Application Note

Metagenomics Service for Profiling Microbial Communities (16S rDNA & ITS)

Introduction

Amplicon deep-sequencing using next generation sequencing (NGS) technologies has become a powerful tool to study the diversity of microbial communities. By sequencing parts of the ribosomal DNA (16S rDNA or ITS) derived from environmental samples NGS can generate unprecedented amounts of sequence data permitting rapid and profound analysis of microbial communities. However, the processing and evaluation of NGS sequence data is a challenge due to the large amount of data generated.

Importance of Selecting the Appropriate Barcoding Locus

Metagenomic studies are commonly performed by analyzing amplicons from the 16S rDNA in prokaryotes or the internal transcribed spacer regions (ITS) in fungi (Figure 1). Both loci form a mosaic of highly conserved and hypervariable regions, the latter being used for phylogenetic classifications.

It is yet not possible to sequence amplicons spanning the entire 16S rDNA (~1.3 kb) or ITS regions because the read length of current NGS technologies is limited. Which parts of the 16S rDNA or ITS are best amplified for the profiling is still under debate, and varies depending on the study objectives, experimental design and type of sample. In general, if you aim to get a fine-scale taxonomic resolution, your objective should be to cover as many variable regions as possible in your PCR. The challenge of any profiling project is, however, to carefully balance the trade-off between the taxonomic resolution and possible bias from PCR amplification resulting in an under-representation or loss of taxonomic entities [1] [2]. Often only a pilot study will help to find the best primer set for a specific research question.

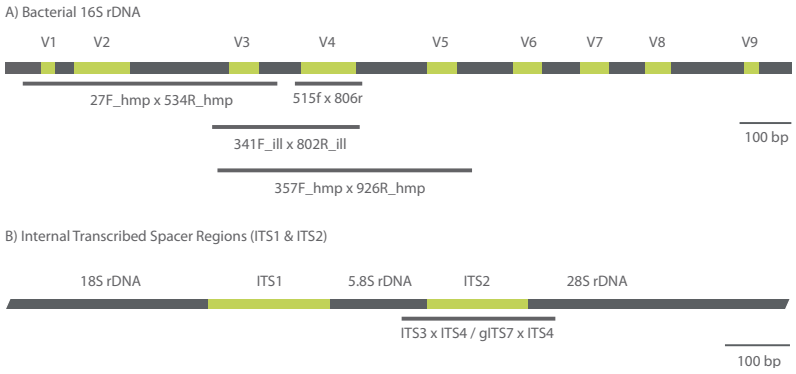


Figure 1. Overview of the ribosomal gene loci most often used for the analysis of microbial communities. **1A**. Structure of the 16S rDNA for bacterial species showing the 9 different variable regions (V1-V9) and common regions often used for community analysis. **1B**. Structure of the internal transcribed spacer regions (ITS) for fungi showing the two variable regions ITS1 and ITS2. Most often ITS2 is used for profiling fungal communities. Variable regions are green whereas conserved regions are grey.

Microsynth Competences and Services

One of Microsynth's core competence areas is the profiling of microbial communities based on 16S rDNA analysis. Microsynth is able to offer its customers a full service covering the entire process from experimental design planning, DNA isolation, PCR amplification and sequencing up to a detailed bioinformatics analysis of the generated data (Figure 2).

DNA Isolation: Either the customer provides isolated DNA or outsources this critical step to Microsynth (>13 years of experience in DNA/RNA isolation from various and demanding matrices like plant material, food or stool).

PCR Amplification: The PCR amplification will follow a two-step PCR protocol using a state-of-the-art high-fidelity polymerase. This two-step PCR is applied in order to increase reproducibility and to improve the production of high-quality multiplex amplicon libraries. PCR products are purified, quantified with fluorescence spectroscopy using Picogreen and pooled in equimolar amounts.

Microsynth AG | Schützenstrasse 15 | 9436 Balgach | Switzerland
Phone: +41 71 722 83 33 | E-mail: info@microsynth.ch | Web: www.microsynth.ch

NGS Sequencing: Sequencing is done using Illumina MiSeq sequencing technology. MiSeq allows high-throughput profiling at low costs with the advantage of long reads (up to 570 bp).

