UVIC
UNIVERSITAT DE VIC
UNIVERSITAT CENTRAL
DE CATALUNYA

Master of Science in Omics Data Analysis

Master Thesis

Impact of the Nutritional Supplementation with DHA on Cystic Fibrosis Microbiota

by

Luz Balsalobre

Supervisor: Rosa del Campo Moreno, Department of Microbiology, University Hospital Ramón y Cajal, and Instituto Ramón y Cajal de Investigaciones Sanitarias (IRYCIS)

Department of Systems Biology

University of Vic – Central University of Catalonia

27 September 2018

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Cystic fibrosis (CF) is a genetic disease characterized by the formation of thick secretions in the gut and airways, among other anatomical locations, which leads to an altered microbiota. Understanding the significance of bacterial species that colonize and persist in CF patients requires a detailed examination of bacterial community structures. In addition, deficiency in fatty acids, such as docosahexaenoic acid (DHA), is a common co-morbidity in CF patients who are malnourished. Thus, the restoration of DHA levels may have influence in the bacterial composition of CF microbiota.

The aims of this study were, first, to analyze the gut and airway microbiota of CF patients using next generation sequencing (NGS) tools, secondly, to determine the DHA supplementation impact in the microbiota of both ecosystems, and finally, to associate particular bacterial profiles to different clinical variables as nutritional status and lung function.

Forty-eight CF patients were randomized into two groups and receiving either DHA supplemented diet or placebo for 6 months. Fecal or sputum samples were collected before and after finishing the DHA or placebo supplementation (two samples per each patient). We performed V3-V4 16S rDNA NGS analysis to characterize the gut and lung microbiota at basal status in these patients. The results showed that gut and airways CF microbiota was aberrant, with significant inter-individual variability in their community structure. The core CF-gut microbiota consisted of *Lachnospiraceae* family and *Streptococcus* and *Blautia* genera. Airways in children were dominated by *Streptococcus* and *Haemophilus* whereas in adults *Streptococcus* was the only genera over 10% abundance. *Lachnospiraceae* abundance in CF-gut and children airways was linked to the effect of DHA supplementation in the bacterial composition. Only considering the top-28 most abundant genera in adult CF-airways, we were able to detect an association between strict anaerobes, such as *Prevotella*, *Veillonella* and *Lachnospiraceae,* and CF patients with lower lung function.

1. INTRODUCTION

1.1. Cystic fibrosis

Cystic fibrosis (CF) is the most common life-limiting autosomal recessive disease in Caucasians, with a reported incidence of approximately 1 in 3,000-4,000 live births (1). The disease is caused by the absence or dysfunction of the cystic fibrosis transmembrane conductance regulator (CFTR) protein, a transmembrane cAMP-activated chloride channel which is expressed primarily at the apical plasma membrane of secretory epithelia in the airways, pancreas, intestine and other organs. Both copies of the *CFTR* gene, located on chromosome 7q31.2, are mutated in clinical disease and there are over 2,000 different mutations that can cause disease. These mutations are divided into 6 categories according to the different mechanism by which they are known to disrupt CFTR production and function: defective protein synthesis, defective protein processing, disordered regulation, defective chloride conductance, reduced amount of protein and accelerated channel turnover (2). The absence or dysfunction of CFTR protein results in poor chloride and bicarbonate transport, which causes dehydration of secretions with viscous mucus and leads to different pathologies. The classic phenotype of the CF, or mucoviscidosis, is characterized by chronic bacterial lung colonization/infection, pancreatic insufficiency, malabsorption, liver disease, male infertility, and loss of sweat electrolytes. Although many organs are affected in CF, lung disease is the major cause of morbidity and mortality due to the accumulation of mucus in the lungs, that makes CF airways vulnerable to recurrent infections and permanent inflammation (1). In addition, most CF patients (≥85%) have pancreatic exocrine insufficiency, even in early infancy, due to ductal obstruction, which could lead to fat malabsorption and malnutrition (3). Regarding fat malabsorption, CF individuals display fatty acid profiles characterized by decreased linoleic acid (LA) and docosahexaenoic acid (DHA) levels (4). Animal models and research in patients with CF demonstrated that an oral administration of DHA corrects this lipid imbalance and ameliorates some CF pathological manifestations (5, 6). Moreover, the association between nutritional status and lung function has been long recognized, as numerous studies have noted the negative effect of malnutrition on respiratory function and prognosis (7-9). However, although malnutrition is a preventable or potentially treatable disorder, 9.5% of the patients with CF in Europe still suffer from chronic severe malnutrition (10).

1.2. Airway microbiota in CF patients

The microbiota colonizing/infecting the airway is particularly relevant in CF. Actually, bacterial lung infections are the main cause of reduced life expectancy in CF patients, causing a progressive loss of lung function accompanied by an increased burden of chronic infections (1, 11). *Pseudomonas aeruginosa*, methicillin resistant *Staphylococcus aureus* (MRSA), and *Mycobacterium abscessus* play an essential role in chronic CF lung infections (12). Other bacteria frequently isolated from the sputum of patients with CF by standard microbiological methods are *Haemophilus influenzae*, *Burkholderia cepacia* complex, *Stenotrophomonas maltophilia* and *Achromobacter* spp. The prevalence of these species in the airways changes over time; *S. aureus* is often the first respiratory pathogen identified in the respiratory secretions in CF young children, whereas *P. aeruginosa* is the most common pathogen found in adults and its appearance, especially the mucoid phenotype, is associated with the course of lung disease in CF (13). The advent of high-throughput sequencing technology for microbial community analysis has shown the diversity and complexity of the airway microbiota in CF patients, revealing that the aforementioned small group of CF airway-associated microbes is, in fact, part of a larger and more complex CF lung bacterial community. 16S rDNA gene sequencing has allowed the detection and identification of strict anaerobes, such as *Prevotella* spp. and *Veillonella* spp. but also non-cultivable microorganisms, which would not have been detected using currently recommended diagnostic culture methods for CF respiratory samples. Although a large number of studies have revealed a significant anaerobic burden in CF airways (14), the role of anaerobic bacteria in the pathogenesis of infection and inflammation in the CF airways remains unclear. A recent study demonstrated a high prevalence of anaerobes in the CF airways and showed that its prevalence increases with age (15). This study also mentioned that strict anaerobes were able to secrete short-chain fatty acids, molecules that induce a significant IL-8 response in bronchial epithelial cells, a phenomenon which is more pronounced in CF than in normal bronchial epithelium (15). Based on the knowledge about CF lung microbiota and community ecology, Conrad *et al*. proposed the Climax-Attack Model (CAM) in which characteristic bacterial communities appear during exacerbations and times of stable disease (16). The CAM describes two functional microbial communities in CF airways: the attack and the climax. The attack community consists of microorganisms like *Streptococcus* spp.,

Staphylococcus spp. and non-mucoid *P. aeruginosa*. This community elicits strong host immune responses and scarring in the lungs. Then, the scarring tissue is colonized by the climax community, which consists of typical CF pathogens such as mucoid *P. aeruginosa*, *S. maltophilia* and *Achromobacter* spp. at the periphery and strict anaerobes such as *Prevotella*, *Veillonella* and *Fusobacterium* at the center of the scar. Therefore, the attack community would be a transient bacterial population associated with exacerbations of symptoms while the climax community would be a bacterial community associated with stable clinical states, more resilient to chronic host defenses and perturbations, such as those produced by antimicrobial therapy. This model predicts that with advancing disease, climax communities become increasingly prevalent. In relation with these findings, a recent report described an increase in the proportion of samples in which *Pseudomonas*, *Burkholderia*, or *Achromobacter* become the dominant taxon in airways communities sampled from CF patients with advancing disease stage (17). In addition, it has been described an increase in the relative abundance of anaerobes during lung exacerbations, at least during the early and intermediate stages of lung disease (18).

1.3. Next-generation sequencing and cystic fibrosis

Next-generation sequencing (NGS) is the reference methodology to decipher the structure and composition of human microbiota. The emergence of high-throughput sequencing approaches and the development of new bioinformatics pipelines designed to cope with metagenomics data, has revolutionized the study of complex bacterial communities. Massive sequencing of the bacterial 16S rDNA gene allows the identification and relative enumeration of the bacterial taxa within a clinical sample in a way that would be simply unachievable using culture-based approaches, while analyzing all of the genes in a sample (shotgun metagenomic sequencing) defines the functional potential encoded within the collective genomes present in a sample.

Mutations in the *CFTR* gene are known to alter the airway and the gastrointestinal microbiota of CF patients (51, 52). Several metagenomics studies have described how bacterial communities change over the course of CF disease and have revealed a different airway bacterial community structure in exacerbations (characterized by a reduced bacterial diversity and an increasingly conserved community composition) coupled with a decline in lung function (19), antibiotic treatment (20), and patient age (21). Changes

in total bacterial density was not observed with exacerbations, suggesting that shifts in the relative abundance of bacterial community members, rather than changes in total bacterial density, are more likely to be associated with alterations in clinical status (22). Similarly, NGS techniques were also used to describe gastrointestinal microbiota of CF patients. Burke *et al*. demonstrated that the CF gut microbiota had reduced microbial diversity, an increase in *Firmicutes* and a reduction in *Bacteroidetes* comparing with healthy controls. This study also reported individual particularities in the CF gut microbiota composition, even though grouped by clinical parameters including percentage predicted FEV_1 (measure of lung dysfunction) and number of intravenous antibiotic courses in the previous 12 months (23).

2. OBJECTIVES

Massive sequencing is revolutionizing the scientific knowledge of the microbial communities, showing that their structure and composition have profound effects in a variety of health and disease conditions. In this context, the objectives of this work were:

1- To analyze the gut and airway microbiota of CF patients in the basal status using NGS tools.
2- To determine the DHA supplementation impact in the airway and fecal bacterial ecosystems.
3- To associate particular bacterial profiles in both airway and fecal ecosystems to clinical variables as nutritional status and lung function.

3. MATERIAL AND METHODS

3.1. Patients and samples

Forty-eight CF patients from five different Spanish hospitals were enrolled in a blind clinical trial with DHA supplementation. Patients were randomized into two groups receiving either DHA supplemented diet or placebo for 6 months. Fecal or sputum samples were collected before and after finishing the DHA or placebo supplementation (two samples per each patient). Total DNA of each sample was obtained by the QiaAMP kit (Qiagen). Then, the V3-V4 region of the bacterial 16S rDNA gene was amplified, and the PCR products were pooled equally by following the Illumina 16S Metagenomic Sequencing Library Preparation guide. Finally, 16S rDNA massive sequencing was performed in a MiSeq (Illumina) platform at FISABIO (Valencia, Spain). In total, 76 sputum (34 samples from 17 patients receiving DHA and 38 samples from 19 patients with placebo) and 24 stool samples (10 from 5 patients with DHA and 14 samples from 7 patients with placebo) were finally collected (Fig 3.1).

Anthropometric and clinical variables such as sex, age, lung function, nutritional status and airway colonization by *P. aeruginosa* or *S. aureus* were collected from the clinical chart. Patients were categorized as children if age ≤10 years and adults if age >10 years. Malnutrition was defined by a body mass index (BMI) below 18.5 kg/m^2 in adults and by a BMI percentile <10th for children. The severity of lung disease was based on forced expiratory volume in 1 second (FEV_1) and categorized as low if FEV_1 ≤75% or normal if FEV_1 >75%.

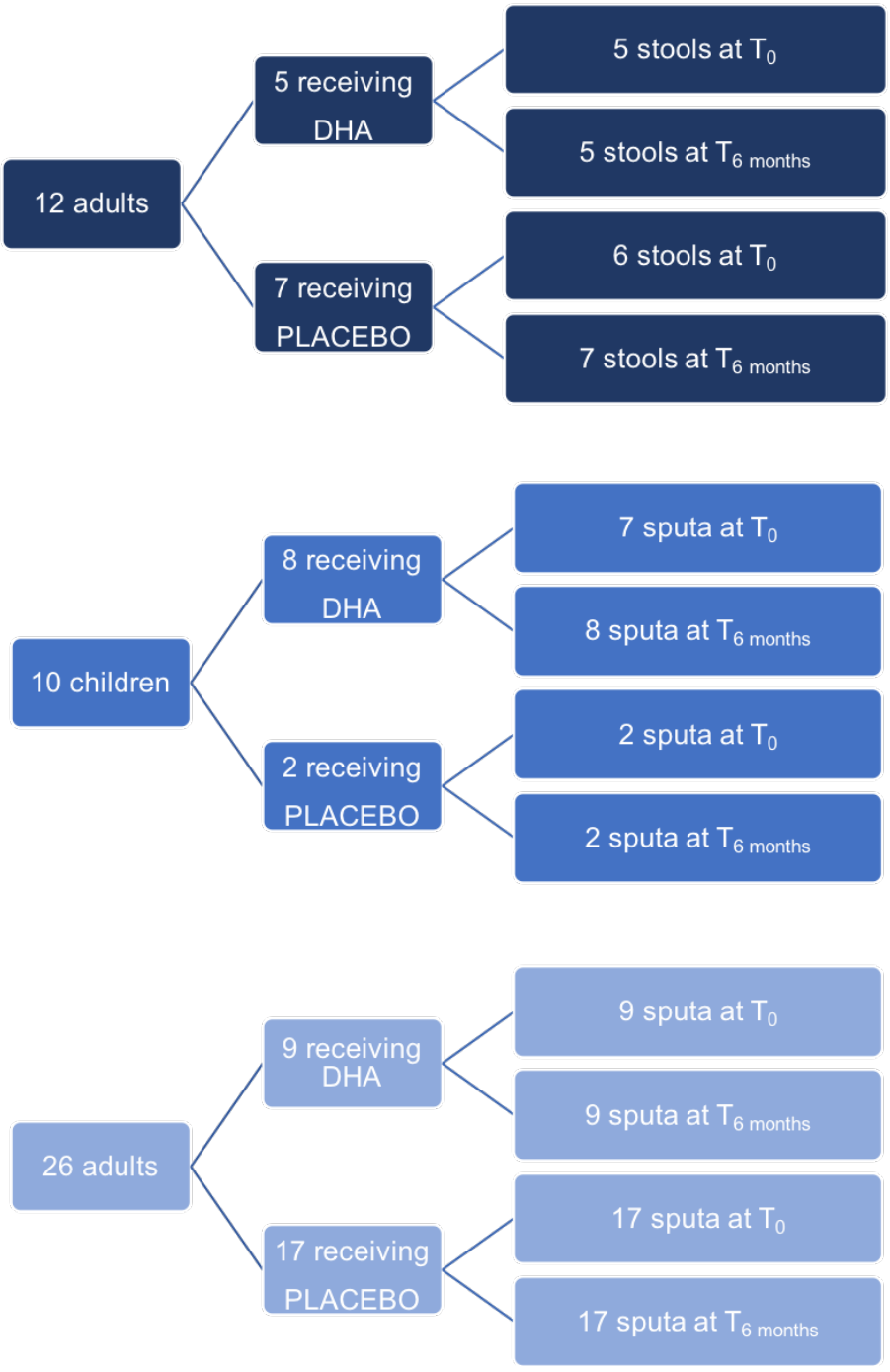

Figure 3.1. Patients and samples analyzed. Due to the low number of reads (<1,000) two samples were removed, a stool collected after 6 months of placebo and a sputum from a child before receiving DHA.

