leading-edge genomics services and solutions).

Generation of Amplicon:

The bacterial genomic DNA concentration in leaf tissue samples was normalized to 10 ng/L. The conserved regions of 16S rRNA were amplified using PCR (V5-V7-WBI-NV2018010942). Phusion® High-Fidelity PCR Master Mix was used to prepare the PCR library (New England Biolabs). Briefly, 25 µL PCR reaction comprises DNA (6 µL), 12.5 µL of (2x) Master KAPA Hi-Fidelity DNA polymerase (1 U), primer (10 µM) 1.5 µL (each), and distilled autoclaved water. PCR reactions were

initiated with 95°C for 3 min (denaturation cycle) followed by 24 cycles at 98°C for 20 sec, 55°C for 15 sec, and 72°C for 10 sec, and ended at 72°C for 1 min (Extension step). Mix the same amount of 1X loading buffer (with SYB green) with PCR products and run electrophoresis on 2% agarose gel for detection. Samples with a bright main strip between 400-450bp were selected for further studies. PCR products were Gel purified by using Qiagen Kit Manufactured by (Qiagen, Germany).

Library preparation and sequencing:

Sequence libraries were generated using the TruSeq® DNA PCR-Free Sample Preparation Kit (Illumina, USA) following the instructions given. The quality of the library was analyzed by the Qubit@ 2.0 Fluorometer (Thermo Scientific) and the Agilent Bioanalyzer 2100 system. To conclude, the library was sequenced on the Illumina HiSeq 2500 platform and 250 bp paired-end reads were produced. A preliminary study of the illustration and base call was performed on the HiSeq instrument. Hi Seq (ultra-high-throughput) was used to de-multiplex data and exclude reads in FASTQ format that failed the Illumina purity filter (PF = 0). The forward and reverse reads of raw data were combined using the mother pipeline alignment method. Following that, they were trimmed and filtered by deleting the bases with rating scores less than or equal to 2, the maximum number of N accepted = 4, the maximum number of homopolymers accepted = 8, and the contaminant removed. All tests were performed using the Mothur pipeline program software (http://www.mothur.org/wiki/).

Classification of bacteria

The SILVA rRNA database and the Silva database were used to assign operational taxonomic units (OTUs) were assigned to the retrieved read sequences produced from the leaf samples. We used the mother pipeline's “splitting by classification “process to assign OTU.

Statistical analysis

All of the data was processed using one-way ANOVA. The Statistical Package for Social Science (SPSS) was used to conduct the analysis, Tukey's Studentized Range Test HSD (0.05) has been used to compare the means, and p values less than 0.05 which were considered statistically significant.

Diversity Analysis

Alpha and beta are two methods of diversity analysis that are commonly used to find diversity using NGS. Alpha diversity (AD) is used to analyze the complexity of species diversity in the experiment by diversity indices, including Observed-Species, Chao1, Shannon, Simpson, ACE, and Good-coverage. All of these indices were measured with QIIME and viewed with the R program. Beta Diversity (BD) Analysis was used to assess differences in sample species complexity. Beta Diversity was measured using QIIME software Unit fraction metrics (unifrac), as weighted and unweighted. Unifrac is a method of calculating the phylogenetic distance between taxonomic groups in a tree as a percentage of the length of the branch that contributes to ancestors from either one or both origins. Arithmetic Means in an Unweighted Pair-group Method (UPGMA) QIIME was used to perform clustering, a hierarchical clustering technique that uses average linkage to interpret the distance matrix.

RESULTS

DNA extraction for Next-Generation Sequencing (Hi seq)

Leaf samples of both healthy and symptomatic *Citrus paradisi*, *C. reticulata* Blanco, and *C. aurantifolia* were obtained from the backyard of the Institute of Agricultural Sciences and preserved at-80^{o}C. The CTAB method was used to separate DNA from the leaf samples. To access the diversity of cultivable vs. non-cultivable bacteria, isolated DNA was electrophoresed in a 1% agarose gel (to verify DNA) and, quantified through nanodrop before further processing. sent for Illumina Hi seq NGS technique.

Sequencing and data processing

The Illumina paired-end network was used to sequence the PCR amplicon yielding raw reads (Raw PE) with paired ends of 250 bp that were then extracted and Clean Tags were obtained after being pretreated. To obtain Effective Tags, Clean Tags that included chimeric sequences were identified and excluded. The data output indicates data interpretation and QC status (Table 1).

Table 1: **Data processing and QC (quality control) stats of citrus (***Citrus aurantifolia* **,** ***Citrus paradisi*** **,** *Citrus reticulata* Blanco**) samples.**

Sample abbreviations used in this study	Raw PE (#)	Raw Tags (#)	Clean Tags (#)	Effective Tags (#)	Base (nt)	AvgLen (nt)	Q20	Q30	GC %	Effective %
C.aurantifolia Healthy **(SM1)**	42,914	41,860	39,641	26,813	9,982,254	372	98.60	97.08	53.71	62.48
C. aurantifolia Infected **(SM2)**	35,655	34,844	33,190	22,880	8,21,549	372	98.66	97.22	53.77	64.17
C. paradisi Healthy **(MK1)**	68,722	67,168	63,631	55,999	20,820,815	372	98.58	97.05	53.36	81.49
C. paradisi Infected **(MK2)**	57,447	56,216	53,447	47,782	17,789,958	372	98.58	97.07	53.57	83.18
C. reticulata Blanco Healthy **(MA1)**	33,837	33,037	31,226	24,496	9,149,614	374	98.53	96.98	55.43	72.39
C.reticulata Blanco Infected **(MA2)**	30,903	30,223	28,359	25,189	9,440,842	375	98.45	96.82	55.99	81.51

OTU Clustering and species annotation

All Successful Tags were grouped into OTUs based on 97% DNA sequence similarity to evaluate the species diversity in each sample. Detailed information gathered from a variety of samples, such as Tag annotation data, effective Tags data, and low-frequency Tags data was collected during the construction of OTUs. The statistical data set is organized as follows (Fig. 1).

Figure 1: Statistical analysis of the tags and operational taxonomic units of each tested citrus leaf sample

Phylogenetic Tree

R and D software was used to select independently the most common top ten genera of specific species with high relative abundance by default) for the construction of a phylogenetic tree [51]. Actinobacteria, Cyanobacteria, and Firmicutes, as well as Proteobacteria, were identified that belong to the phylum (Alpha, beta, and gamma). The research samples were found to be infected with

eight orders and nine groups of bacteria (figure tree of particular species in samples SM-1/SM-2 (Asymptomatic/Symptomatic *Citrus aurantifolia*); Mk-1/MK-2 (Asymptomatic/ Symptomatic *Citrus paradisi*); MA1/MA2 (Asymptomatic/Symptomatic *Citrus reticulata* Blanco). In this diagram, the four major phyla are represented (Fig. 2).

Figure 2: Taxonomy tree of specific species in citrus leaf samples.