Bioinformatics Analyses: Depending on customer requirements Microsynth offers either a basic bioinformatics analysis package or an advanced bioinformatics analysis package. Both packages are based on the Qiime software [3]. The basic package provides taxonomic profiles and α-diversity measurements (i.e. diversity of organisms in one sample). The advanced package builds on the basic package and adds β-diversity measurements (i.e. diversity of organisms across samples). Guided by experimental parameters (e.g. different sample conditions like temperature, pH, etc.), the comparative analysis of the advanced package allows to test the experimental hypothesis.

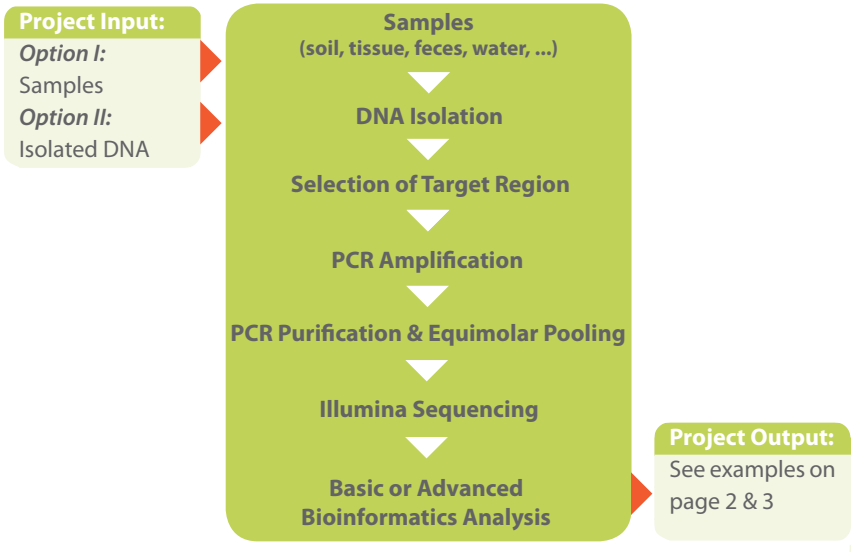


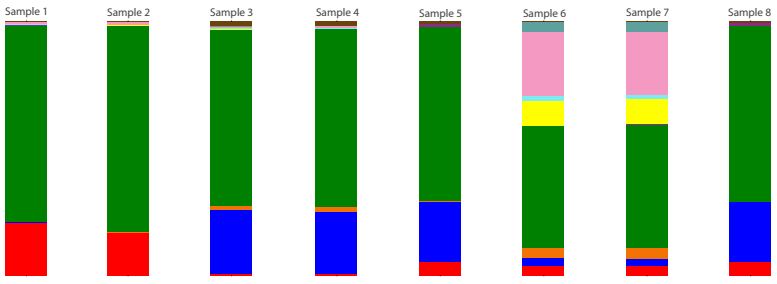
Figure 2. Typical steps in a profiling project. Depending on researcher needs Microsynth can deliver a one-stop service starting from DNA isolation and PCR amplification & sequencing right through to data analysis.

Microbial Profiling with Basic Bioinformatics Package

Data Processing: Input for further data analysis are demultiplexed, stitched (combining forward and reverse reads), quality filtered and Chimera cleaned fasta reads.

Data Enrichment: The Qiime software in combination with a reference database is used to define operational taxonomic units (OTUs) and to assign each OTU to a taxonomic entity at different taxonomic levels (Figure 3). Output may serve as source for down-stream analysis done by the customer.

α-Diversity: Chao1 and Shannon diversity measurements are calculated. Rarefaction analysis is also performed to estimate if sampling has been exhaustive (Figure 4).