3.2. DNA Sequence pre-processing

Analysis of raw sequences (FASTQ files) was done with mothur software v.1.39.5 (24). Initial pre-processing and quality control was performed according to standard operating procedure (SOP) Miseq SOP (http://mothur.org/wiki/ MiSeq_SOP) (25) accessed on March 2018. Briefly, the two sets of reads for each sample (forward and reverse) were combined using *make.contigs* command. Sequences that failed to fulfil any one of the following criteria were excluded: maximum length of 600 bases, the presence of any ambiguities, maximum homopolymer length of 10 bases, and more than one nucleotide mismatch to the primer using *screen.seqs* command. Duplicated sequences were removed using *unique.seqs* command. SILVA v132 database (26, 27) was customized to the targeted V3-V4 region of 16S rDNA gene using the *pcr.seqs* command. Unique sequences were aligned using the SILVA customized database as reference. Sequences that did not align to the predicted V3-V4 region were removed from further analysis. Columns in the alignment that only contains gap characters were filtered out using *filter.seqs* command. Then, sequences were preclustered using *pre.cluster* command which could permit at most five nucleotide differences among sequences. Chimeras were detected using UCHIME algorithm and SILVA gold database using *chimera.uchime* command (28) and then removed using *remove.seqs* command. Then, singletons (n=1) were filtered out using *split.abund* command. Remaining sequences were taxonomically classified using the Naïve Bayesian classification with 80% confidence threshold by using *classify.seqs* command (29) and SILVA v.132 databases. Sequences that were not classified to any one of the domains (unknown) or classified in Chloroplast, Mitochondria, Eukaryota, and Archea were eliminated using *remove.lineage* command. A distance matrix was created and the sequences were clustered into operational taxonomic units (OTUs) at 3% dissimilarity cut-off (97% similarity) using *dist.seqs*, *cluster*, and *cluster.split* commands. Finally, the number of sequences in each OTU was determined using *make.shared* command and the taxonomy for each OTU was specified using *classify.otu* command.

3.3. Alpha and beta diversity measurements

Taxonomic alpha diversity and community richness were estimated using Shannon and Chao1 indices, respectively. Default parameters were applied using the phyloseq R package (30). The bacterial composition between different groups (beta-diversity) was

estimated using phylogenetic Bray-Curtis distance. Dissimilarity between samples was explored using NMDS analysis, calculated in R using the Bray-Curtis distance among samples. The Bray-Curtis dissimilarity was determined to measure how similar two communities are based on taxa present and their relative abundances. Heat maps of relative abundance were generated using the Heatmap.plus R package (31).

3.4. Statistical analysis

Statistical significance of the differences in mean alpha diversity and richness between samples was calculated using a Welch two-sample T-test in R. Statistical differences, at *phylum* and *genus* level, between samples before and after DHA supplementation were measured by Wilcoxon rank sum test with Benjamini-Hochberg correction to avoid increases of type error I due to multiple testing. Both similarities and dissimilarities in taxonomic composition among samples were visually assessed using non-metric multidimensional scaling ordinations (NMDS) based on the Bray-Curtis dissimilarity matrix after 20 runs. A permutational non-parametric multivariate analysis of variance was performed with the adonis function included in the vegan R package (32) to determine differences in dissimilarities among communities according to different variables. All statistical tests were conducted using R 3.4.3 version. Linear discriminate analysis effect size (LEfSe) was used in taxonomic comparisons (33). LEfSe combines the standard tests for statistical significance (Kruskal-Wallis test and pairwise Wilcoxon test) with linear discriminate analysis for feature selection. In addition to detecting significant features, it also ranks features by effect size, which put features that explain most of the biological difference at top. Alpha value for the factorial Kruskal-Wallis test was 0.05 and the threshold on the logarithmic LDA score for discriminative features was 2.0.

4. RESULTS

4.1. Sequencing data analysis

A total of 5,481,036 sequences were obtained from the 96 samples, and after trimming, dereplication, denoising, removal of chimeras, short sequences and singletons, 8,161 unique sequences were used for downstream analysis. The average number of processed sequences per sample was (mean ± SD) 51,659 ± 46,192. Sequences were classified into 1,784 OTUs that were distributed in 25 *phyla*, 52 classes, 124 orders, 214 families and 505 genera. In order to statistically adjust the number, samples with less than 1,000 reads were removed. Then, the OTU table containing the remaining samples was rarefied, before alpha-diversity metrics were performed, to 5,000 reads, to remove any bias from variation in sample read numbers. After filtering and rarefying, the final OTU table contained 90 samples and 1,164 OTUs.

4.2. Ordination analysis

As a preliminary description of samples, and to assess differences in bacterial community compositions, all the samples were ordinated in two-dimensional nonmetric multidimensional scaling (NMDS) plots based on Bray-Curtis dissimilarity measures.

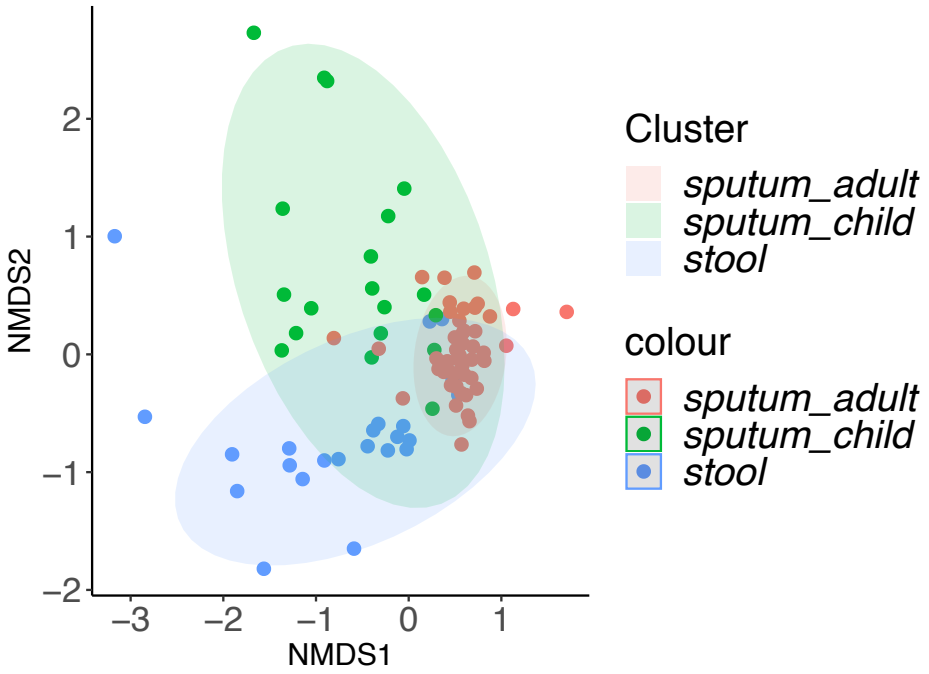

Figure 4.1: Non-metric multidimensional scaling (NMDS) plot displaying the spatial ordination among samples. NMDS plot was generated based on the Bray-Curtis distance. Each dot represents the overall bacterial community in each sample. Spatial ordination among 23 stool samples collected from adults (blue circles), 19 sputum samples collected from children (green circles) and 52 sputum samples from adults (red circles).

The NMDS plot showed that samples clustered by sample type, as well as by age category (Fig 4.1). Adonis was used to characterize the impact of different sample characteristics (age, sex, lung function, nutritional status and colonization by *P. aeruginosa* or *S. aureus*) on bacterial community structure. The largest proportion of explained variation within our data was due to differences in sample type (11.4%, $p<0.001$) and age category (12.2%, $p<0.001$), so that the airway samples analysis of children and adults was separately analyzed. However, it should be noted that the majority of the variation (~70%) remained unexplained. Figure 4.2 showed that the median number of OTUs was quite similar among the 3 groups of samples.

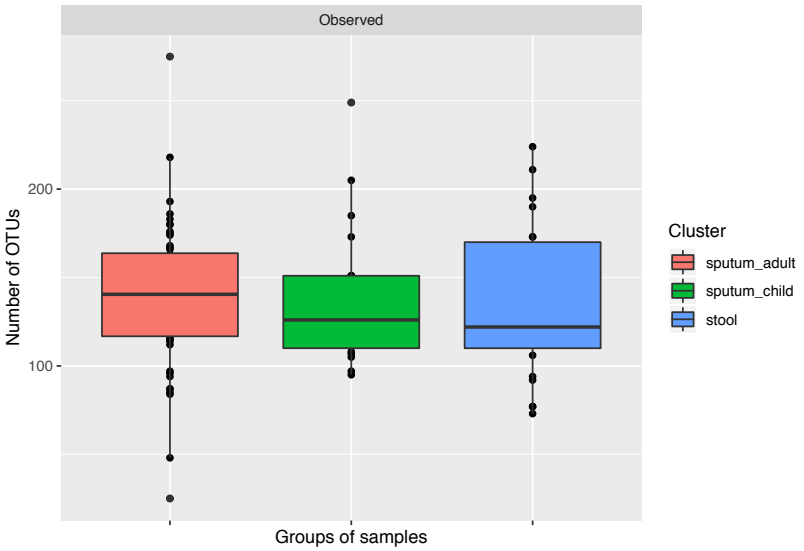

Figure 4.2. Number of OTUs detected among gut and airway samples.

4.3. Fecal samples at basal status

Fecal samples were collected from 12 patients without *P. aeruginosa* or *S. aureus* lung colonization. The samples analyzed in this section were collected from patients who received a placebo supplement or from patients before receiving a DHA supplemented diet.

4.3.1. Alpha-diversity of fecal samples

The demographic and clinical characteristics of patients that contributed with fecal samples are depicted in Table 4.1. Significant differences of the gut microbiota with clinical variables as sex, lung function, nutritional status, taxonomic richness (number of OTUs detected), or taxonomic diversity (metric based on richness and abundance) were not detected.

Table 4.1. Relationship among demographic or clinical patient characteristics and fecal bacterial alpha-diversity metrics. Data are presented as n (%), mean ± SD and p-value.

Patient characteristics No. samples =16		Alpha-diversity metrics	
		Richness (Chao1 index)	Alpha-diversity (Shannon index)
Sex			
Female	n=6 (37.5%)	83.94 ± 21.86	2.68 ± 0.52
Male	n=10 (62.5%)	76.91 ± 11.86	2.76 ± 0.27
p-value		0.4923	0.7278
Lung function			
FEV_1 <=75%	n=9 (56.2%)	78.06 ± 13.25	2.78 ± 0.27
FEV_1 >75%	n=7 (43.8%)	81.46 ± 19.97	2.67 ± 0.49
p-value		0.7058	0.6166
Nutritional status			
Normal	n=14 (87.5%)	80.59 ± 16.90	2.72 ± 0.39
Malnutrition	n=2 (8.7%)	72.25 ± 2.48	2.81 ± 0.06
p-value		0.1084	0.4595

4.3.2. Beta-diversity of fecal samples

OTUs were grouped according to *phylum* and genus to determine the median relative abundances at these taxonomic levels. Microbial taxonomic composition is shown in Figure 4.3.

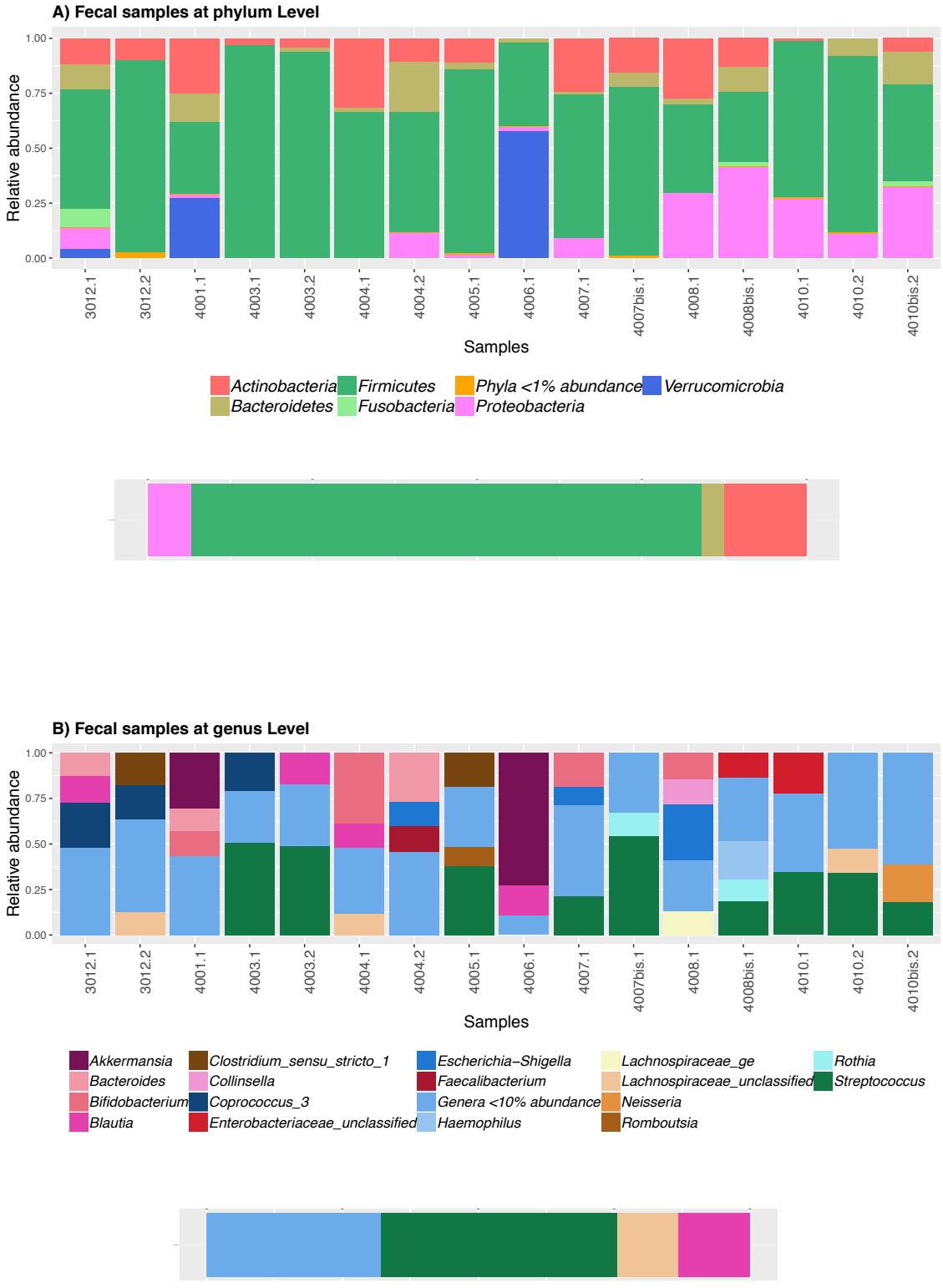

Figure 4.3. *Phyla* and genera distribution in fecal samples. Vertical bars represent the relative abundance of main taxa for each sample whereas horizontal bars represent the median relative abundance of main taxa in the studied population.