Relative Abundance of Species

To structure the scattering of relative abundance of species in histograms, the top 10 species in each taxonomic rank were chosen. The distribution of the phyla can be seen in (Fig. 3) and the relative abundance of bacterial species in normal vs. infected leaves revealed that SM1/SM2 (*Citrus aurantifolia* asymptomatic/symptomatic) has a higher proportion of Proteobacteria, whereas the infected one has a smaller proportion of other phyla, with Cyanobacteria dominating among them. MKI/MK2 (Citrus paradisi asymptomatic /symptomatic) showed a similar pattern. MA1/MA2 (*Citrus reticulata* Blanco asymptomatic/ symptomatic) had 40% Proteobacteria and 60% Cyanobacteria, while MA2 had 20% Proteobacteria and the remaining 80% Cyanobacteria and another phylum. The relative abundance of bacteria is calculated by integrating both symptomatic and asymptomatic bacteria into one group were represented through; Bac-1 (Citrus aurantifolia) community revealed 90% Proteobacteria and just around 10% cyanobacteria. While in Bac-2 (*Citrus paradisi*) group only Proteobacteria was found in abundance. Bac-3 (Citrus reticulata Blanco) community, on the other hand, had a 25% proportion of Proteobacteria and a 75% proportion of Cyanobacteria and others.

Figure 3: Relative abundance of bacterial Species at phylum level from citrus leaves

The Phylogenetic tree

The top hundred taxa have been selected, and the evolutionary tree was built by aligning the sequences. Each genus' relative abundance was measured as shown in (Fig. 4).

Figure 4: The evolutionary tree based on the genus of Bacterial endophytes from citrus leave

Venn diagrams were also constructed based on operational taxonomic units of the identified bacteria from citrus leaf samples as shown in (Fig.5).

Figure 5: Venn diagram constructed based on operational taxonomic units of the bacterial diversity from citrus leaf samples

Alpha and Beta diversity Analysis

OTUs with 97% sequence identity are assumed to be homologous among species and statistical indices of AD are listed in (Table 2).

Table 2: Statistical analysis of alpha diversity (AD) indices from NGS data of citrus leaves

Sample Abbreviations used in this study	**No of species observed**	**Simpson**	**Shannon**	**Chao1**	**ACE**	**Good coverage**	**PD whole tree**
C.aurantifolia Healthy **(SM1)**	128	0.484	2.120	141.571	138.761	0.999	9.691
C. aurantifolia Infected **(SM2)**	131	0.567	2.290	138.241	144.296	0.999	9.428
C. paradisi Healthy **(MK1)**	92	0.163	0.751	113.136	130.639	0.998	8.407
C. paradisi Infected **(MK2)**	87	0.307	1.245	102.812	105.703	0.999	8.125
C. reticulata Blanco Healthy **(MA1)**	104	0.741	2.419	120.714	130.361	0.999	8.699
C.reticulata Blanco Infected **(MA2)**	69	0.539	1.602	84.833	88.361	0.999	6.379

Beta Diversity Indices and heat map

Unweighted vs. Weighted Unifrac distances, which are phylogenetic indicators that are commonly used in current bacterial community sequencing projects, were chosen to quantify the dissimilarity coefficient between pairwise samples. In this graph, a heat map centered on the weighted vs. unweighted Unifrac distances is plotted (Fig.6). The red section of the triangle suggests that there is less beta variety among samples, whereas the yellow portion indicates that there is more beta diversity among samples (SM2, MK1, MK2, and MA1).

Figure 6: Illustrates beta diversity analysis (Heat map) based on Weighted/Unweighted Unifrac distances.

Unweighted Pair-group Method with Arithmetic Mean (UPGMA)

Clustering analysis and the construction of a clustering tree were used to investigate the similarities between different samples. The (UPGMA) procedure with arithmetic mean is a type of hierarchical clustering method used for classifying ecosystem samples. The following are fundamental concepts of UPGMA methods. The samples with the shortest distance were being grouped, and then a new sample is generated. It has a branching point in the middle of the two initial samples. After computing the average distance between the newly created "sample" and other samples, the closest two samples can be used to repeat the procedures adopted earlier in this section. Until all of the samples are clustered together, a complete clustering tree can be obtained. Before conducting UPGMA cluster analysis, the Weighted Unifrac distance matrix and the Unweighted Unifrac distance matrix were calculated. They could be seen in a graph that included the clustering results as well as every sample's phylum-specific relative abundance (Fig. 7a and 7b).

The SM1/SM2 and MK1/MK2 clusters in the same clade had more or less similar bacterial diversity, according to the UPGMA cluster tree based on the Weighted Unifrac distance tree, however, MA1/MA2 displayed a distinct configuration and is in a different clade, indicating that *Citrus reticulata* Blanco has a different bacterial diversity than *Citrus aurantifolia* and *Citrus paradisi*. The UPGMA cluster tree based on unweighted unifrac distance displays a variable pattern if compared to the weighed unifrac distance tree. MK1/MK2 is in a distinct clade in unweighted unifrac distance trees, whereas the other two groups are all in the same clade, as seen in (Fig.7b). MK1/MK2 had a higher proportion of Bacteroides than the others. As a whole, the most common genera found in three samples were Proteobacteria, Cyanobacteria, and Actinobacteria.

Figure 7: UPGMA cluster tree based on a) Weighted Unifrac distance b) Un Weighted Unifrac distance showing the relative abundance of bacterial species at phyla level.

DISCUSSION

A microbial community study is a fast way to learn regarding the structure and functioning of bacterial communities, and it could contribute to the isolation and detection of new bacteria [38]. This research explores the diversity and composition of microbial communities in the leaf midribs of both HLB-affected and healthy citrus plants. Our research discovered that the Illumina sequencing protocol can be used to evaluate the bacterial endophytes present in plant tissues. The sequencing can be improved with a good choice of primer pair to amplify a longer stretch of the 16S rRNA gene. Our empiric findings illustrate the importance of this platform for accurate and high-resolution microbiota profiling (N90% at species level) of endophytic populations or may be extended to other resources/samples. It was critical to design multiple testing procedures to minimize the bias introduced by host DNA (chloroplast) and chimaera, which were both removed without changing the overall read quality. We have a good likelihood of executing the read sequence efficiently on a specific platform with the Mothur pipeline. This reduced the possibility of read contamination.

The total diversity and complexity of microbiome populations in plant tissues, which include both cultivable and non-cultivable endophytic bacteria, were also exposed using the novel NGS shotgun 16S rRNA gene. Alpha diversity (AD), comprised of species abundance boxplots, species richness curves, and statistical analysis indices, is a common technique for evaluating bacterial diversity within populations [39]. The spreading of bacterial species across tissues and the overall mutual richness is illustrated in this Venn diagram. The Venn diagram (map) of the OTU distribution exposed a colonization pattern of Acinetobacteria 1.41%, Cyanobacteria 28.56%, Firmicutes 1.31%, and Proteobacteria 63.76% of microbes contained in plant leaves were also identified in three samples.

The phylum cyanobacteria were found to be more common in *C. reticulata* Blanco as compared to other phyla. On the contrary, the other two samples presented a greater fraction of Proteobacteria, and few phyla were not observed by culture-based methods, illustrating the importance of NGS. This also

led to the fact that these microbes can spread through a variety of channels that penetrate plant tissues [40, 41]. Finally, Proteobacteria, Firmicutes, Actinobacteria, Cyanobacteria, and Bacteriodetes were found to colonize citrus plant leaf tissues, they have been demonstrated to produce useful bioactive chemicals A comparison of bacterial species based on their structure is referred to as beta diversity. As a result, the differences in microbial populations are measured using beta-diversity metrics. A square "distance" or dissimilarity matrix, such as Unweighted Unifrac, was calculated to reflect the contrast among test plant leaves to compare microbial communities between each pair of group samples [42, 43] and Weighted Unifrac distances [44].