Legend	Taxonomy	Total		Sample1		Sample2		Sample3		Sample4		Sample5		Sample6		Sample7		Sample8	
		count	%	count	%	count	%	count	%	count	%	count	%	count	%	count	%	count	%
	Bacteria;Actinobacteria;Coriobacteriia	185101	6.70%	68641	20.90%	43430	16.80%	2784	0.70%	2727	0.70%	19580	5.40%	13319	3.90%	13461	3.80%	21159	5.70%
	Bacteria;Bacteroidetes;Bacteroidia	378227	13.70%	347	0.10%	272	0.10%	93468	25.10%	91165	24.50%	86392	23.80%	10735	3.10%	10202	2.90%	85646	23.20%
	Bacteria;Firmicutes;Bacilli	44996	1.60%	725	0.20%	636	0.20%	6884	1.80%	6918	1.90%	665	0.20%	13543	3.90%	14955	4.30%	670	0.20%
	Bacteria;Firmicutes;Clostridia	1819457	65.90%	253478	77.10%	209401	81.00%	257739	69.10%	259693	69.80%	248809	68.50%	166074	48.10%	169459	48.40%	254804	69.00%
	Bacteria;Firmicutes;Erysipelotrichi	9608	0.30%	239	0.10%	184	0.10%	300	0.10%	288	0.10%	4110	1.10%	24	0.00%	26	0.00%	4437	1.20%
	Bacteria;Proteobacteria;Alphaproteobacteria	71317	2.60%	601	0.20%	490	0.20%	19	0.00%	22	0.00%	131	0.00%	34607	10.00%	35302	10.10%	145	0.00%
	Bacteria;Proteobacteria;Betaproteobacteria	15941	0.60%	755	0.20%	422	0.20%	1556	0.40%	1508	0.40%	446	0.10%	5488	1.60%	5299	1.50%	467	0.10%
	Bacteria;Proteobacteria;Deltaproteobacteria	186441	6.80%	3811	1.20%	3629	1.40%	2438	0.70%	2385	0.60%	66	0.00%	87152	25.20%	86878	24.80%	82	0.00%
	Bacteria;Proteobacteria;Epsilonproteobacteria	29008	1.10%	0	0.00%	0	0.00%	340	0.10%	326	0.10%	0	0.00%	14186	4.10%	14156	4.00%	0	0.00%
	Unassigned;Other;Other	20211	0.70%	109	0.00%	73	0.00%	7318	2.00%	7038	1.90%	3013	0.80%	393	0.10%	310	0.10%	1957	0.50%

Figure 3. Example for an interactive HTML taxonomy plot on the Class level. The taxonomy is given for different taxonomic levels (phylum, class, order, family, genus, species).

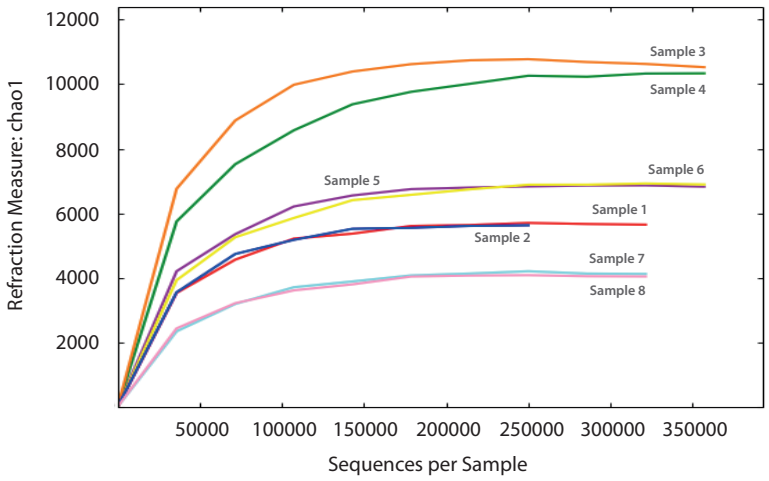


Figure 4: Rarefaction curves are calculated for each sample based on the OTU computations. Rarefaction curves help to estimate whether bacterial communities were sampled properly, i.e. enough sequence reads per sample where collected. Rarefaction curves are expected to reach a plateau if sampling has been exhaustive [4].

Microbial Profiling with Advanced Bioinformatics Package

The Advanced Bioinformatics Analysis includes the results of the basic analysis. In addition, the microbial community structures are compared against the environmental parameters provided by the customer in a phylogenetic context.

The OTU dataset obtained with the basic bioinformatics package serves as input for further analysis.