The dominant *phyla* in CF-fecal microbiota were *Firmicutes* (77.4%), *Actinobacteria* (12.5%), *Proteobacteria* (6.6%) and *Bacteroidetes* (3.4%), together accounting for 99.9% of all sequences. The *Firmicutes/Bacteroidetes* ratio was 22.8 (Fig 4.3A). At the genus level, only 3 genera displayed over 10% abundance, *Streptococcus* (43.6%), *Blautia* (13.2%) and unclassified *Lachnospiraceae* (11.1%), together accounting for 67.9% of all sequences. The remaining 32.1% comprised 47 genera. Fecal samples showed a high variability, showing very different genera distribution patterns (Fig. 4.3B). In fact, unsupervised clustering analysis did not indicate a clear clustering of these samples at the genus level.

4.4. Airway samples at basal status

4.4.1. Analysis of airway samples in children at basal status

A total of 10 children contributing with two sputum samples (placebo) or only one sputum just before receiving a DHA supplemented diet.

4.4.1.1. Alpha-diversity of airway samples in children

The main characteristics of children were shown on Table 4.2. Sputum carried *P. aeruginosa* were not detected in any children, whereas 4 out 10 were chronically colonized by *S. aureus*. The airway bacterial community have not significant association with clinical variables as sex, nutritional status, *S. aureus* colonization, and taxonomic richness and diversity. Lung function was not analyzed due to lack of FEV_1 data.

Table 4.2. Relationship among demographic or clinical patient characteristics and airway bacterial alpha-diversity metrics in children. Data are presented as n (%), mean ± SD or p-value.

Patient characteristics No samples =10		Alpha-diversity metrics	
		Richness (Chao1 index)	Alpha-diversity (Shannon index)
Sex			
Female	n=8 (80.0%)	88.038 ± 62.360	1.922 ± 1.458
Male	n=2 (20.0%)	117.611 ± 13.435	3.019 ± 0.021
p-value		0.2533	0.071
Nutritional status			
Normal	n=6 (60.0%)	112.767 ± 67.967	2.452 ± 1.648
Malnutrition	n=4 (40.0%)	65.747 ± 11.542	1.675 ± 0.772
p-value		0.1535	0.3480
S. aureus colonization			
Yes	n=4 (40.0%)	96.552 ± 25.550	2.370 ± 0.750
No	n=6 (60.0%)	92.219 ± 73.209	1.989 ± 1.719
p-value		0.8979	0.6460

4.4.1.2. Beta-diversity of airway samples in children

OTUs were grouped according to *phylum* and genus to determine the median relative abundances at these taxonomic levels. Microbial taxonomic composition is shown in Figure 4.4.

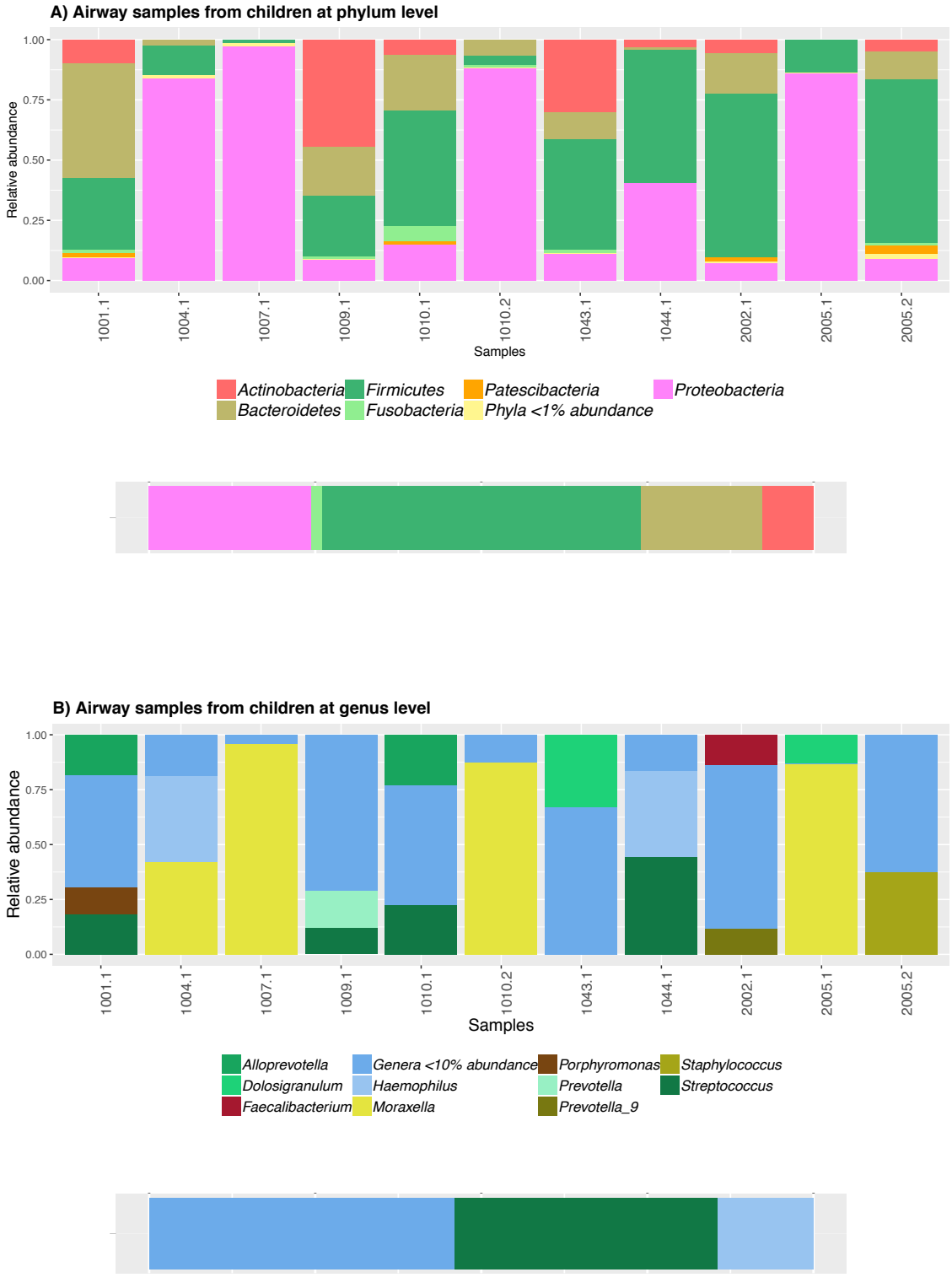

Figure 4.4. *Phyla* and genera distribution in airway samples in children. Vertical bars represent the relative abundance of main taxa for each sample whereas horizontal bars represent the median relative abundance of main taxa in the studied population.

Dominant *phyla* in children airways were *Firmicutes* (47.8%), *Proteobacteria* (24.4%), *Bacteroidetes* (18.4%), *Actinobacteria* (7.6%), and *Fusobacteria* (1.7%), together accounting for 99.9% of all sequences. The *Firmicutes/Bacteroidetes* ratio was 2.6 (Fig. 4.4A). At the genus level, only 2 genera displayed an over 10% abundance, *Streptococcus* (39.6%) and *Haemophilus* (14.4%), together accounting for 54.0% of all sequences, the remaining 46.0% was comprised for 51 genera. Among the most abundant genera (median relative abundance >1%), 53.3% were strict anaerobes. Airway microbiota in children showed a high interindividual variability, with unrelated genera patterns (Fig. 4.4B). In fact, unsupervised clustering analysis did not indicate a clear clustering of children airway samples at the genus level.

4.4.2. Analysis of airway samples in adults at basal status

Twenty-six adults were included and for those supplemented with DHA only their first sputum was considered as basal. Chronic lung colonization by *P. aeruginosa* (10 patients) and *S. aureus* (20 patients) was a common feature of this subpopulation.

4.4.2.1. Alpha-diversity of sputum samples in adults

The clinical data of the 26 adults are detailed in Table 4.3. As previously occurred in the fecal samples and the sputum from children, significant associations between the microbiota composition and the clinical characteristics of the patients were not found.

Table 4.3. Relationship among demographic or clinical patient characteristics and airway bacterial alpha-diversity metrics in adults. Data are presented as n (%), mean ± SD or p-value.

Patient characteristics No samples =42		Alpha-diversity metrics	
		Richness (Chao1 index)	Alpha-diversity (Shannon index)
Sex			
Female	n=23 (54.8%)	89.907 ± 33.644	2.777 ± 0.539
Male	n=19 (45.2%)	93.806 ± 28.116	2.781 ± 0.522
p-value		0.6846	0.9795
Lung function			
FEV_1 <=75%	n=17 (40.5%)	98.649 ± 30.545	2.886 ± 0.448
FEV_1 >75%	n=25 (59.5%)	86.926 ± 30.938	2.706 ± 0.568
p-value		0.2327	0.6000
Nutritional status			
Normal	n=30 (71.4%)	93.494 ± 32.120	2.780 ± 0.555
Malnutrition	n=12 (28.6%)	87.115 ± 28.639	2.776 ± 0.462
p-value		0.5355	0.9800
P. aeruginosa colonization			
Yes	n=10 (23.8%)	84.463 ± 25.908	2.588 ± 0.466
No	n=32 (76.2%)	93.924 ± 32.412	2.838 ± 0.535
p-value		0.3561	0.1704
S. aureus colonization			
Yes	n=20 (47.6%)	98.038 ± 29.695	2.915 ± 0.499
No	n=22 (52.4%)	85.883 ± 371.614	2.655 ± 0.528
p-value		0.2063	0.1086

4.4.2.2. Beta-diversity of airway samples in adults

Microbial taxonomic composition is shown in Figure 4.5 considering the *phyla* and genera composition.

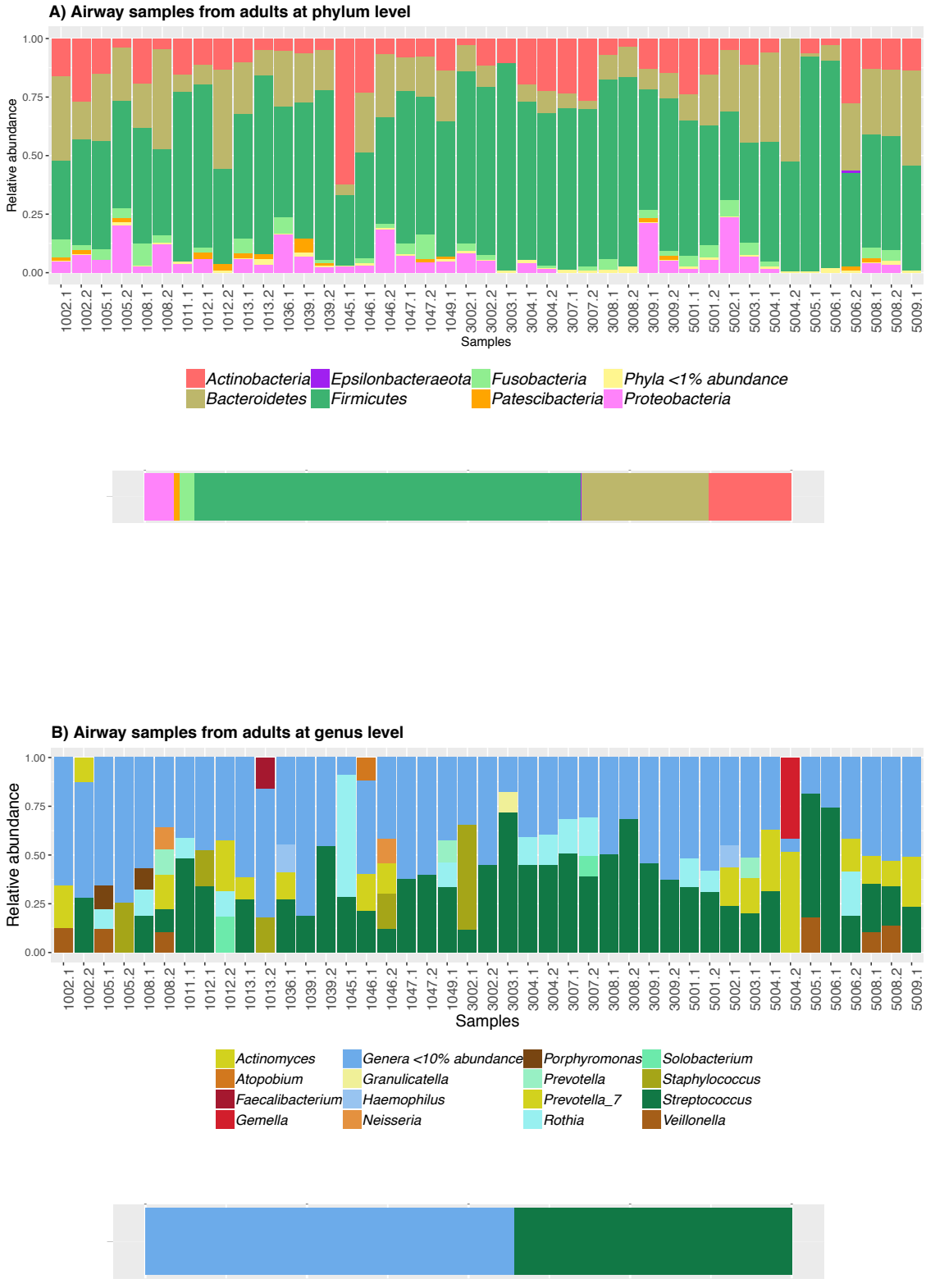

Figure 4.5. *Phyla* and genera distribution in sputum from adults. Vertical bars represent the relative abundance of main taxa for each sample whereas horizontal bars represent the median relative abundance of main taxa in the studied population.

Dominant *phyla* were *Firmicutes* (59.7%), *Bacteroidetes* (19.7%), *Actinobacteria* (12.7%), *Proteobacteria* (4.6%) and *Fusobacteria* (2.4%), together accounting for 99.1% of all sequences. The *Firmicutes/Bacteroidetes* ratio was 3.0 (Fig. 4.5A). At the genus level, only *Streptococcus* (42.9%) displayed an over 10% abundance, and the remaining 57.1% comprised for 58 genera. Among the most abundant genera (median relative abundance >1%), 53.8% were strict anaerobes. Lung microbiota composition in adults showed a high variability among the patients (Fig. 4.5B).

Interestingly, the unsupervised hierarchical cluster analysis of the most abundant genera (28 genera with median relative abundance among all samples over 0.1%) showed that samples were clustered into two separated groups (Fig. 4.6). One cluster grouped 25 samples (48.1%) and the other one grouped 22 samples (42.3%). The remaining 5 samples were considered outliers since they did not belong to any cluster. A different proportion of samples from patients with FEV_1 <75% was identified between both clusters (8 out of 25 or 32% *vs* 13 out of 22 or 59%, Chi-squared test p=0.06231) (Fig 4.6). Thus, the differences found in both clusters of samples could be attributed to lung function status. Compositional structure within these top-28 genera found in airway samples from adults was displayed in a heat map confirming that both clusters were distinguishable from one another based on their bacterial genera assemblage composition (Fig 4.6).