At the phylum level, Actinobacteria accounted for 26.47 percent, Cyanobacteria for 2.94 percent, Firmicutes for (23.52%), and Proteobacteria for (47.05%), which was significantly higher than the fraction of other phyla. Though we found 100 genera among them most common were Staphylococcus, Pseudomonas, Lactobacillus, Sphingomonas, Bacillus, Streptomyces, and Pantoea. Bacillus and Lactobacillus, as well as Streptomyces, have previously been found in the roots or leaves of infected (CLas) or infected citrus trees [46-49]. Pantoea, Curtobacterium, and Methylobacterium were also detected in citrus leaves in this analysis. All of these have previously been characterized in terms of bud wood, leaves, and roots [50].

Through studying the PCR products of 16S rDNA sequences covering two specific regions (V3–V4 regions), [45] discovered the diversity of bacterial endophytes from Aloe vera plant leaves, stems, and roots using the NGS by illumina Hi seq technology. The most popular genera identified were Proteobacteria, Firmicutes, Actinobacteria, and Bacteriodetes. This research was identical to the findings of the current study, but we looked for diversity in the V5-V7 region of 16S r RNA. Illumina for next-generation sequencing Hiseq is a relatively new method, with only a limited amount of literature available on it. The discovery of novel bacterial endophytes from citrus illustrates the

significance of this study. There has been no comparable work being done with this technique in citrus in other regions of the world, not yet in Pakistan.

Conclusion

The predominant bacterial groups in the leaf of citrus varieties were Proteobacteria, Actinobacteria, Cyanobacteria, Firmicutes, and Bacteroides, although other groups were commonly found to be less prevalent. Through the culture-dependent method, we find changes in bacterial diversity of endophytes from a citrus leaf but in comparison with an uncultured method, no significant variations existed in relative abundance and diversity of bacteria among taxa from both symptomatic and asymptomatic leaf samples. Some genera such as Staphylococcus, Enterococcus, Enterobacter, Pseudomonas, Bacillus, and Burkholderia were also found in the cultured approach (unpublished data). Although the type of strains has a significant influence on their functional characterization in terms of plant growth-promoting traits rather than their source of isolation either from bulk soil or rhizosphere soil. These genera have been widely found in most of the diversity-related studies of different parts of plants and soils. Some of the isolated strains have great potential to enhance plants growth and they can also be utilized as biocontrol agents against different plant diseases. Finally, this study indicates that these endophytic bacteria may be tested in open field conditions on the same host plants to see whether their biocontrol potential or plant growth-promoting action is successful. Furthermore, their effects on plant physiology could be estimated. We may use these endophytes to produce biofertilizers to replace chemical fertilizers if the same results are obtained from field trials.

Acknowledgments

The first author acknowledges the financial grant from Higher Education Commission (HEC) Pakistan.

References

1. Usman M, Ashraf I, Chaudhary KM, Talib. Factors impeding citrus supply chain in Central Punjab, Pakistan. Int J Agri Ext.2018; 6:1-5.

345 **2.** Government of Pakistan. Ministry of National Food Security & Research Islamabad;2016.

346 3. Naz S, Shahzadi K, Rashid S, Saleem F, Zafarullah A, Ahmad S.et al. Molecular
347 characterization and phylogenetic relationship of different citrus varieties of Pakistan. J Anim
348 Plant Sci.2014; 24: 315-320.

349 4. Wang N, Pierson EA, Setubal JC, Xu J, Levy JG, Zhang Y, Li J, Rangel LT, Martins JJ. et al.
350 The Candidatus Liberibacter–host interface: insights into pathogenesis mechanisms and disease
351 control. Annu Rev Phytopathol.201; 55: 451-482.

352 5. Chen J, Deng X, Sun X, Jones D, Irey M, Civerolo E. et al. Guangdong and Florida populations
353 of 'Candidatus Liberibacter asiaticus' distinguished by a genomic locus with short tandem
354 repeats. Phytopathology.2010;100: 567-572.

355 6. Zheng Z, Wu F, Kumagai L , Polek M, Deng X, Chen J.et al. Two 'Candidatus Liberibacter
356 asiaticus' strains recently found in California harbor different prophages. Phytopathology.2017;
357 107: 662-668.

358 7. Kunta M, Zheng Z, Wu F, da Graca JV, Park JW, Deng X, Chen J.et al. Draft whole-genome
359 sequence of "Candidatus Liberibacter asiaticus" strain TX2351 isolated from Asian citrus
360 psyllids in Texas, USA. Genome announc.2017; 5(15).

361 8. Hall D. Incidence of "Candidatus Liberibacter asiaticus" in a Florida population of Asian citrus
362 psyllid. J Appl Entomol.2018; 142: 97-103.

363 9. Passera A, Alizadeh H, Azadvar M, Quaglino F, Alizadeh A, Casati P, Bianco P. et al. Studies of
364 Microbiota Dynamics Reveals Association of "Candidatus Liberibacter Asiaticus" Infection with
365 Citrus (Citrus sinensis) Decline in South of Iran. Int J Mol Sci.2018; 19: 1817.

366 10. Lopez-Buenfil JA, Ramirez-Pool JA, Ruiz-Medrano R, Montes-Horcasitas M del Carmen,
367 Chavarin-Palacio C, Moya-Hinojosa J, Trujillo-Arriaga FJ, Carmona RL, Xoconostle-Cazares B.

et al. Dynamics of Huanglongbing-associated Bacterium Candidatus Liberibacter asiaticus in Citrus aurantifolia Swingle (Mexican Lime). Pak J Biol Sci.2017; 20:113-123.

11. Luo X , Yen AL, Powell KS, Wu F, Wang Y, Zeng L, Yang Y, Cen Y.et al. Feeding behavior of Diaphorina citri (Hemiptera: Liviidae) and its acquisition of' Candidatus Liberibacter asiaticus', on huanglongbing-infected Citrus reticulata leaves of several maturity stages. Fla Entomol.2015;186-192.
12. Yaqub MS, Khan IA, Usman M, Rana IA. Molecular detection of Candidatus Liberibacter asiaticus, the causal organism of huanglongbing (Citrus greening) in Faisalabad, Pakistan for huanglongbing management. Pak J Agri Sci.2017; 54.
13. da Graça JV, Douhan GW, Halbert SE, Keremane ML, Lee RF, Vidalakis G, Zhao H.et al. Huanglongbing: An overview of a complex pathosystem ravaging the world's citrus. J Integr Plant Biol.2016; 58: 373-387.
14. Schlaeppi K, Bulgarelli D. The plant microbiome at work. Mol Plant-Microbe Interact.2015; 28:212-217.
15. Vogel C, Bodenhausen N, Gruissem W, Vorholt JA. The Arabidopsis leaf transcriptome reveals distinct but also overlapping responses to colonization by phyllosphere commensals and pathogen infection with impact on plant health. New Phytol.2016; 212: 192-207.
16. Ginnan NA, Dang T, Bodaghi S, Ruegger PM, Peacock BB, McCollum G, England G, Vidalakis G, Roper C, Rolshausen P, Borneman J.et al. Bacterial and Fungal Next Generation Sequencing Datasets and Metadata from Citrus Infected with 'Candidatus Liberibacter asiaticus'. Phytobiomes.2018; 2(2): 64-70.
17. Rodríguez-Ezpeleta N, Hackenberg M, Aransay AM. "Bioinformatics for high throughput sequencing," Springer Science & Business Media;2011.