Data Enrichment: In a first step a phylogenetic tree is calculated for all OTUs in the dataset (UniFrac)[5]. The phylogenetic tree serves as basis for the comparative analysis and calculation of β-diversity.

β-Diversity: A qualitative overview of intersample diversity is obtained by means of principal coordinate analysis (PCoA, Figure 5A). In addition, significance of pairwise phylogenetic differences among communities is computed and hierarchical UPGMA based clustering is performed (Figure 5B).

Comparative Analysis: A quantitative assessment is performed to test the experimental hypothesis which answers (i) whether the sample categories differ from each other and (ii) whether OTUs are differentially represented based on sample categories (e.g. different sample conditions).

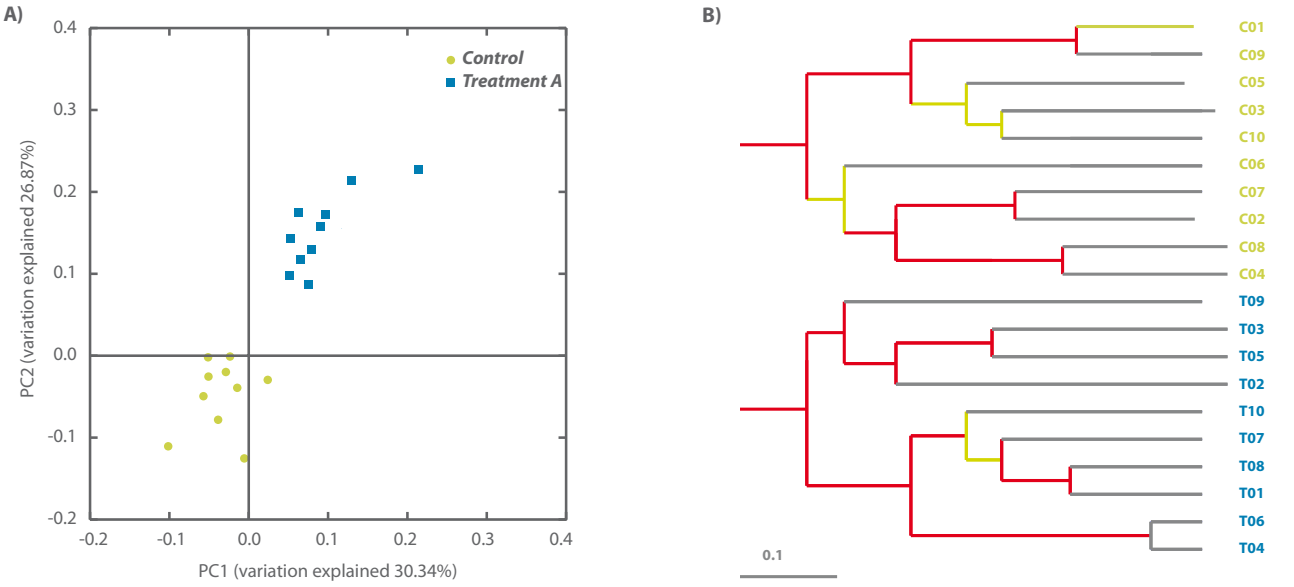


Figure 5. Example output of the comparative analyses (β-diversity). **5A.** 2D plot of a PCoA analysis based on the UniFrac distance matrix. PCoA reduces the dimensionality of a complex dataset and can be used to visualize the relationship of each sample to any other sample in the set. **5B.** UPGMA clustering based on the UniFrac distance matrix. UPGMA clustering discovers the hierarchical relationships that underlie the samples.

Primer Systems and Their Possible Effects on Profiling Results

Primer Name	Sequence (5'-3')	Region	Size (bp)	Source
515f	GTG CCA GCM GCC GCG GTA A	V4	~300	EMP
806r	GGA CTA CHV GGG TWT CTA AT			
27F_hmp	AGA GTT TGA TCC TGG CTC AG	V123	~510	HMP
534R_hmp	ATT ACC GCG GCT GCT GG			
357F_hmp	CCT ACG GGA GGC AGC AG	V345	~570	HMP
926R_hmp	CCG TCA ATT CMT TTR AGT			
341F_ill	CCT ACG GGN GGC WGC AG	V34	~460	Illumina
802R_ill	GAC TAC HVG GGT ATC TAA TCC			
ITS3	GCA TCG ATG AAG AAC GCA GC	ITS2	~300-400	
ITS4	TCC TCC GCT TAT TGA TAT GC			
gITS7	GTG ART CAT CGA RTC TTT G	ITS2	~230-330	
ITS4	TCC TCC GCT TAT TGA TAT GC			