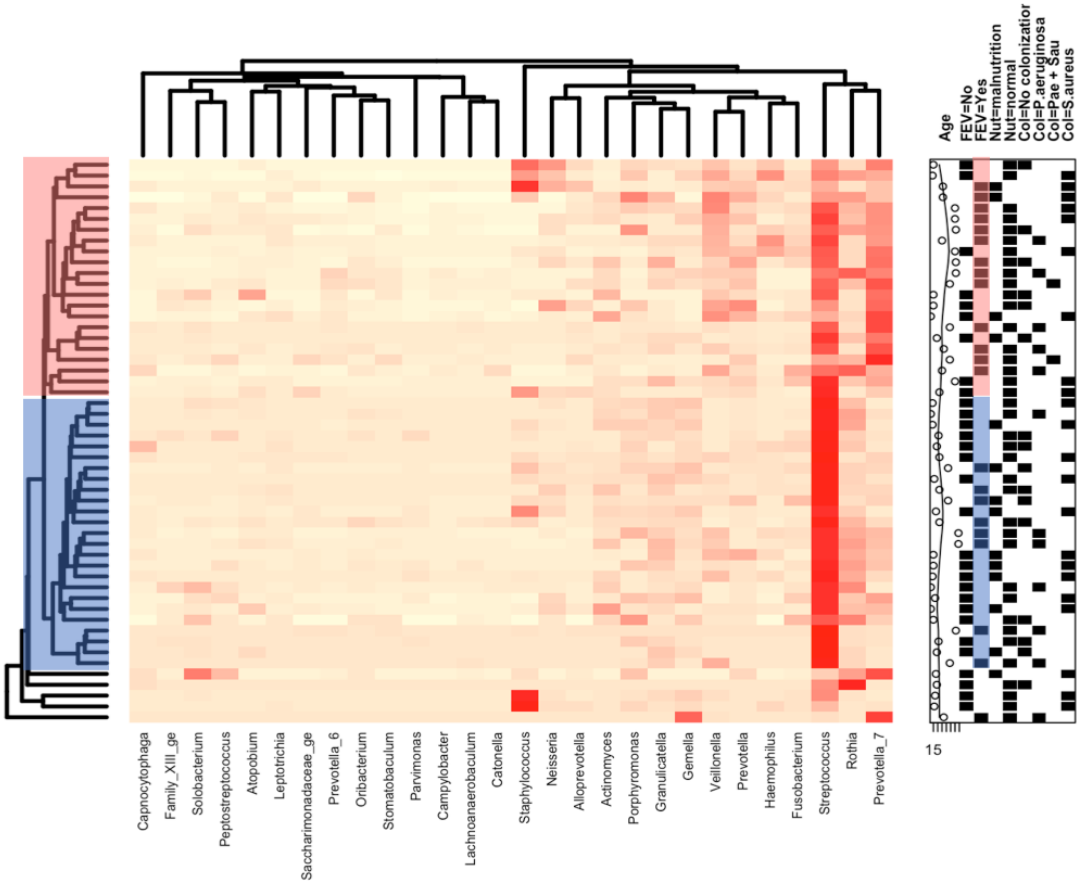

Figure 4.6. Heat map of the relative abundance and distribution for the most abundant genera found in sputum samples collected in adults. Heat map was constructed following a cluster analysis based on Bray-Curtis distances. The different color intensities represent the median relative abundance in each sample, genera are displayed in columns and rows represent individual samples. Two sample clusters were identified with blue and pink boxes.

The heat map showed that sputum samples clustered in the group dominated by strict anaerobes such as *Prevotella-7* and *Veillonella* were collected from a group of patients with lower lung function than patients clustered in the *Streptococcus* group (Fig 4.6). The main characteristics of the group with higher lung function were a median age of 19 years (range from 12 to 44), 36% of them have malnutrition, and finally 44.0% of them were colonized by *S. aureus* whereas 24.0% by *P. aeruginosa*. The contrary group, with a lower lung function, has a median age of 26.5 years (range from 12 to 41), only 18.2% of

them suffered from malnutrition, and lung colonization by *S. aureus* and *P. aeruginosa* occurred in 45.5% and 31.8% of samples (Fig 4.6).

The differences found between airways bacterial communities from both clusters of patients provided an opportunity to identify particular bacteria consistently associated with a worse respiratory condition. Using LEfSe, an all-against-all multiclass comparison was performed to identify clades specifically modulated by a low lung function. The results showed significant enrichments of *Bacteroidetes, Epsilonbacteraeota* and *Negativicutes phyla,* corresponding at the genus taxa with *Prevotella, Campylobacter* and *Veillonella,* respectively. On the other hand, samples clustered into the group with a higher lung function showed a significantly increase of the *Firmicutes/Bacteroidetes* ratio, with enrichment of *Firmicutes*, including the following genera: *Streptococcus*, *Granulicatella*, *Gemella* and *Parvimonas* (Fig. 4.7 and 4.8).

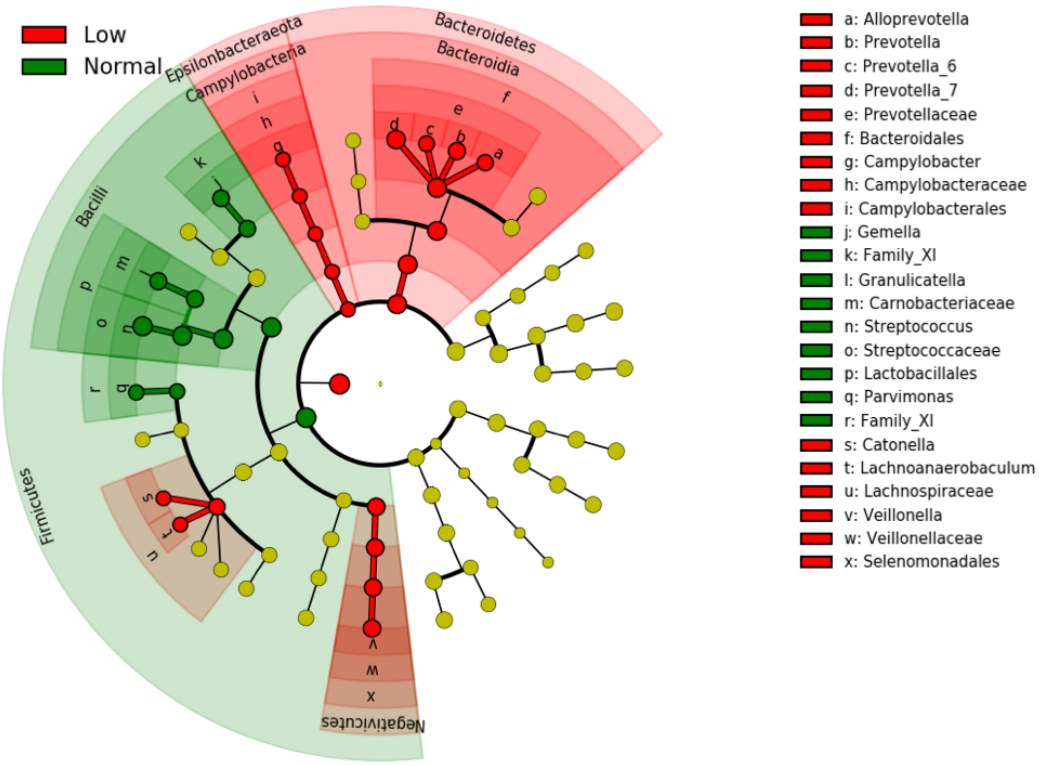

Figure 4.7. LEfSe cladogram of differentially abundant microbial clades in sputum collected from adults. Only 28 most abundant genera were analyzed. Samples were grouped according to the clusters identified in the hierarchical analysis. Outliers were removed for more clarity. Dark-yellow circles represent non-significant microbial clades.

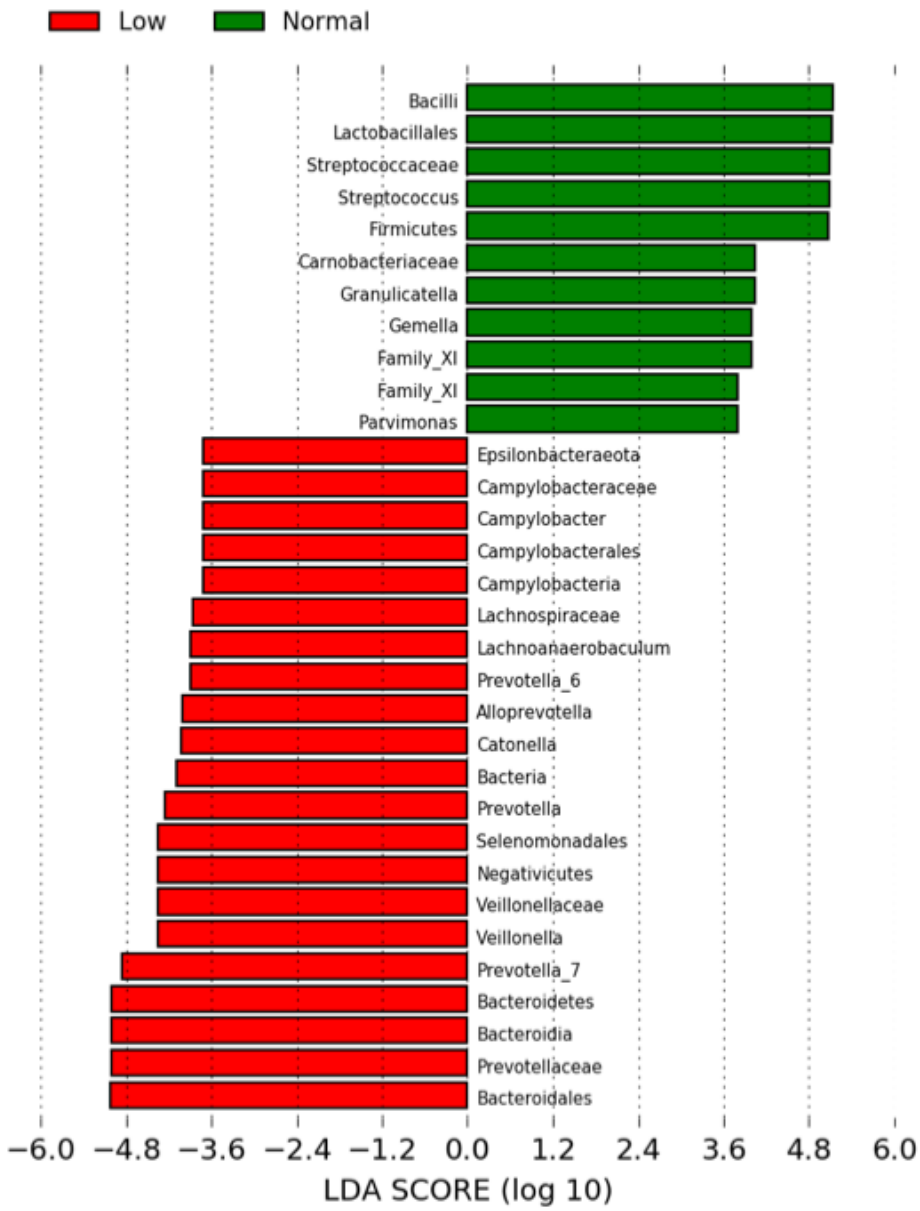

Figure 4.8. LEfSe bar graph showing LDA scores. Normal lung function-enriched taxa are indicated with a positive LDA score (green) and low lung function-enriched taxa have a negative score (red). Only taxa with a significant LDA threshold value >2 are shown.

4.5. Effect of DHA supplementation on CF microbiota

4.5.1. Effect of DHA on gut ecosystem

The influence of DHA supplementation in the gut microbiota composition was explored in 5 CF-patients, and compared with 7 patients receiving placebo. *Phyla* (Fig 4.9) and genera (Fig. 4.10) median relative abundances were compared between the basal and the last sample after 6 months of placebo or DHA consumption.

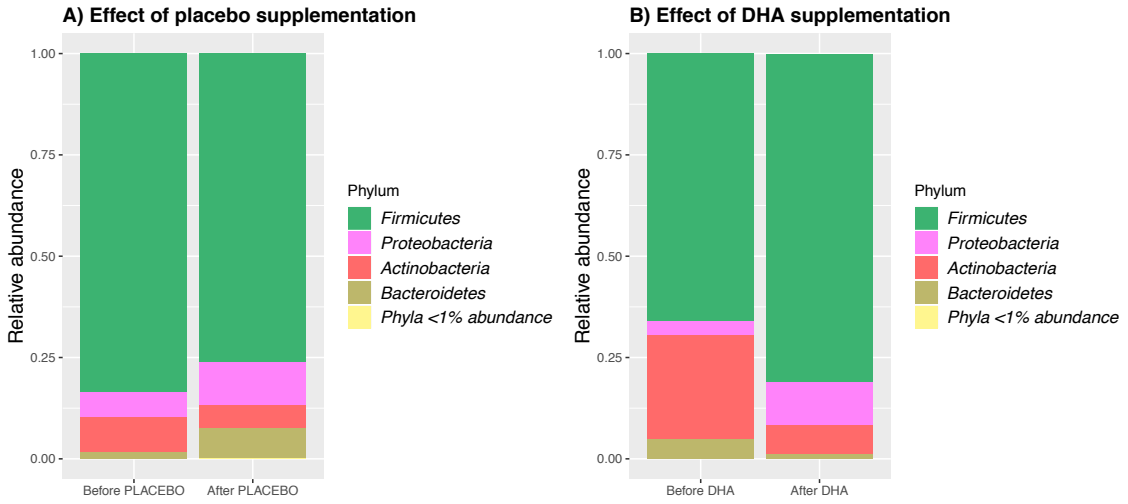

Figure 4.9. Effect of nutritional supplementation in *phyla* distribution in fecal samples: A) Placebo, B) DHA.

No significant differences, measured by Wilcoxon rank sum test with Benjamini-Hochberg correction, were found between samples collected before and after placebo or DHA supplementation. However, an increase in the *Firmicutes*/*Bacteroidetes* ratio from 13.2 to 67.4 was identified after DHA supplementation (Fig 4.9B).