18. Bybee SM, Bracken-Grissom H, Haynes BD, Hermansen RA, Byers RL, Clement MJ, Udall JA, Wilcox ER, Crandall KA.et al.Targeted amplicon sequencing (TAS): a scalable next-gen approach to multilocus, multitaxa phylogenetics. Genome Biol Evol.2011;3:1312-1323.
19. Karagöz MA,Nalbantoglu OU.Taxonomic classification of metagenomic sequences from Relative Abundance Index profiles using deep learning. Biomed Signal Process.2021; 67, 102539.
20. Chakravorty S, Helb D, Burday M, Connell N, Alland D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. J Microbiol Methods.2007;69:330-339.
21. Mizrahi-Man O, Davenport ER, Gilad Y. Taxonomic classification of bacterial 16S rRNA genes using short sequencing reads: evaluation of effective study designs. PloS one.2013; 8: e53608.
22. Alindonosi A, Baeshen M, Elsharawy N. Prospects For Diatoms Identification Using Metagenomics: A Review. Appl Ecol Environ Res.2021;19(6):4281-4298.
23. Vilo C, Dong Q. Evaluation of the RDP classifier accuracy using 16S rRNA gene variable regions. Metagenomics.2012;1: 1-5.
24. Hugerth LW, Andersson AF. Analysing microbial community composition through amplicon sequencing: from sampling to hypothesis testing. Front Microbiol.2017;8:1561.
25. Shokralla S, Spall JL, Gibson JF, Hajibabaei M. Next-generation sequencing technologies for environmental DNA research. Mol Ecol.2012; 21:1794-1805.
26. Takahashi S, Tomita J, Nishioka K, Hisada T, Nishijima M. Development of a prokaryotic universal primer for simultaneous analysis of bacteria and archaea using next-generation sequencing. PloS one.2014; 9:e105592.

27. Thao ML, Baumann P. Evolutionary relationships of primary prokaryotic endosymbionts of whiteflies and their hosts. Appl Environ Microbiol .2004;70:3401-3406.

28. Lennon JT, Muscarella ME, Placella SA, Lehmkuhl BK. How, when, and where relic DNA affects microbial diversity. MBio.2018; 9(3): e00637-18.

29. de Melo Pereira GV, Magalhaes KT, Lorenzetii ER, Souza TP, Schwan RF. A multiphasic approach for the identification of endophytic bacterial in strawberry fruit and their potential for plant growth promotion. Microb Ecol.2012; 63: 405-417.

30. Kim YC, Leveau J, Gardener BBM, Pierson EA, Pierson LS, Ryu Cm.et al. The multifactorial basis for plant health promotion by plant-associated bacteria. Appl Environ Microbiol.2011; **77**: 1548-1555.

31. Polz MF, Cavanaugh CM. Bias in template-to-product ratios in multitemplate PCR. Appl Environ Microbiol.1998;64: 3724-3730.

32. Schneider H. Anatomy of greening-diseased sweet orange shoots. Phytopathology.1968; 58: 1555-1160.

33. Chelius M, Triplett E. The Diversity of Archaea and Bacteria in Association with the Roots of Zea mays L. Microb Ecol.2001;41: 252-263.

34. Singh BK, Millard P, Whiteley AS, Murrell JC. Unravelling rhizosphere–microbial interactions: opportunities and limitations. Trends Microbiol.2004;12:386-393.

35. Chandler D, Fredrickson J, Brockman F. Effect of PCR template concentration on the composition and distribution of total community 16S rDNA clone libraries. Mol Ecol.1997; 6:475-482.

36. Azevedo JL, Maccheroni JW, Pereira JO, de Araújo WL. Endophytic microorganisms: a review on insect control and recent advances on tropical plants. Electron J Biotechnol.2000;3: 15-16.

37. Wilson K. Preparation of genomic DNA from bacteria. Curr. Protoc. Mol. Biol.1987;2-4.

437 38. Costa LE, Queiroz MVD, Borges AC, Moraes CAD, Araujo EFD . Isolation and characterization
438 of endophytic bacteria isolated from the leaves of the common bean (Phaseolus vulgaris). Braz J
439 Microbiol.2012;43:1562-1575.
440 39. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Gormley N.et al.
441 Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq
442 platforms. ISME J.2012;6(8):1621-1624.
443 40. Huang JS. Ultrastructure of bacterial penetration in plants. Annu Rev Phytopathol.1986; 24:141-
444 157.
445 41. Quadt-Hallmann A, Kloepper J, Benhamou N. Bacterial endophytes in cotton: mechanisms of
446 entering the plant. Can J Microbiol.1997; 43:577-582.
447 42. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial
448 communities. Appl Environ Microbiol.2005; 71: 8228-8235.
449 43. Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. UniFrac: an effective distance
450 metric for microbial community comparison. ISME J.2011; 5:169.
451 44. Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and qualitative β diversity
452 measures lead to different insights into factors that structure microbial communities. Appl
453 Environ Microbiol.2007; 73: 1576-1585.
454 45. Akinsanya MA, Goh JK, Lim SP,Ting ASY. Metagenomics study of endophytic bacteria in
455 Aloe vera using next-generation technology. Genomics data.2015;6:159-163.
456 46. Trivedi P, Duan Y, Wang N. Huanglongbing, a systemic disease, restructures the bacterial
457 community associated with citrus roots. Appl Environ Microbiol.2010;76(11):3427-3436.
458 47. Trivedi P, He Z, Van Nostrand JD, Albrigo G, Zhou J, Wang N.et al. Huanglongbing alters the
459 structure and functional diversity of microbial communities associated with citrus rhizosphere.
460 ISME.2012: J 6:363.

48. Trivedi P, Spann T, Wang N. Isolation and characterization of beneficial bacteria associated with citrus roots in Florida. Microb Ecol.2011; 62: 324-336.

49. Zhang M, Powell CA, Benyon LS, Zhou H, Duan Y. Deciphering the bacterial microbiome of citrus plants in response to 'Candidatus Liberibacter asiaticus'-infection and antibiotic treatments. PloS one.2013;8:e76331.

50. Blaustein RA, Lorca GL, Meyer JL, Gonzalez CF,Teplitski M. Defining the core citrus leaf-and root-associated microbiota: Factors associated with community structure and implications for managing huanglongbing (citrus greening) disease. Appl. Environ. Microbiol.2017; 83(11): e00210-17.

51. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie E, Keller K, Andersen GL.et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol.2006;72(7):5069-5072.