Table 1. Summary of most common primer systems for profiling bacterial and fungal communities. Only template-specific sequences and its spanning variable parts are shown. These sequences combined with Illumina adaptor sequences are used in the Illumina Nextera 2-step protocol for the library preparation. EMP = Earth Microbiome Project (http://www.earthmicrobiome.org/); HMP = Human Microbiome Project (http://www.hmpdacc.org/).

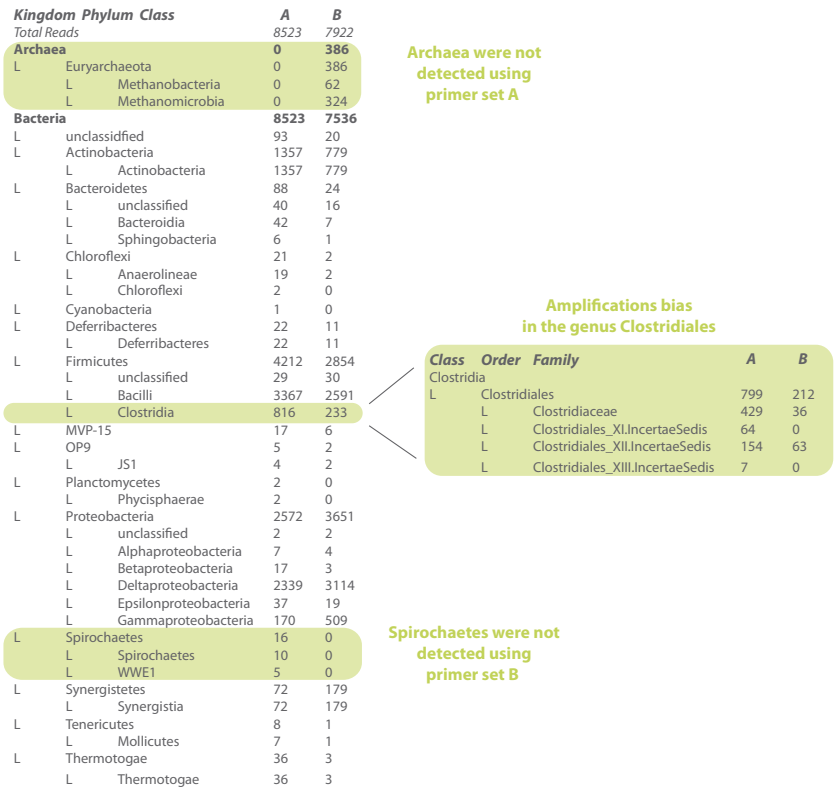


Figure 6. Analysis of the same environmental sample using different primer systems. In the first case (A) the V123 region was amplified using primers 9F & 534R (A) and in the second case the V345 region was amplified using primers 341F & 909R (B). This analysis clearly demonstrates that some of the taxa were not amplified using one of the primer sets, and also reveals differences in the amplification efficiency. For example the Archaea were detected using system B but not A. In contrast, the Spirochaetes were detected with primer system A but not B.

Literature

[1] Berry et al. (2011) Barcoded primers used in multiplex amplicon pyrosequencing bias amplification, App Env Microb, 77: 7846-7849.
[2] Klindworth et al. (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based divesity studies. Nucleic Acids Res 41, e1.
[3] Caporaso et al. (2010) Qiime allows analysis of high-throughput community sequencing data, Nature Methods, 7: 335–336.
[4] Colwell, R.K. & Coddington, J.A. (1994) Estimating terrestrial biodiversity through extrapolation. Philos T Roy Soc B, 345: 101–118.
[5] Lozupone, C. & Knight, R. (2005) UniFrac: a New Phylogenetic Method for Comparing Microbial Communities, Appl Environ Microbiol, 71: 8228-8235.