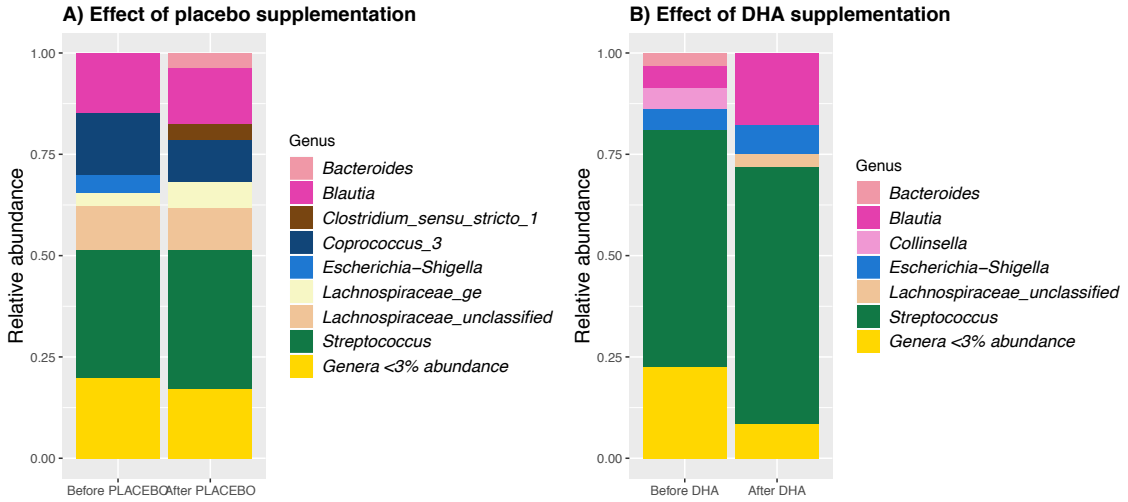

Figure 4.10. Effect of nutritional supplementation in genera distribution in fecal samples: A) Placebo, B) DHA.

At genus level, *Bacteroides* and *Collinsella* were only identified in the basal samples, whereas *Lachnospiraceae unclassified* was only identified after DHA supplementation. Among the shared genera, no significant differences at genus level were found between both groups of patients (Fig 4.10).

4.5.2. Effect of DHA on children airway microbiota

Microbial taxonomic compositions of sputum samples collected from children before and after receiving a placebo or a DHA supplemented diet are shown in Figures 4.11 and 4.12.

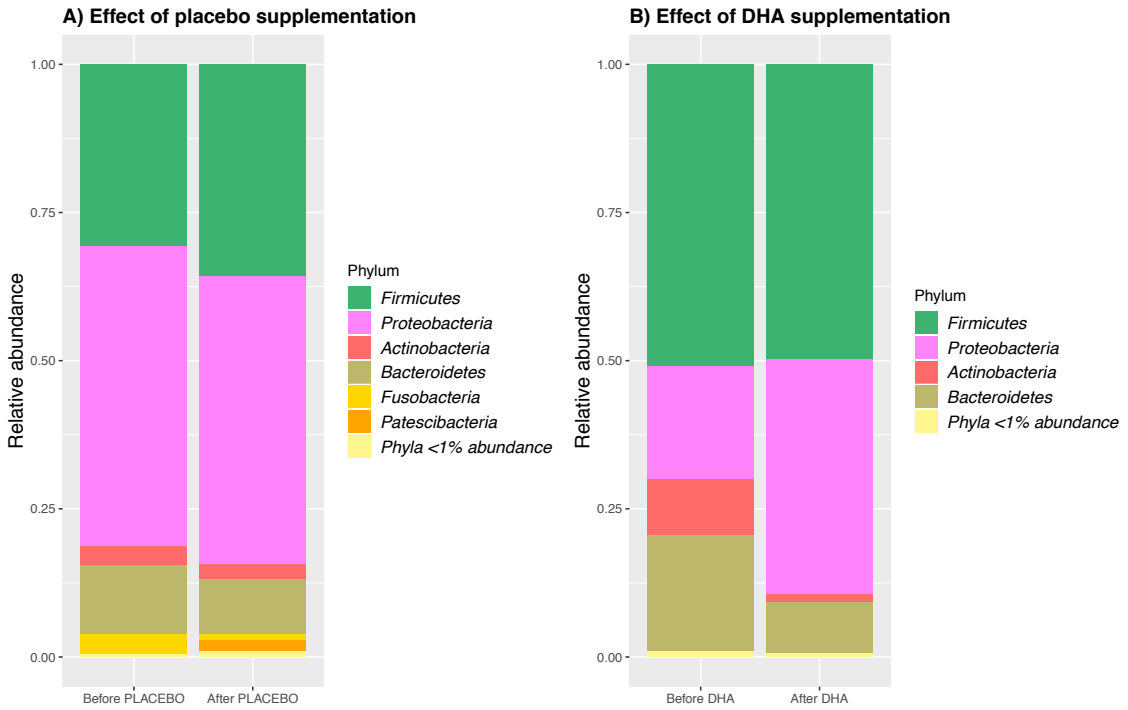

Figure 4.11. Effect of nutritional supplementation in *phyla* distribution in children airways: A) Placebo, B) DHA.

No significant differences were found between samples collected before and after placebo or DHA supplementation. However, an increase in the *Firmicutes*/*Bacteroidetes* ratio from 2.6 to 5.8 was identified after DHA supplementation due to a reduction in *Bacteroidetes* relative abundance (Fig 4.11B).

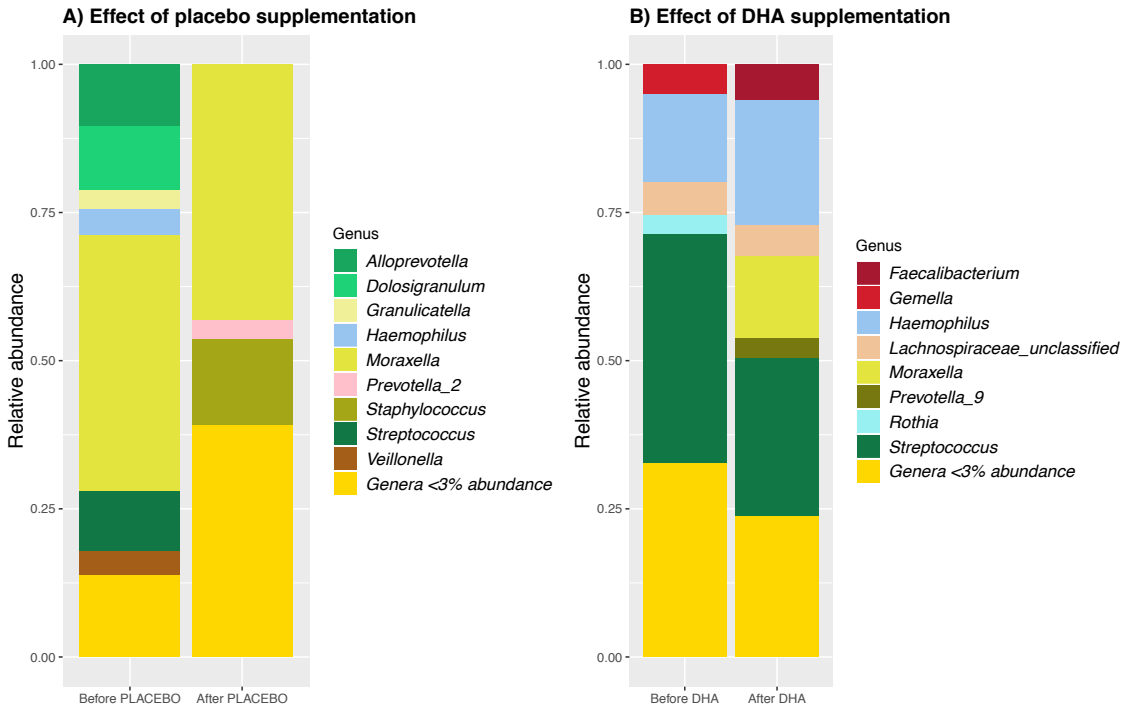

Figure 4.12. Effect of nutritional supplementation in *genera* distribution in children airways: A) Placebo, B) DHA.

At genus level, *Gemella* and *Rothia* were only identified in the airway microbiota before DHA supplementation, whereas *Faecalibacterium, Moraxella* and *Prevotella_9* were only identified after DHA supplementation. Among the remaining shared genera, no significant differences at genus level were found between both groups of samples (Fig 4.12). In this case, only 2 patients were supplemented with placebo. Due to the small sample size and the high individual variability was very difficult to make any meaningful conclusion. However, LEfSe analysis detected a significant association among *Selenomonas* and *Lachnospiraceae* with DHA supplementation.

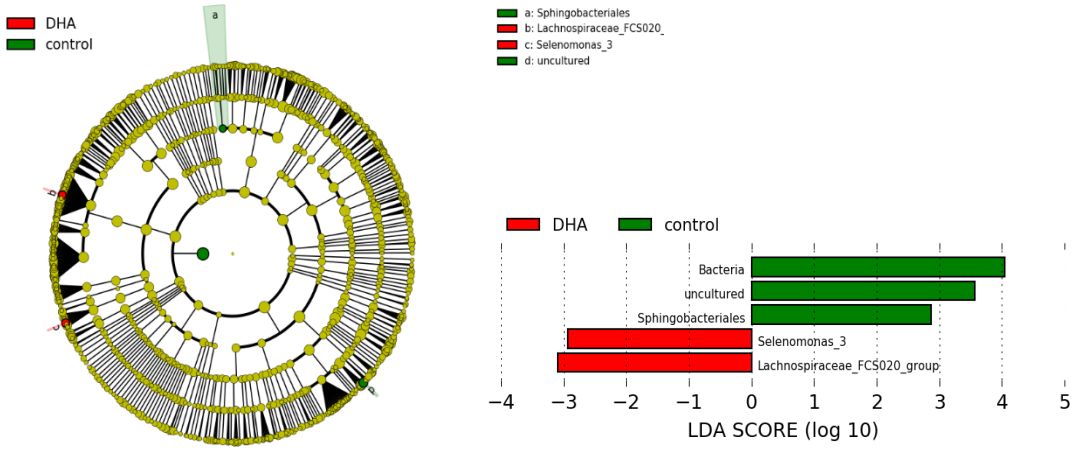

Figure 4.13. LEfSe analysis in children airways. LEfSe cladogram (left) and bar graph showing LDA scores (right).

4.5.3. Effect of DHA on adult airway microbiota

Microbial taxonomic compositions of sputum samples collected from adults before and after receiving a placebo or a DHA supplemented diet are shown in Figures 4.14 and 4.15.

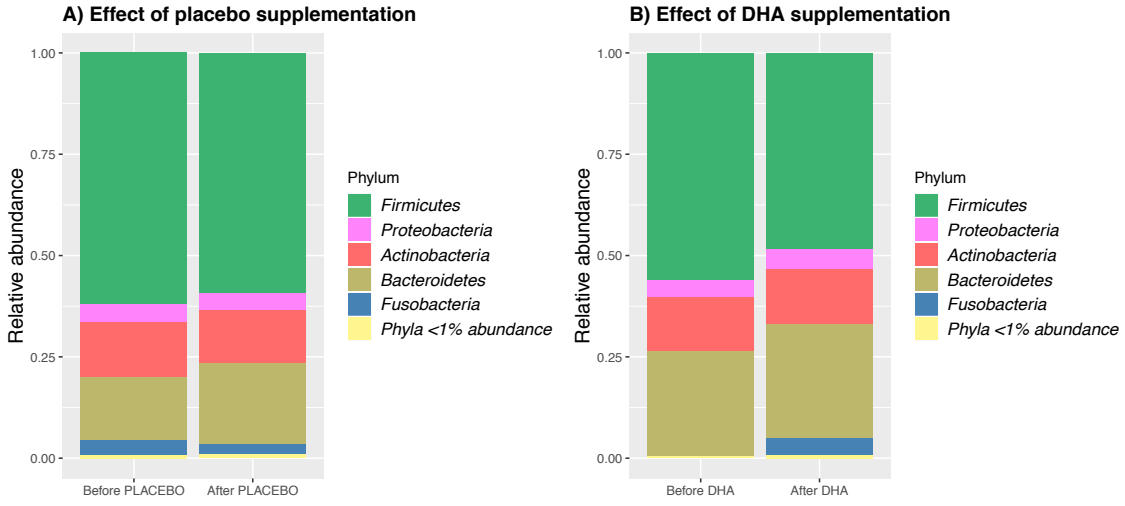

Figure 4.14. Effect of nutritional supplementation in *phyla* distribution in adults' airways: A) Placebo, B) DHA.

No significant differences were found at *phylum* level between samples collected before and after DHA supplementation, including the *Firmicutes*/*Bacteroidetes* ratio which changed from 2.2 to 1.7 after DHA supplementation (Fig 4.14). At genus level, *Gemella* and *Granulicatella* were only identified before DHA supplementation, whereas *Fusobacterium* and *Porphyromonas*, both of them strict anaerobes, were only identified after DHA supplementation. The proportion of genera with a relative abundance lower 3% increased from 9.5% to 18.2% after DHA supplementation. However, among the shared genera, no significant differences at genus level were found between both groups of samples (Fig 4.15).

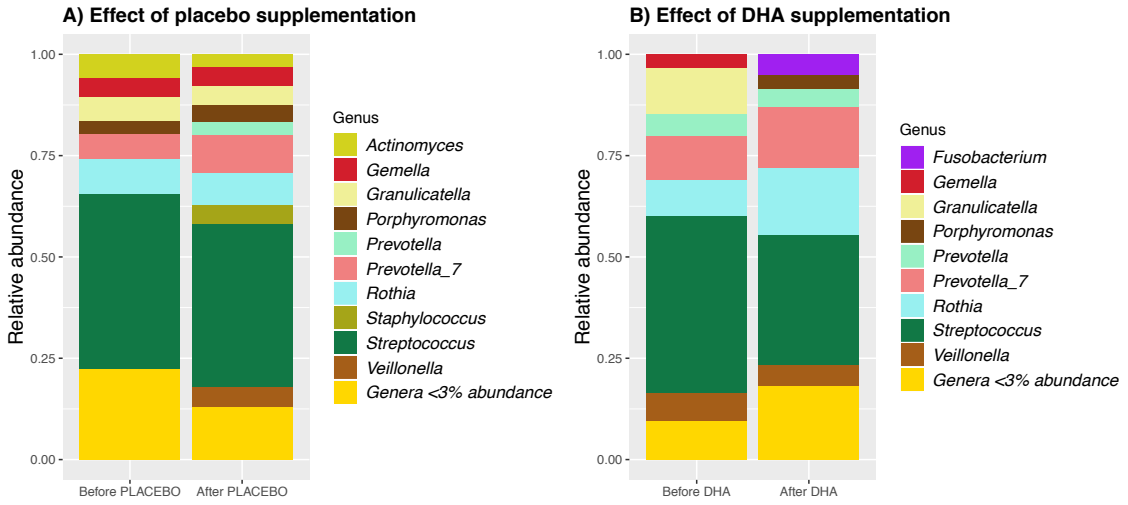

Figure 4.15. Effect of nutritional supplementation in *genera* distribution in adults' airways: A) Placebo, B) DHA.

4.6. Association between nutritional status and lung function with CF microbiota

This objective has been already accomplished earlier in the Results section. As a summary and according to adonis analysis, there was no effect of nutritional status and lung function on the composition of airway or fecal microbiota ($p>0.2$, $R2<0.01$ in both cases). In addition, significant differences in α-diversity and richness were not detected (Tables 4.1, 4.2 and 4.3). Only the unsupervised hierarchical cluster analysis of the top-28 most abundant airway genera in adults allowed us to detect a slightly association with lung function (Fig 4.6).