475

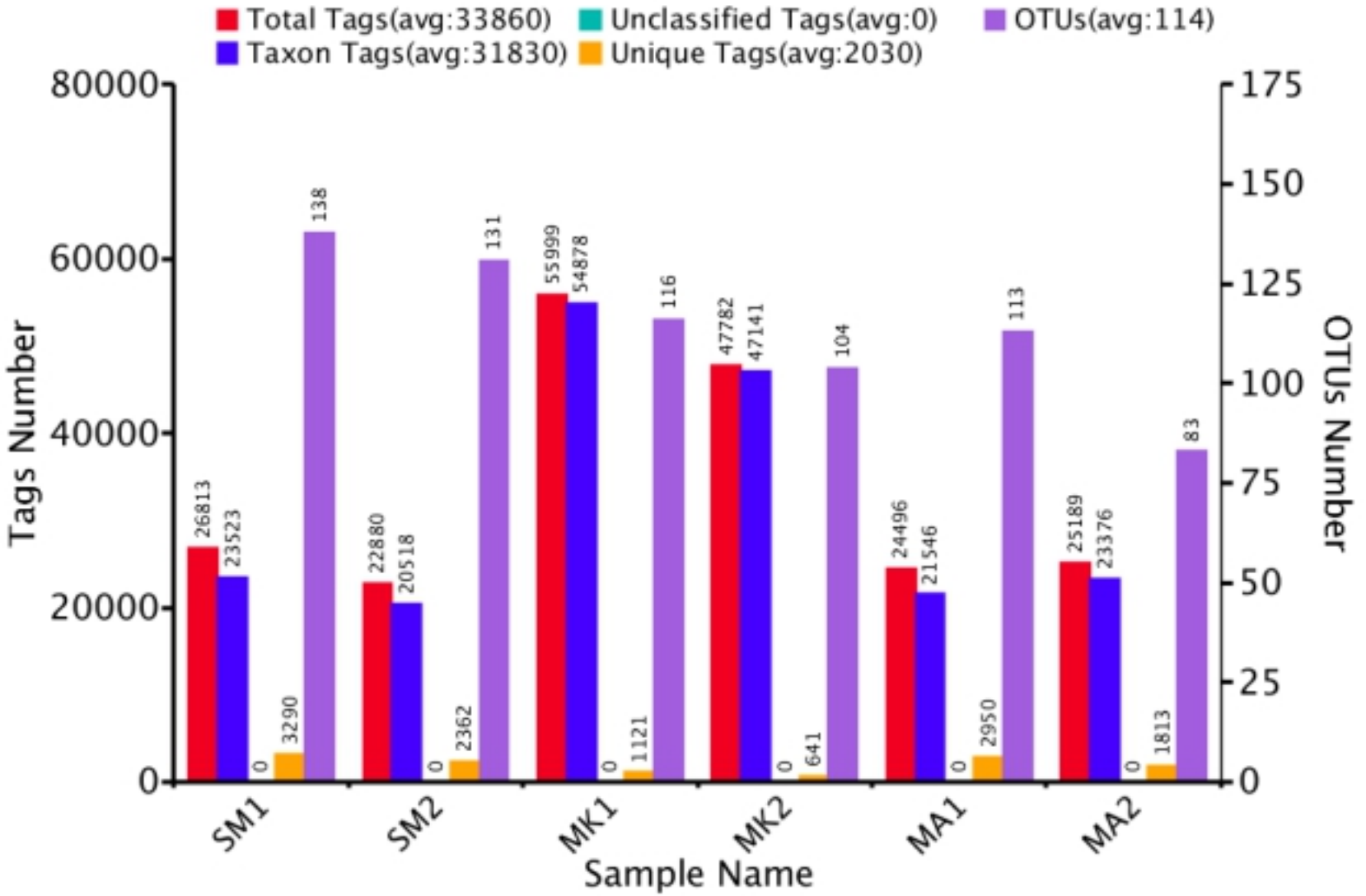

Total Tags(avg:33860)
Taxon Tags(avg:31830)
Unclassified Tags(avg:0)
Unique Tags(avg:2030)
OTUs(avg:114)
Tags Number
OTUs Number
Sample Name
80000
60000
40000
20000
0
175
150
125
100
75
50
25
0
SM1
SM2
MK1
MK2
MA1
MA2
26813
23523
0
3290
138
22880
20518
0
2362
131
55999
54878
0
1121
116
47782
47141
0
641
104
24496
21546
0
2950
113
25189
23376
0
1813
83

Fig. 1

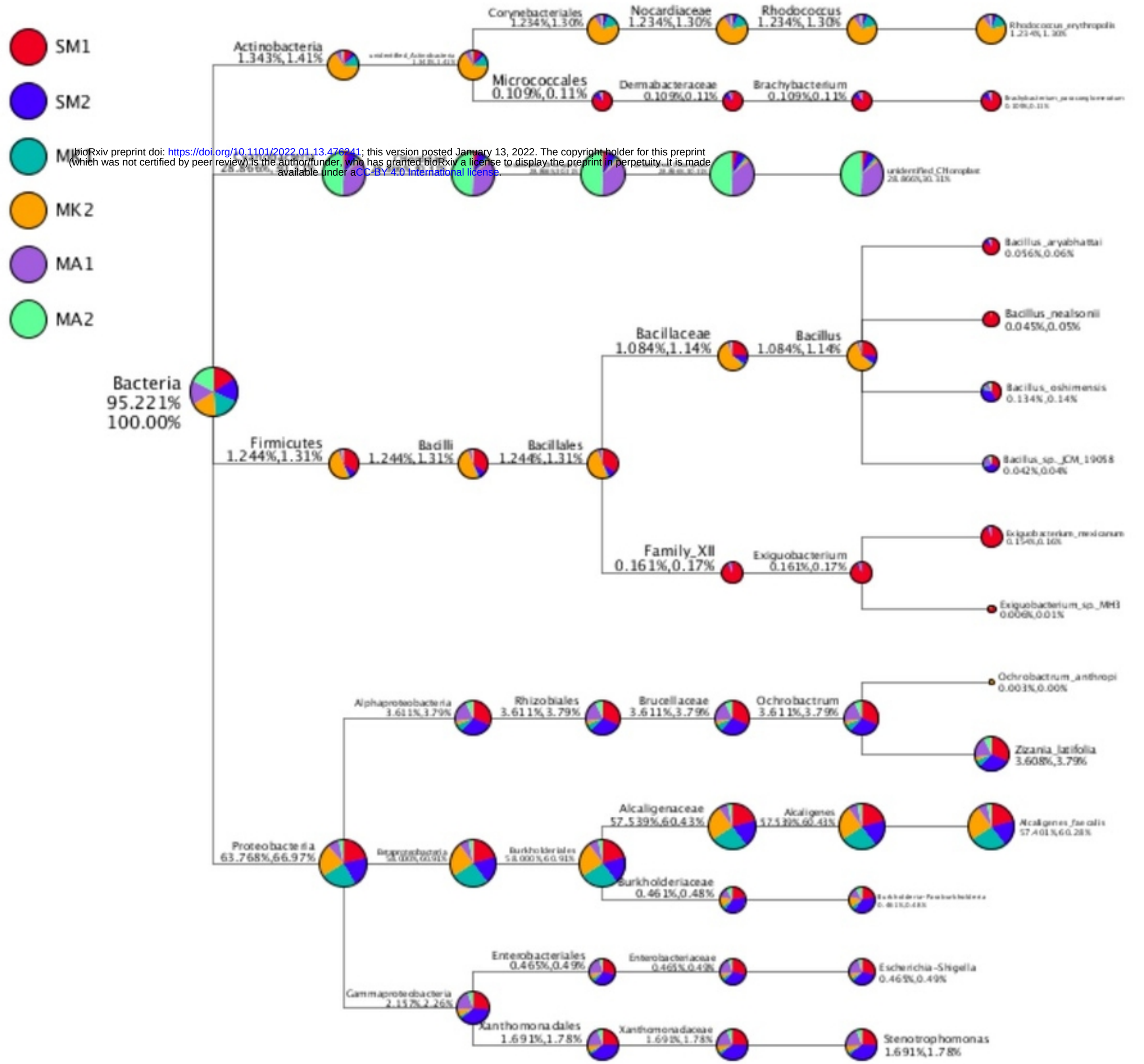
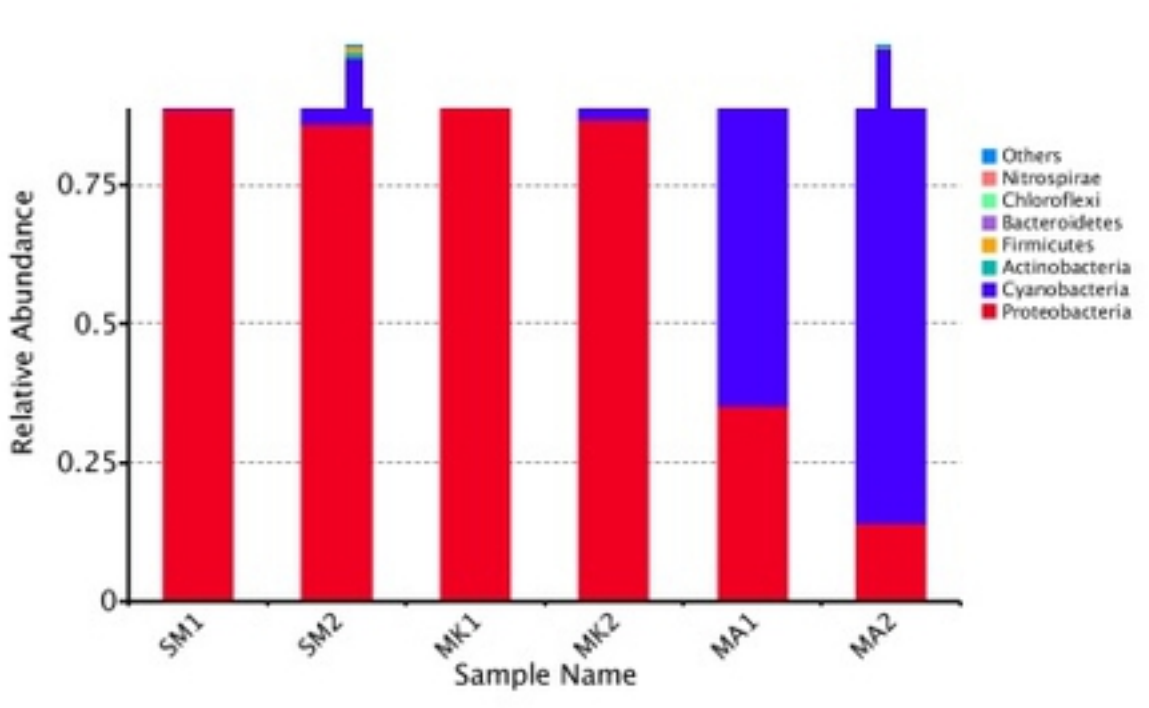