5. DISCUSSION

The gut microbiota of healthy adults is dominated by *Bacteroidetes*, *Firmicutes* and *Actinobacteria phyla*, being the predominant genera *Bacteroides*, *Eubacterium*, *Bifidobacterium*, *Peptostreptococcus*, *Ruminococcus*, *Clostridium* and *Propionibacteriun* whereas the *Enterobacteriaceae* family and the *Streptococcus*, *Enterococcus*, *Lactobacillus*, *Fusobacterium*, *Desulfovibrio* and *Methanobrevibacter* genera are minority populations (34, 35). CF is a genetic disease which ions interchange deficiency affects the mucosal surfaces producing high-density mucosal secretions. In consequence, CF microbiota is usually aberrant, with limited microbial diversity and higher *Firmicutes/Bacteroidetes* ratio that of healthy controls. The results obtained in the present work confirmed that CF gut microbiota presents a *Firmicutes* increase in detriment of *Bacteroidetes* compared to those reported for non-CF individuals (23). At the genus level, there was reported a decreased of the *Faecalibacterium, Roseburia* and *Bifidobacterium* genera density (36). Contrary to that, we found *Faecalibacterium* and *Bifidobacterium* among the most abundant gut microbiota genera of our patients. At this taxonomic level, it was also remarkable the high variability in microbiota structure within individual samples, that preclude the establishment of common rules for all patients.

Regarding the airway microbiota, the NGS tools allowed us to confirm the presence of cultivable CF-pathogens of the *Proteobacteria phylum* including *Pseudomonas*, *Haemophilus* and *Burkholderia*, along with an additional outgrowth of the *Actinobacteria phylum*. Furthermore, it is well established that airway microbiota is modified by age in CF patients (37), finding that was corroborated in this study. We also detected in adults a significant reduction of *Proteobacteria* population whereas no differences were found in the *Firmicutes/Bacteroidetes* ratio or anaerobes proportion. At genus level, *Streptococcus* was the most abundant genus in both, children and adults CF patients, followed by *Haemophilus* and *Gemella* in children and by *Rothia* and *Prevotella* in adults. Different studies of the CF microbiota have identified anaerobes as important bacteria in lung colonization/infection, reporting that some anaerobes are associated with CF exacerbations (18, 38). Despite these studies, the role of anaerobes in pathogenesis of CF lung disease remains poorly understood. Anaerobic growth reduces the pH of the extracellular environment, due to the excretion of acidic fermentation products, which promotes the growth of other fermentative bacteria, such as *Lactobacillus*, *Prevotella*,

and *Veillonella*, or facultatively fermentative bacteria, such as *Streptococcus*, *Rothia* and *Granulicatella* (39). Our results associated a group of strict anaerobes, such as *Prevotella*, *Veillonella* and *Lachnospiraceae,* with the CF patients with lowest lung function. Nevertheless, we are aware of our limitations in the lung function categorization (FEV1 >75 or ≤75%), although the use of a numerical variable could have helped to improve the sample clustering.

It has been reported a strong correlation among the nutritional status and the lung function (7–9), with influence also in the bacterial colonization. Surprisingly, significant association of lung function and the nutritional status or the colonization by pathogens such as *P. aeruginosa* or *S. aureus* was not detected among our data. The high inter-individual variability of the airway microbiota at the genus level detected in both children and adults, combined with the reduced number of patients studied; preclude the extraction of solid conclusions with our results. However, the most relevant result of the present work is the absence of a common CF microbiota pattern in both the airway and the gut compartments. These findings have been also confirmed by different authors, supporting the individual character of CF microbiota (40, 41). In fact, Renwick *et al.,* have proposed that the airway community in each CF patient is unique since very early in lifetime (41).

Concerning to the DHA supplementation, it is well known that diet is one of the strongest selective pressures for microbial communities within the gastrointestinal tract, although its relation with the lung microbiota remains unexplored. Despite the abundance of published data about the gut diet effect, omega-3 fatty acids influence is poorly understood. Recently, it has been reported a *Faecalibacterium* decrease, often associated with an increase in the *Lachnospiraceae* family, genera such as *Roseburia*, *Coprococcus* and *Bacteroides* (42–46). All of these studies highlighted the increased abundance of butyrate-producing bacterial genera after omega-3 fatty acids supplementation. The importance of butyrate, and short chain fatty acids in general, are linked to anti-inflammatory properties. In particular, butyrate is the preferred energy source for the colon epithelial cells, contributing to the maintenance of the gut barrier functions, and has immunomodulatory and anti-inflammatory properties (47). With regard to CF disease, omega-3 fatty acids also provide multiple benefits, decreasing the number of exacerbations and duration of antibiotic therapy and improving inflammatory and

anthropometric parameters in adults with CF (48, 49). To the best of our knowledge, this is the first study to describe the effect of omega-3 polyinsatured fatty acids, specifically DHA, on CF gastrointestinal and airway bacterial ecosystems. In this study, the DHA intervention revealed slightly differences at the genus level rather than the *phyla* level. At genus level, an increase in butyrate-producing bacteria belonging to the *Lachnospiraceae* family have been observed, confirming the findings previously mentioned. With regard to DHA effects on CF airway microbiota, *Lachnospiraceae* family along with an increase in the *Firmicutes/Bacteroidetes* ratio was detected in sputum from children after DHA supplementation. This ratio was reported to be a valuable marker in signaling human gut microbiota status. A decreased in this ratio has been related to weight loss (50) and advanced age (34). In adults, we found an increase in anaerobes associated with a decrease in facultative anaerobes in sputum collected from adults after 6 months of DHA diet supplementation. However, as far as our knowledge, this result is not supported by previous scientific literature and more analysis must be done to confirm it.

In summary, in this work we have applied the most recent sequencing and bioinformatics tools to characterize the gut and airway microbiota composition of CF-patients, detecting an aberrant composition of both ecosystems although with considerable individual particularities. The second objective of this work was to explore the effect of DHA supplementation in the bacterial content, and in this regard, some small changes have been detected, minimizing the interest of this fatty acid for the treatment of CF. However, it is important to note that the nutritional status of the patients at the beginning of the study was almost optimal, which also can dismiss the real effect of the DHA supplementation in a malnutrition scenario. Finally, no strong association was found between the clinical variables and the composition of the microbiota, which may be partly due to the high variation among patients, and the low number of subjects recruited.

6. CONCLUSIONS

1. The gut and airway CF microbiota is aberrant compared with that reported in non-CF subjects.
2. Both ecosystems show a significant inter and individual variability in their community structures.
3. The core CF-gut microbiota comprises *Lachnospiraceae* family and *Streptococcus* and *Blautia* genera.
4. Airways microbiota in children is dominated by *Streptococcus* and *Haemophilus* whereas in adults *Streptococcus* was the only genera over 10% abundance.
5. DHA supplementation provokes a slight increase of *Lachnospiraceae* family in both the gut and the airway microbiota of children.
6. There are not significant associations between the nutritional status or the lung function and the microbiota composition. Only considering the top-28 most abundant genera in adult CF-airways, we are able to detect an association between strict anaerobes, such as *Prevotella*, *Veillonella* and *Lachnospiraceae,* with CF patients with a low lung function.

7. REFERENCES

1. O'Sullivan BP, Freedman SD. 2009. Cystic fibrosis. Lancet 373:1891–1904.
2. Elborn JS. 2016. Cystic fibrosis. Lancet 388:2519–2531.
3. Wilschanski M, Novak I. 2013. The cystic fibrosis of exocrine pancreas. Cold Spring Harb Perspect Med 3:a009746.
4. Freedman SD, Blanco PG, Zaman MM, Shea JC, Ollero M, Hopper IK, Weed DA, Gelrud A, Regan MM, Laposata M, Alvarez JG, O'Sullivan BP. 2004. Association of cystic fibrosis with abnormalities in fatty acid metabolism. N Engl J Med 350:560–569.
5. Mimoun M, Coste TC, Lebacq J, Lebecque P, Wallemacq P, Leal T, Armand M. 2009. Increased tissue arachidonic acid and reduced linoleic acid in a mouse model of cystic fibrosis are reversed by supplemental glycerophospholipids enriched in docosahexaenoic acid. J Nutr 139:2358–2364.
6. Leggieri E, De Biase RV, Savi D, Zullo S, Halili I, Quattrucci S. 2013. Clinical effects of diet supplementation with DHA in pediatric patients suffering from cystic fibrosis. Minerva Pediatr 65:389–98.
7. Sanders DB, Fink AK. 2016. Background and epidemiology. Pediatr Clin North Am 63:567–84.
8. Konstan MW, Butler SM, Wohl MEB, Stoddard M, Matousek R, Wagener JS, Johnson CA, Morgan WJ, Investigators and coordinators of the epidemiologic study of cystic fibrosis. 2003. Growth and nutritional indexes in early life predict pulmonary function in cystic fibrosis. J Pediatr 142:624–630.
9. Lai HJ, Shoff SM, Farrell PM, Wisconsin Cystic fibrosis Neonatal Screening group WCNS. 2009. Recovery of birth weight z score within 2 years of diagnosis is positively associated with pulmonary status at 6 years of age in children with cystic fibrosis. Pediatrics 123:714–22.
10. Kerem E, Viviani L, Zolin A, Macneill S, Hatziagorou E, Ellemunter H, Drevinek P, Gulmans V, Krivec U, Olesen H, Kerem E. 2014. Factors associated with FEV_1 decline in cystic fibrosis: analysis of the ECFS Patient Registry on behalf of the ECFS Patient Registry Steering Group. Eur Respir J 43:125–133.
11. Lyczak JB, Cannon CL, Pier GB. 2002. Lung infections associated with cystic fibrosis. Clin Microbiol Rev 15:194–222.
12. Gilligan PH. 2014. Infections in patients with cystic fibrosis: diagnostic

microbiology update. Clin Lab Med 34:197–217.

13. Harrison F. 2007. Microbial ecology of the cystic fibrosis lung. Microbiology 153:917–923.
14. Tunney MM, Field TR, Moriarty TF, Patrick S, Doering G, Muhlebach MS, Wolfgang MC, Boucher R, Gilpin DF, McDowell A, Elborn JS. 2008. Detection of anaerobic bacteria in high numbers in sputum from patients with cystic fibrosis. Am J Respir Crit Care Med 177:995–1001.
15. Mirković B, Murray MA, Lavelle GM, Molloy K, Azim AA, Gunaratnam C, Healy F, Slattery D, McNally P, Hatch J, Wolfgang M, Tunney MM, Muhlebach MS, Devery R, Greene CM, McElvaney NG. 2015. The role of short-chain fatty acids, produced by anaerobic bacteria, in the cystic fibrosis airway. Am J Respir Crit Care Med 192:1314–24.
16. Conrad D, Haynes M, Salamon P, Rainey PB, Youle M, Rohwer F. 2013. Cystic fibrosis therapy: a community ecology perspective. Am J Respir Cell Mol Biol 48:150–156.
17. Goddard AF, Staudinger BJ, Dowd SE, Joshi-Datar A, Wolcott RD, Aitken ML, Fligner CL, Singh PK. 2012. Direct sampling of cystic fibrosis lungs indicates that DNA-based analyses of upper-airway specimens can misrepresent lung microbiota. Proc Natl Acad Sci U S A 109:13769–74.
18. Carmody LA, Caverly LJ, Foster BK, Rogers MAM, Kalikin LM, Simon RH, VanDevanter DR, LiPuma JJ. 2018. Fluctuations in airway bacterial communities associated with clinical states and disease stages in cystic fibrosis. PLoS One 13:e0194060.
19. Bacci G, Paganin P, Lopez L, Vanni C, Dalmastri C, Cantale C, Daddiego L, Perrotta G, Dolce D, Morelli P, Tuccio V, De Alessandri A, Fiscarelli EV, Taccetti G, Lucidi V, Bevivino A, Mengoni A. 2016. Pyrosequencing unveils cystic fibrosis lung microbiome differences associated with a severe lung function decline. PLoS One 11:e0156807.
20. Smith DJ, Badrick AC, Zakrzewski M, Krause L, Bell SC, Anderson GJ, Reid DW. 2014. Pyrosequencing reveals transient cystic fibrosis lung microbiome changes with intravenous antibiotics. Eur Respir J 44:922–930.
21. Zemanick ET, Wagner BD, Robertson CE, Ahrens RCC, Chmiel JF, Clancy JP, Gibson RL, Harris WT, Kurland G, Laguna TA, McColley SA, McCoy K, Retsch-

Bogart G, Sobush KT, Zeitlin PL, Stevens MJ, Accurso FJ, Sagel SD, Harris JK. 2017. Airway microbiota across age and disease spectrum in cystic fibrosis. Eur Respir J 50:1700832.

22. Carmody LA, Zhao J, Kalikin LM, LeBar W, Simon RH, Venkataraman A, Schmidt TM, Abdo Z, Schloss PD, LiPuma JJ. 2015. The daily dynamics of cystic fibrosis airway microbiota during clinical stability and at exacerbation. Microbiome 3:12.
23. Burke DG, Fouhy F, Harrison MJ, Rea MC, Cotter PD, O'Sullivan O, Stanton C, Hill C, Shanahan F, Plant BJ, Ross RP. 2017. The altered gut microbiota in adults with cystic fibrosis. BMC Microbiol 17:58.
24. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–41.
25. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl Environ Microbiol 79:5112–20.
26. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glöckner FO. 2014. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Res 42:D643–D648.
27. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:D590–D596.
28. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27:2194–200.
29. Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 73:5261–7.
30. McMurdie PJ, Holmes S. 2013. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8:e61217.
31. Ploner A. 2015. Heatplus: Heatmaps with row and/or column covariates and

colored clusters. R package version 2.24.0.

32. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, Dan McGlinn PR, Minchin, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H. 2018. vegan: community ecology package. R package version 2.5-2.
33. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C. 2011. Metagenomic biomarker discovery and explanation. Genome Biol 12:R60.
34. Mariat D, Firmesse O, Levenez F, Guimarães V, Sokol H, Doré J, Corthier G, Furet J-P. 2009. The *Firmicutes/Bacteroidetes* ratio of the human microbiota changes with age. BMC Microbiol 9:123.
35. Harmsen HJM, Raangs GC, He T, Degener JE, Welling GW. 2002. Extensive set of 16S rRNA-based probes for detection of bacteria in human feces. Appl Environ Microbiol 68:2982–90.
36. Flint HJ, Duncan SH, Scott KP, Louis P. 2007. Interactions and competition within the microbial community of the human colon: links between diet and health. Environ Microbiol 9:1101–1111.
37. Salsgiver EL, Fink AK, Knapp EA, LiPuma JJ, Olivier KN, Marshall BC, Saiman L. 2016. Changing epidemiology of the respiratory bacteriology of patients with cystic fibrosis. Chest 149:390–400.
38. Worlitzsch D, Rintelen C, Böhm K, Wollschläger B, Merkel N, Borneff-Lipp M, Döring G. 2009. Antibiotic-resistant obligate anaerobes during exacerbations of cystic fibrosis patients. Clin Microbiol Infect 15:454–460.
39. McLean JS, Fansler SJ, Majors PD, McAteer K, Allen LZ, Shirtliff ME, Lux R, Shi W. 2012. Identifying low pH active and lactate-utilizing taxa within oral microbiome communities from healthy children using stable isotope probing techniques. PLoS One 7:e32219.
40. Stressmann FA, Rogers GB, van der Gast CJ, Marsh P, Vermeer LS, Carroll MP, Hoffman L, Daniels TWV, Patel N, Forbes B, Bruce KD. 2012. Long-term cultivation-independent microbial diversity analysis demonstrates that bacterial communities infecting the adult cystic fibrosis lung show stability and resilience. Thorax 67:867–873.
41. Renwick J, McNally P, John B, DeSantis T, Linnane B, Murphy P, Shield CF. 2014. The microbial community of the cystic fibrosis airway is disrupted in early

life. PLoS One 9:e109798.

42. Watson H, Mitra S, Croden FC, Taylor M, Wood HM, Perry SL, Spencer JA, Quirke P, Toogood GJ, Lawton CL, Dye L, Loadman PM, Hull MA. 2017. A randomised trial of the effect of omega-3 polyunsaturated fatty acid supplements on the human intestinal microbiota. Gut gutjnl-2017-314968.
43. Pu S, Khazanehei H, Jones PJ, Khafipour E. 2016. Interactions between obesity status and dietary intake of monounsaturated and polyunsaturated oils on human gut microbiome profiles in the canola Oil Multicenter Intervention Trial (COMIT). Front Microbiol 7:1612.
44. Noriega BS, Sanchez-Gonzalez MA, Salyakina D, Coffman J. 2016. Understanding the impact of omega-3 rich diet on the gut microbiota. Case Rep Med 2016:1–6.
45. Menni C, Zierer J, Pallister T, Jackson MA, Long T, Mohney RP, Steves CJ, Spector TD, Valdes AM. 2017. Omega-3 fatty acids correlate with gut microbiome diversity and production of N-carbamylglutamate in middle aged and elderly women. Sci Rep 7:11079.
46. Younge N, Yang Q, Seed PC. 2017. Enteral high fat-polyunsaturated fatty acid blend alters the pathogen composition of the intestinal microbiome in premature infants with an enterostomy. J Pediatr 181:93–101.e6.
47. Rivière A, Selak M, Lantin D, Leroy F, De Vuyst L. 2016. *Bifidobacteria* and butyrate-producing colon bacteria: importance and strategies for their stimulation in the human gut. Front Microbiol 7:979.
48. Olveira G, Olveira C, Acosta E, Espíldora F, Garrido-Sánchez L, García-Escobar E, Rojo-Martínez G, Gonzalo M, Soriguer F. 2010. La suplementación con ácidos grasos mejora parámetros respiratorios, inflamatorios y nutricionales en adultos con fibrosis quística. Arch Bronconeumol 46:70–77.
49. Hanssens L, Thiébaut I, Lefèvre N, Malfroot A, Knoop C, Duchateau J, Casimir G. 2016. The clinical benefits of long-term supplementation with omega-3 fatty acids in cystic fibrosis patients - A pilot study. Prostaglandins Leukot Essent Fatty Acids 108:45–50.
50. Ley RE, Turnbaugh PJ, Klein S, Gordon JI. 2006. Microbial ecology: Human gut microbes associated with obesity. Nature 444:1022–1023.
51. Renwick J, McNally P, John B, DeSantis T, Linnane B, Murphy P, Shield CF.

2014. The microbial community of the cystic fibrosis airway is disrupted in early life. PLoS One 19;9(12):e109798.

52. de Freitas MB, Moreira EAM, Tomio C, Moreno YMF, Daltoe FP, Barbosa E, Ludwig Neto N, Buccigrossi V, Guarino A. 2018. Altered intestinal microbiota composition, antibiotic therapy and intestinal inflammation in children and adolescents with cystic fibrosis. PLoS One 13(6):e0198457.