Fig. 2


Relative Abundance
0.75
0.5
0.25
0
SM1
SM2
MK1
MK2
MA1
MA2
Sample Name
Others
Nitrospirae
Chloroflexi
Bacteroidetes
Firmicutes
Actinobacteria
Cyanobacteria
Proteobacteria

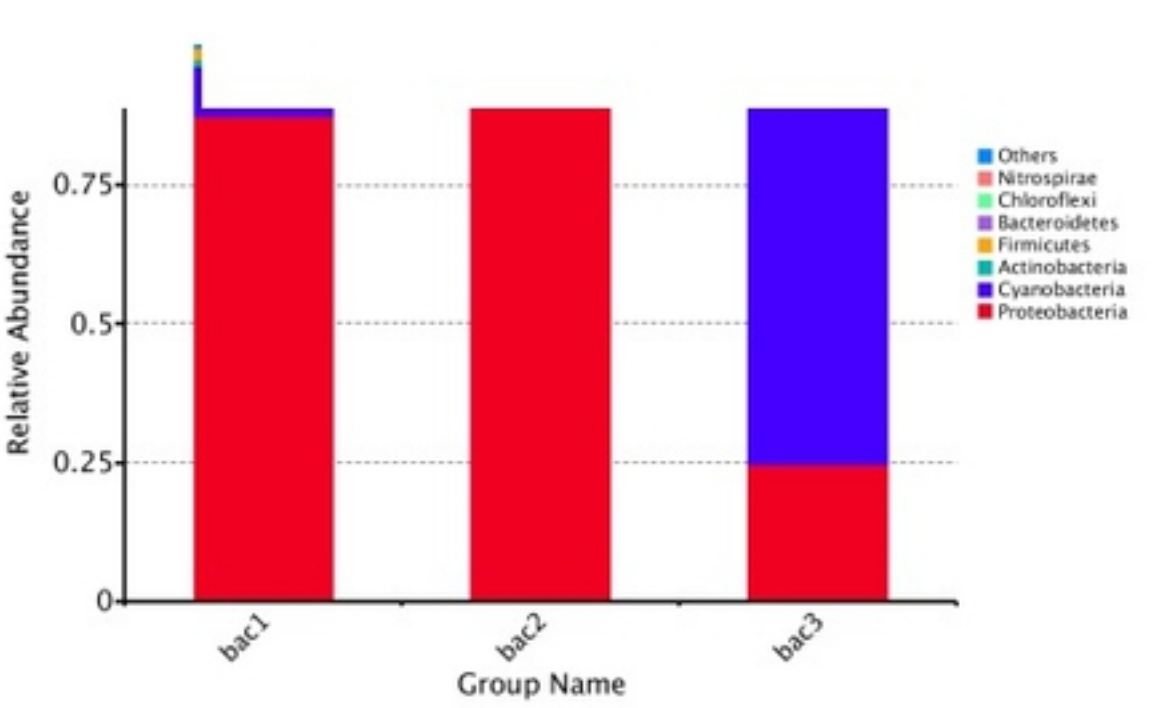

Relative Abundance
0.75
0.5
0.25
0
bac1
bac2
bac3
Group Name
Others
Nitrospirae
Chloroflexi
Bacteroidetes
Firmicutes
Actinobacteria
Cyanobacteria
Proteobacteria

Fig. 3

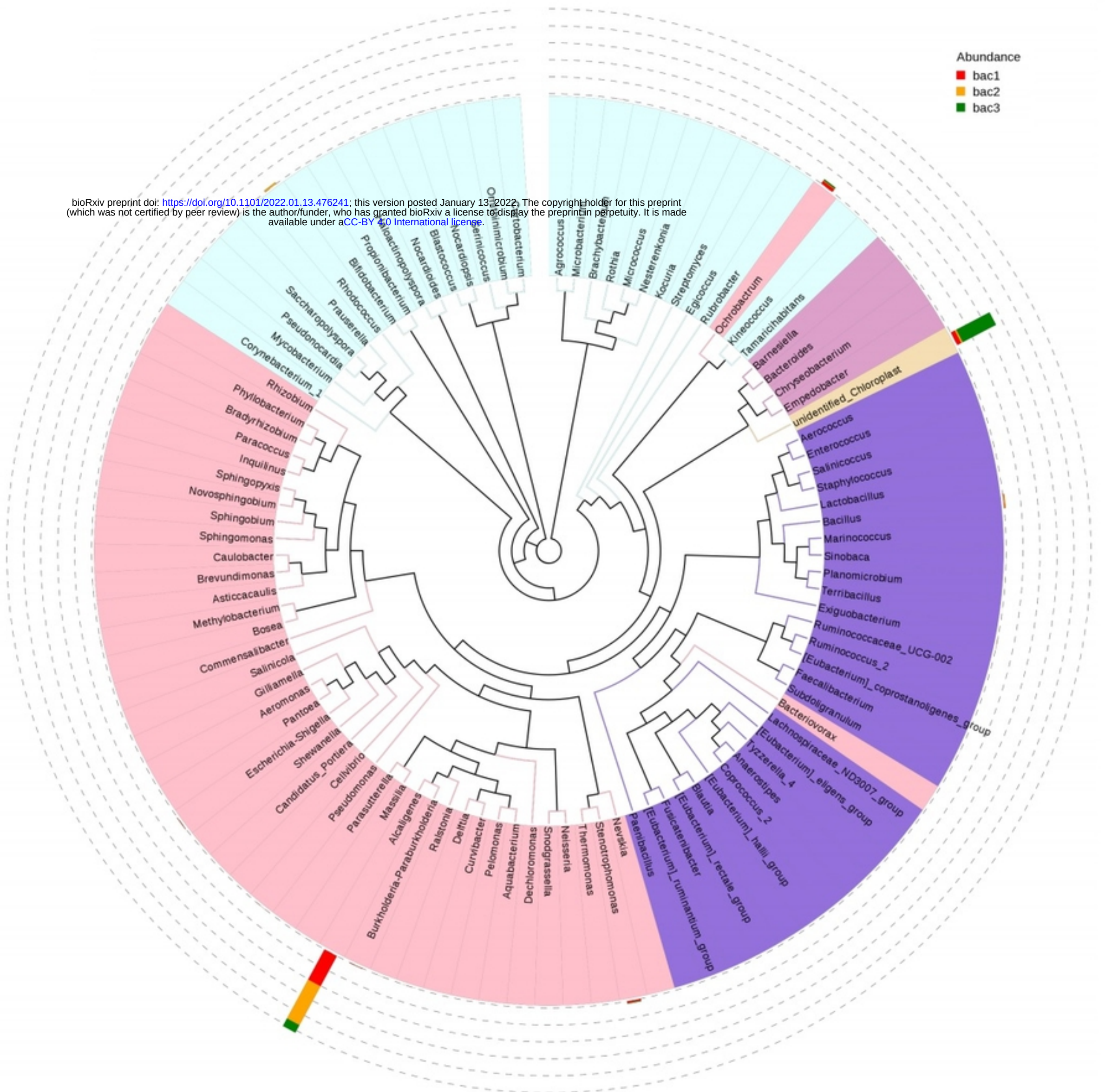


Fig. 4

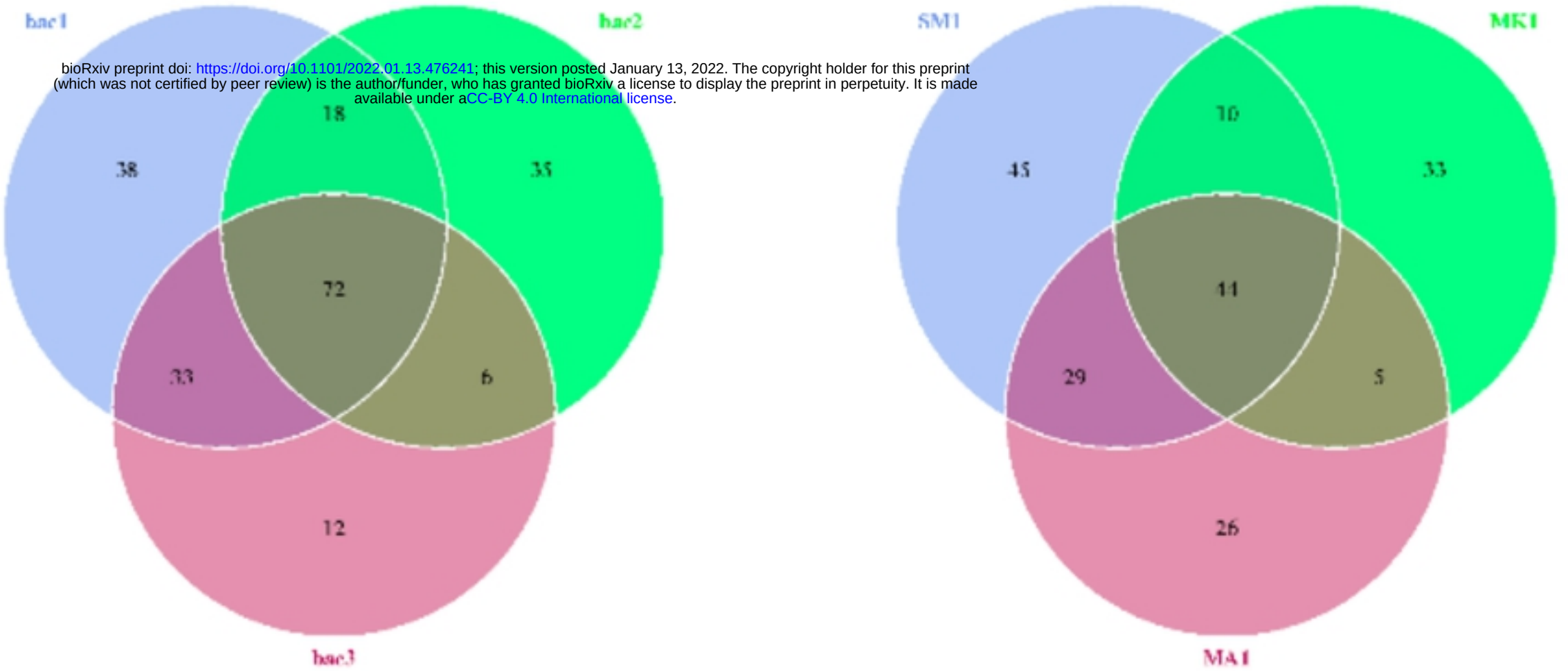


Fig. 5

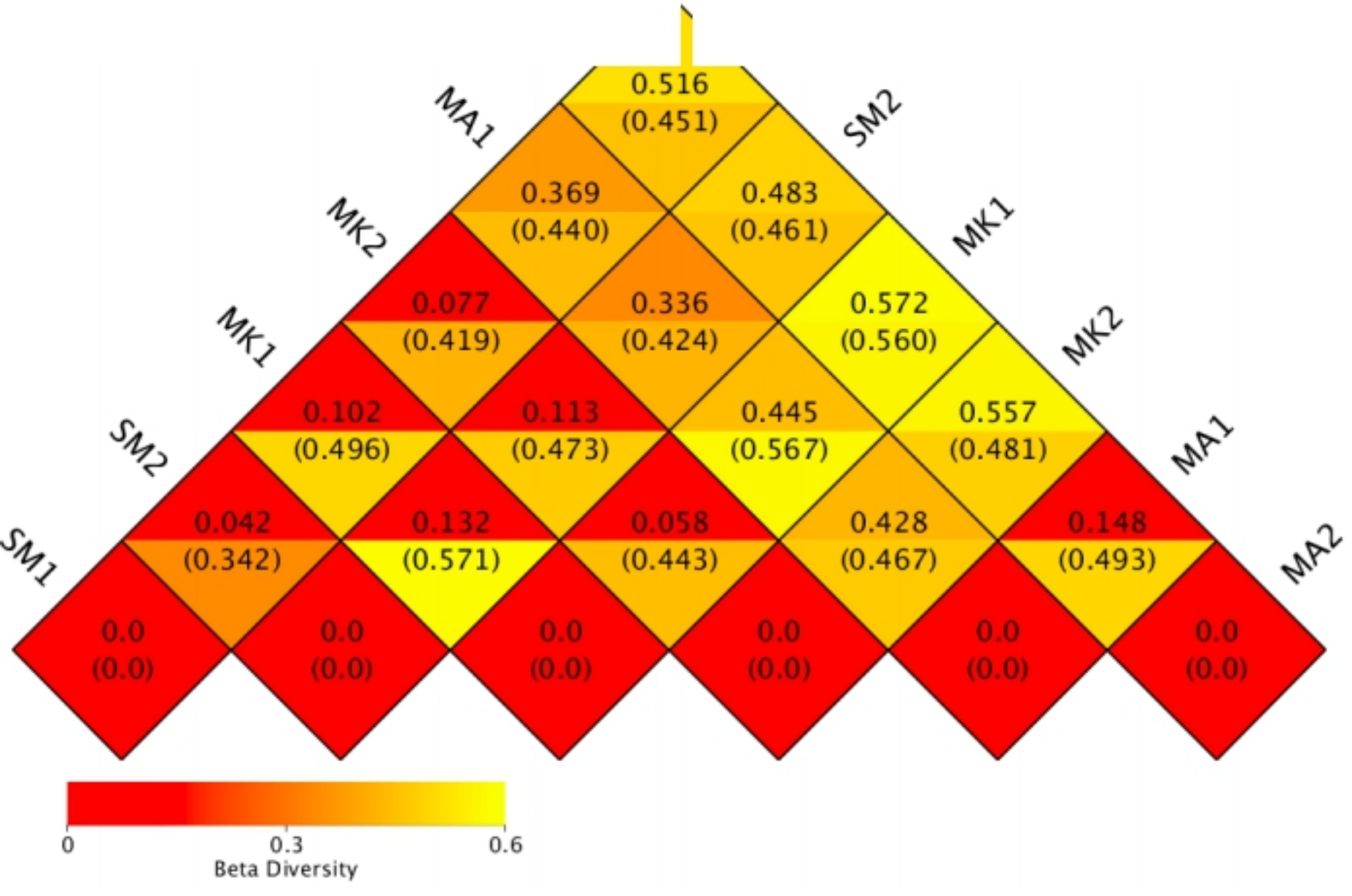

MA1
SM2
MK2
MK1
MK1
MK2
SM2
MA1
SM1
MA2
0.516
(0.451)
0.369
(0.440)
0.483
(0.461)
0.077
(0.419)
0.336
(0.424)
0.572
(0.560)
0.102
(0.496)
0.113
(0.473)
0.445
(0.567)
0.557
(0.481)
0.042
(0.342)
0.132
(0.571)
0.058
(0.443)
0.428
(0.467)
0.148
(0.493)
0.0
(0.0)
0.0
(0.0)
0.0
(0.0)
0.0
(0.0)
0.0
(0.0)
0.0
(0.0)
0
0.3
0.6
Beta Diversity

Fig. 6

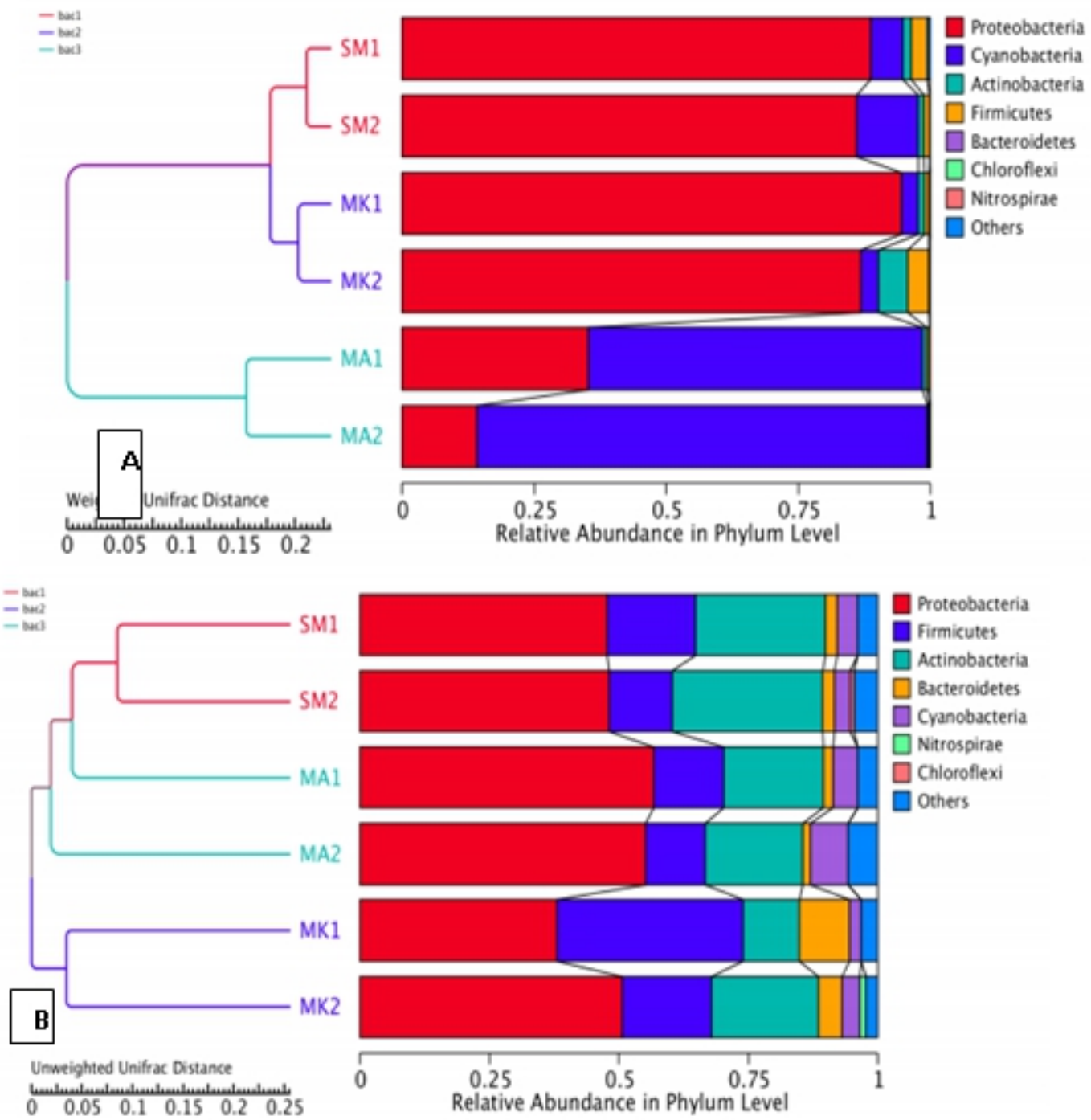

bac1
bac2
bac3
SM1
SM2
MK1
MK2
MA1
MA2
A
Unifrac Distance
0
0.05
0.1
0.15
0.2
Proteobacteria
Cyanobacteria
Actinobacteria
Firmicutes
Bacteroidetes
Chloroflexi
Nitrospirae
Others
0
0.25
0.5
0.75
1
Relative Abundance in Phylum Level
B
Unweighted Unifrac Distance
0
0.05
0.1
0.15
0.2
0.25

Fig. 7