8. Appendix: R pipeline

```
#####################################################################
                        PACKAGES INSTALLATION
#####################################################################
#Remove all from environment.
rm(list=ls())
ls()

#Set the directory.
setwd("/Users/luzbalsalobre/desktop/Project_TFM")
getwd()

#Install the following packages:
library(tidyr)
library(phyloseq)
library(vegan)
library(ggplot2)
library(DESeq2)
library(DESeq)
require(doBy)
require(RColorBrewer)
library(gdata)
source("http://bioconductor.org/biocLite.R")
biocLite("Heatplus")

#####################################################################
                          Read OTU Data
#####################################################################
# Since we are using 16S amplicon design it will be possible to
taxonomically  classify down to the genus level.
AvailableRanks<-
c("Kingdom","Phylum","Class","Order","Family","Genus")

# Define the minimum number of OTU counts for a sample to be
further analyzed.
minSampleCountsB<-1000

# Read the metadata.
Metadata <- read.csv("~/Desktop/Project_TFM/Data/Metadata.csv",
header=TRUE, sep=";")
View(Metadata)

# Let's inspect the metadata.
dim(Metadata) #96 samples y 13 variables

# We will do some processing of some variables to facilitate
further analysis.
rownames(Metadata)<-Metadata$SampleID
Metadata$Time<-as.factor(Metadata$Time)
Metadata$Agecat[Metadata$Age<=10] <- "child"
Metadata$Agecat[Metadata$Age>10] <- "adult"
Metadata$Agecat <- factor(Metadata$Agecat)
Metadata$FEV <- factor(Metadata$FEV)
```

```
#I need these percentiles values to categorize IMC values in
children.
IMC_P10 <- quantile(Metadata$IMC,0.10)
IMC_P85 <- quantile(Metadata$IMC,0.85)

Metadata$IMCcat[Metadata$Age>=18 & Metadata$IMC<=18.5] <-
"malnutrition"
Metadata$IMCcat[Metadata$Age>=18 & Metadata$IMC>18.5] <- "normal"
Metadata$IMCcat[Metadata$Age<18 & Metadata$IMC>IMC_P10] <- "normal"
Metadata$IMCcat[Metadata$Age<18 & Metadata$IMC<IMC_P85] <- "normal"
Metadata$IMCcat[Metadata$Age<18 & Metadata$IMC<IMC_P10] <-
"malnutrition"
Metadata$IMCcat[Metadata$Age<18 & Metadata$IMC>IMC_P85] <-
"malnutrition"
Metadata$IMCcat <- factor(Metadata$IMCcat)
Metadata$Cluster[Metadata$Agecat=="adult" &
Metadata$Sample=="stool"] <- "stool_adult"
Metadata$Cluster[Metadata$Agecat=="adult" &
Metadata$Sample=="sputum"] <- "sputum_adult"
Metadata$Cluster[Metadata$Agecat=="child" &
Metadata$Sample=="sputum"] <- "sputum_child"

Metadata$Colonization[Metadata$P_aeruginosa=="Yes"&
Metadata$S_aureus=="No"] <- "P.aeruginosa"
Metadata$Colonization[Metadata$P_aeruginosa=="No" &
Metadata$S_aureus=="Yes"] <- "S.aureus"
Metadata$Colonization[Metadata$P_aeruginosa=="No" &
Metadata$S_aureus=="Yes"] <- "S.aureus"
Metadata$Colonization[Metadata$P_aeruginosa=="Yes" &
Metadata$S_aureus=="Yes"] <- "Pae + Sau"
Metadata$Colonization[Metadata$P_aeruginosa=="No" &
Metadata$S_aureus=="No"] <- "No colonization"
Metadata$Colonization <- factor(Metadata$Colonization)

# Now we will import data derived from Mothur. For this, the
minimal information is an OTU table and a taxonomical
classification for each of the OTUS. In addition we can also import
a phylogenetic tree containing evolutionary information between all
OTUs. We will use a cut-off of 97% similarity to define an OTU in
agreement with parameters used for Mothur pipeline.

xTFM<-
import_mothur(mothur_constaxonomy_file="~/Desktop/Project_TFM/Data/
Fibrosis.0.03.cons.taxonomy",

mothur_shared_file="~/Desktop/Project_TFM/Data/Fibrosis.shared",
                    cutoff=0.03)

Otus <- otu_table(xTFM)
dim(Otus) #1784 Otus y 96 samples

tax <- tax_table(xTFM)
dim(Otus) #1784 Otus y 96 samples

xTree<-
import_mothur(mothur_tree_file="~/Desktop/Project_TFM/Data/Fibrosis
.rep.phylip.tre")
```

```
# To change ID labels in the tree and then merge the data.
taxa_names(xTree) <- taxa_names(xTFM)
xTFM <- merge_phyloseq(xTFM, xTree)
xTFM

# Attach the metadata.
sample_data(xTFM)<-Metadata

# Define available ranks in taxonomy file.
colnames(tax_table(xTFM))<-AvailableRanks

# Some data before filtering.
ordered(colSums(otu_table(xTFM))) # To see the minimum and maximun
reads before filtering
sum(colSums(otu_table(xTFM))) # 4,959,311 is the Total number of
reads
mean(colSums(otu_table(xTFM))) # 51,650 is the Average reads among
samples
sd(colSums(otu_table(xTFM))) #46,192 is the SD reads reads among
samples

# Filter out samples that are below 1000 counts (minSampleCountsB)
xTFM<-prune_samples(sample_sums(xTFM)>minSampleCountsB,xTFM)
xTFM #I removed 2 samples with < 1000 reads/counts

# Let's plot the number of counts for the remaining 94 samples to a
pdf file
pdf("CoverageBarplots.pdf",paper="A4r")
barplot(colSums(otu_table(xTFM)),las=2,cex.names=0.6,main="#Counts/
Sample")
dev.off()

#To count taxa per taxonomic rank
length(get_taxa_unique(xTFM, taxonomic.rank = "Genus"))#505
length(get_taxa_unique(xTFM, taxonomic.rank = "Family"))#214
length(get_taxa_unique(xTFM, taxonomic.rank = "Order"))#124
length(get_taxa_unique(xTFM, taxonomic.rank = "Class"))#52
length(get_taxa_unique(xTFM, taxonomic.rank = "Phylum"))#25
length(get_taxa_unique(xTFM, taxonomic.rank = "Kingdom"))#1

######################################################################
                    Start Ordination Analysis
######################################################################
# For this we need to use ecological distance. We will be using
Bray-Curtis distance
# To put our data aside
x.4.0<-xTFM
x.4.0<-tax_glom(x.4.0,taxrank="Genus")

# We transform the data to proportion abundances
x.4.0 = transform_sample_counts(x.4.0, function(x) ((x/sum(x))))

# We will be using NMDS which used an initial random seed for
iteration. Since we want to be able to repeat the same analysis we
fix a seed for analysis
set.seed(12345)
```

```
# NMDS Ordination with Bray-Curtis distance
x.4.0.ord<-ordinate(x.4.0,"NMDS",distance="bray",trymax=200)
capture.output(file="NMDS_Bray_proportions_ordinfo.txt",x.4.0.ord)
pdf("NMDS_Bray_proportions_stressplot.pdf")
stressplot(x.4.0.ord)
dev.off()

# Simple Plot
p.4.0.samples=plot_ordination(x.4.0,x.4.0.ord)
p.4.0.samples

# Let's map some metadata with coloured and ellipses
p.4.0.samples=plot_ordination(x.4.0,x.4.0.ord)

p.4.0.samples  + geom_point(size = 2.5,aes(color=Cluster))+
theme_bw()+
theme(panel.border=element_blank(),panel.grid.major=element_blank()
,panel.grid.minor=element_blank(),axis.line=element_line(colour="bl
ack"),axis.text.x=element_text(size=16),axis.text.y=element_text(si
ze=16))+
theme(axis.title.x=element_text(size=14),axis.title.y=element_text(
size=14))+
theme(legend.text=element_text(size=12,face="italic"),legend.title=
element_text(size=12))+
theme(plot.title=element_text(lineheight=1,face="bold",size=19))
ggsave("NMDS_Bray_Proportion_Colour_withClusters_ByCluster.pdf")

#####################################################################
                          Adonis TESTs
#####################################################################
x.6.0<-xTFM
x.6.0.genus<-tax_glom(x.6.0,taxrank="Genus")
x.6.0<-x.6.0.genus
x.6.0 = transform_sample_counts(x.6.0, function(x) (((1*
(x))/sum(x))))

# Non-parametric analysis of variance using NPMANOVA through adonis
function {vegan}package for full dataset (all genera and all
samples)
x.6.0.response<-t(otu_table(x.6.0))
metadata<-data.frame(sample_data(x.6.0))
x.6.0.explanatory<-
metadata[,c("Sample","Agecat","Group","FEV","IMCcat","P_aeruginosa"
,"S_aureus")]

for(var in colnames(x.6.0.explanatory)){
  explanatory=data.frame(x.6.0.explanatory[,eval(var)])
  myadonis<-adonis(x.6.0.response~.,data=explanatory)
  capture.output(paste("Adonis on:
",var),file="adonis_tests.txt",append=T)
  capture.output(myadonis,file="adonis_tests.txt",append=T)
}

adonis(x.6.0.response~.,data=x.6.0.explanatory)
x.6.0.adonis<-adonis(x.6.0.response~.,data=x.6.0.explanatory)
simper(x.6.0.response,x.6.0.explanatory[,"Sample"])
```

```
simper(x.6.0.response,x.6.0.explanatory[,"Agecat"])


#####################################################################
         Start Alpha - Diversity Analysis in STOOL samples
#####################################################################
# We will characterize alpha-diversity indices using a rarefied
subset of 5000 counts in order to balance sampling between samples.
We randomly subset 5000 counts from each sample and discard all
samples with fewer counts.

x.2.0 <- xTFM
set.seed(1234)
x.2.0 <- rarefy_even_depth(x.2.0, 5000)  ##### Rarefy to 5000
counts
x.2.0 #90 samples and 1164 taxa
x.2.0.stool <- subset_samples(x.2.0, Sample=="stool")

# Generate Diversity Stats for STOOL samples and capture them in a
txt file
stool.sampleData<-data.frame(sample_data(x.2.0.stool))
summary(stool.sampleData)

er.x.2.0.stool<-estimate_richness(x.2.0.stool)
rownames(er.x.2.0.stool)<-rownames(stool.sampleData)

er.x.2.0.stool<-
merge(stool.sampleData,er.x.2.0.stool,by="row.names",all.x=T)
rownames(er.x.2.0.stool)<-rownames(stool.sampleData)
sample_data(x.2.0.stool)<-er.x.2.0.stool

capture.output(file="DivAndRichnessStoolStats.txt",paste())
for(covar in c("Sex","FEV","IMCcat","Group")){
  for (measure in c("Chao1","Shannon")){

capture.output(file="DivAndRichnessStoolStats.txt",paste("#########
###",measure,"by",covar),append=T)
    if(length(unique(er.x.2.0.stool[,covar]))==2){
      myttest<-
t.test(er.x.2.0.stool[,measure]~er.x.2.0.stool[,covar])
capture.output(file="DivAndRichnessStoolStats.txt",myttest,append=T
)}else{myanova<-
aov(er.x.2.0.stool[,measure]~er.x.2.0.stool[,covar])
capture.output(file="DivAndRichnessStoolStats.txt",myanova,append=T
)capture.output(file="DivAndRichnessStoolStats.txt",
summary(myanova),append=T)
capture.output(file="DivAndRichnessStoolStats.txt",TukeyHSD(myanova
),append=T)}}}

# To calculate SD values for STOOL samples
stool.Female <- subset(er.x.2.0.stool,
er.x.2.0.stool$Sex=="Female")
sd(stool.Female$Chao1)#30.044
sd(stool.Female$Shannon)#0.964

stool.Male <- subset(er.x.2.0.stool, er.x.2.0.stool$Sex=="Male")
sd(stool.Male$Chao1)#11.644
```

```
sd(stool.Male$Shannon)#0.231

stool.FEV.Yes <- subset(er.x.2.0.stool, er.x.2.0.stool$FEV=="Yes")
sd(stool.FEV.Yes$Chao1)#20.514
sd(stool.FEV.Yes$Shannon)#0.775

stool.FEV.No <- subset(er.x.2.0.stool, er.x.2.0.stool$FEV=="No")
sd(stool.FEV.No$Chao1)#23.563
sd(stool.FEV.No$Shannon)#0.575

stool.IMC.Normal <- subset(er.x.2.0.stool,
er.x.2.0.stool$IMCcat=="normal")
sd(stool.IMC.Normal$Chao1)#22.473
sd(stool.IMC.Normal$Shannon)#0.719

stool.IMC.Malnutrition <- subset(er.x.2.0.stool,
er.x.2.0.stool$IMCcat=="malnutrition")
sd(stool.IMC.Malnutrition$Chao1)#2.474
sd(stool.IMC.Malnutrition$Shannon)#0.057

stool.Group.DHA <- subset(er.x.2.0.stool,
er.x.2.0.stool$Group=="DHA")
sd(stool.Group.DHA$Chao1)#24.850
sd(stool.Group.DHA$Shannon)#1.041

stool.Group.control <- subset(er.x.2.0.stool,
er.x.2.0.stool$Group=="control")
sd(stool.Group.control$Chao1)#16.002
sd(stool.Group.control$Shannon)#0.368

# Analysis of sputa samples was made in a similar way

######################################################################
   Cumulative Stacked Barplots at PHYLUM LEVEL for STOOL samples
######################################################################
# Description of CF microbiota at basal status: I will use patients
Group=Control(that means, placebo patients and Patients treated
with DHA at time 1)
x.9.0<-xTFM
x.9.0.phylum<-tax_glom(x.9.0,taxrank="Phylum")
x.9.0.phylum.stool <- subset_samples(x.9.0.phylum, Sample=="stool")
x.9.0.phylum.stool.control <- subset_samples(x.9.0.phylum.stool,
Group=="control")
ps.melt.x.9.0.phylum.stool.control <-
psmelt(x.9.0.phylum.stool.control)

# Calculate relative abundance
ps.melt.x.9.0.phylum.stool.control$AbundanceProportion <-
ave(ps.melt.x.9.0.phylum.stool.control$Abundance,list(ps.melt.x.9.0
.phylum.stool.control[,"SampleID"]), FUN=function(L) L/sum(L))

# Convert to character
ps.melt.x.9.0.phylum.stool.control$Phylum<-
as.character(ps.melt.x.9.0.phylum.stool.control$Phylum)
View(ps.melt.x.9.0.phylum.stool.control)
```

```
# Rename genera with less than 1% abundance
ps.melt.x.9.0.phylum.stool.control$Phylum[ps.melt.x.9.0.phylum.stoo
l.control$AbundanceProportion<0.01]<-"Phyla <1% abundance"

# We also take care of colors
phylum.stool.colors <-
c("indianred1","darkkhaki","mediumseagreen","lightgreen",
                          "khaki1","orchid1","royalblue")

# Plot the figure
p<-
ggplot(ps.melt.x.9.0.phylum.stool.control,aes(x=sample_ID,y=Abundan
ceProportion))
p+geom_bar(stat="identity",aes(fill=Phylum))+
 theme(axis.text.x = element_text(size=12,angle=90,hjust=1)) +
theme(axis.title.x=element_text(size=14),axis.title.y=element_text(
size=14))+
theme(legend.text=element_text(size=14,face="italic"),legend.title=
element_blank())+
  theme(plot.title=element_text(face="bold",size=14))+
  labs(title="A) Fecal samples at phylum Level")+
  scale_fill_manual(values=phylum.stool.colors)+
  labs(x="Samples",y="Relative abundance")+
  theme(legend.position="bottom")
ggsave("stool.phylum.barplot.pdf", height = 5, width = 11)

######################################################################
     Summary Stacked Barplot at PHYLUM LEVEL for STOOL samples
######################################################################
# To obtain a barplot with the median proportions for stool samples
at phylum level
x.9.0<-xTFM
x.9.0.phylum<-tax_glom(x.9.0,taxrank="Phylum")
x.9.0.phylum.stool <- subset_samples(x.9.0.phylum, Sample=="stool")
x.9.0.phylum.stool.control <- subset_samples(x.9.0.phylum.stool,
Group=="control")

x.9.0.phylum.stool.control.Abundance <-
data.frame(otu_table(x.9.0.phylum.stool.control))
x.9.0.phylum.stool.control.Genera <-
data.frame(tax_table((x.9.0.phylum.stool.control)))
phylum.stool <- cbind(x.9.0.phylum.stool.control.Abundance,
x.9.0.phylum.stool.control.Genera)
dim(phylum.stool)#25 phyla and 22 variables (16 samples + 6 rank
taxa)

phylum.stool <- phylum.stool[,c(18,1:16)]
rownames(phylum.stool) <- phylum.stool$Phylum
phylum.stool <- phylum.stool[,-1]
phylum.stool <- t(phylum.stool)
dim(phylum.stool)#16 samples and 25 phyla

#To transform raw counts to relative abundance within a sample:
phylum.stool.rel <- phylum.stool/rowSums(phylum.stool)
```

```
# To determine the median relative abundance for each
column/phylum.
phylum.stool.medianab <- apply(phylum.stool.rel,2,median)

#To prepare the dataset
data.phylum.stool <- as.data.frame(phylum.stool.medianab)
data.phylum.stool <- cbind(data.phylum.stool,
row.names(data.phylum.stool))
colnames(data.phylum.stool) <- c("median.abundance","Phylum")
data.phylum.stool$median.proportion <-
data.phylum.stool$median.abundance/sum(data.phylum.stool$median.abu
ndance)

#To remove rows without counts
data.phylum.stool <- data.phylum.stool[apply(data.phylum.stool!=0,
1, all),]
dim(data.phylum.stool)#7 rows
data.phylum.stool$sample <- c(rep("stool",7))

#Take care of colors
phylum.stool.colors <-
c("indianred1","darkkhaki","mediumseagreen","lightgreen",
                          "orange","orchid1","royalblue")
#Get the figure
p <- ggplot(data.phylum.stool, aes(x=sample, y=median.proportion))
p + geom_bar(stat="identity", aes(fill=Phylum))+
  scale_fill_manual(values=phylum.stool.colors)+
  theme(legend.position = "none",
        axis.title.x = element_blank(),
        axis.title.y = element_blank(),
        axis.text.x = element_blank(),
        axis.text.y = element_blank())
ggsave("Phylum.stool.median.proportions.pdf")

######################################################################
    Cumulative Stacked Barplots at GENUS LEVEL for STOOL samples
######################################################################
x.9.1<-xTFM

# We apply a more stringent filter, basically we are not interested
in rare OTUs but on general main trends of taxonomical composition
wh0=genefilter_sample(x.9.1,filterfun_sample(function(x) x>5),
A=0.1*nsamples(x.9.1))
x.9.1<-prune_taxa(wh0,x.9.1)
x.9.1.genus<-tax_glom(x.9.1,taxrank="Genus")
x.9.1.genus.stool <- subset_samples(x.9.1.genus, Sample=="stool")
x.9.1.genus.stool.control <- subset_samples(x.9.1.genus.stool,
Group=="control")
ps.melt.x.9.1.genus.stool.control <-
psmelt(x.9.1.genus.stool.control)

# Calculate relative abundance
ps.melt.x.9.1.genus.stool.control$AbundanceProportion <-
ave(ps.melt.x.9.1.genus.stool.control$Abundance,list(ps.melt.x.9.1.
genus.stool.control[,"SampleID"]), FUN=function(L) L/sum(L))
```

```
# Convert to character
ps.melt.x.9.1.genus.stool.control$Genus<-
as.character(ps.melt.x.9.1.genus.stool.control$Genus)

# Rename genera with less than 10% abundance
ps.melt.x.9.1.genus.stool.control$Genus[ps.melt.x.9.1.genus.stool.c
ontrol$AbundanceProportion<0.1]<-"Genera <10% abundance"
View(ps.melt.x.9.1.genus.stool.control)

# We also take care of colors

custom_col42 = c("#781156","#F098A7","#EA6C81","#E43FAD","#784511",
                 "#F098D3","#114578","#D21E2C","#1E78D2","#A5182F",
                 "#6CABEA","#98C4F0","#F7F7C5","#F0C498","#E4913F",
                 "#A55E18","#98F0F0","#117845","#18A55E","#1ED278",
                 "#D2D21E","#E4E43F","#EAEA6C","#F0F098","#F7F7C5",
                 "#784511","#A55E18","#D2781E","#E4913F","#EAAB6C",
                 "#F0C498","#781122","#A5182F","#D21E2C","#E43F5B",
                 "#EA6C81","#F098A7")

# Plot the figure
p<-
ggplot(ps.melt.x.9.1.genus.stool.control,aes(x=sample_ID,y=Abundanc
eProportion))
p+geom_bar(stat="identity",aes(fill=Genus))+
  theme(axis.text.x = element_text(size=12, angle = 90, hjust =
1))+
theme(axis.title.x=element_text(size=14),axis.title.y=element_text(
size=14))+
theme(legend.text=element_text(size=12,face="italic"),legend.title=
element_blank())+
  theme(plot.title=element_text(face="bold",size=14))+
  labs(title="B) Fecal samples at genus Level")+
  scale_fill_manual(values=custom_col42)+
  labs(x="Samples",y="Relative abundance")+
  theme(legend.position="bottom")
ggsave("stool.genus.barplot.pdf", height = 5, width = 11)

######################################################################
      Summary Stacked Barplot at GENUS LEVEL for STOOL samples
######################################################################
# To obtain a barplot with the median proportions for stool samples
at genus level
x.9.2<-xTFM
wh0=genefilter_sample(x.9.2,filterfun_sample(function(x) x>5),
A=0.1*nsamples(x.9.2))
x.9.2<-prune_taxa(wh0,x.9.2)
x.9.2.genus<-tax_glom(x.9.2,taxrank="Genus")
x.9.2.genus.stool <- subset_samples(x.9.2.genus, Sample=="stool")
x.9.2.genus.stool.control <- subset_samples(x.9.2.genus.stool,
Group=="control")
x.9.2.genus.stool.control.Abundance <-
data.frame(otu_table(x.9.2.genus.stool.control))
x.9.2.genus.stool.control.Genera <-
data.frame(tax_table((x.9.2.genus.stool.control)))
genus.stool <- cbind(x.9.2.genus.stool.control.Abundance,
x.9.2.genus.stool.control.Genera)
```

```
dim(genus.stool)#114 genera and 22 variables (16 samples + 6 rank
taxa)

genus.stool <- genus.stool[,c(22,1:16)]
genus.stool <- genus.stool[ ! ( genus.stool$Genus =="uncultured"
),]
genus.stool <- genus.stool[ ! ( genus.stool$Genus =="uncultured_ge"
),]
rownames(genus.stool) <- genus.stool$Genus
genus.stool <- genus.stool[,-1]
genus.stool <- t(genus.stool)
dim(genus.stool) #16 samples and 114 genera

#To transform raw counts to relative abundance within a sample:
genus.stool.rel <- genus.stool/rowSums(genus.stool)

# To determine the median relative abundance for each column/genus.
genus.stool.medianab <- apply(genus.stool.rel,2,median)

#To prepare the data
data.genus.stool <- as.data.frame(genus.stool.medianab)
data.genus.stool <- cbind(data.genus.stool,
row.names(data.genus.stool))
colnames(data.genus.stool) <- c("median.abundance","Genus")
data.genus.stool$median.proportion <-
data.genus.stool$median.abundance/sum(data.genus.stool$median.abund
ance)
sum(data.genus.stool$median.proportion)#1
View(data.genus.stool)

#To remove rows without counts
data.genus.stool <- data.genus.stool[apply(data.genus.stool!=0, 1,
all),]

# Rename genera with less than 10% abundance
levels(data.genus.stool$Genus) <-
c(levels(data.genus.stool$Genus),"Genera <10% abundance")  #Add the
extra level to your factor
data.genus.stool$Genus[data.genus.stool$median.proportion< 0.1]<-
"Genera <10% abundance"

dim(data.genus.stool)#50 rows
data.genus.stool$sample <- c(rep("stool",50))

#Take care of colors
stool.genera.colors2 <- c("#E43FAD","#F0C498","#117845","#6CABEA")

#Get the figure
p <- ggplot(data.genus.stool, aes(x=sample, y=median.proportion))
p + geom_bar(stat="identity", aes(fill=Genus))+
  scale_fill_manual(values=stool.genera.colors2)+
  theme(legend.position = "none",
        axis.title.x = element_blank(),
        axis.title.y = element_blank(),
        axis.text.x = element_blank(),
        axis.text.y = element_blank())
ggsave("Genus.stool.median.proportions.pdf")
```

```
# Plots for sputa samples at phylum and genus level were made in a
similar way

#####################################################################
 Selection of the most abundant genera in sputa samples from ADULTS
#####################################################################
# We select sputa samples from adults
x.7.0.adult <- subset_samples(x.7.0.genus, Sample=="sputum" &
Agecat=="adult")
adult.Abundance <- data.frame(otu_table(x.7.0.adult))
adult.Genera <- data.frame(tax_table((x.7.0.adult)))
adult.all <- cbind(adult.Abundance, adult.Genera)
dim(adult.all)#528 genera and 58 variables (52 samples + 6 rank
taxa)

adult.all <- adult.all[,c(58,1:52)]
adult.all <- adult.all[ ! ( adult.all$Genus =="uncultured" ),]
adult.all <- adult.all[ ! ( adult.all$Genus =="uncultured_ge" ),]
rownames(adult.all) <- adult.all$Genus
adult.all <- adult.all[,-1]
adult.all <- t(adult.all)
dim(adult.all)#52 samples and 503 genera

#To transform raw counts to proportions within a sample:
adult.all.rel <- adult.all/rowSums(adult.all)

#First, determine the MEDIAN relative abundance for each
column/genera.
adult.median.ab <- apply(adult.all.rel,2,median)

#Now, we select the genera with less than 0.1% as their mean
relative abundance
adult.rare <- names(which(adult.median.ab<0.001))#475
adult.abun <- names(which(adult.median.ab>0.001))#28

#This leaves us with 28 genera suggesting that the majority of the
taxa sampled occur at very low relative abundances in adults
data.adult.abun <- adult.all.rel[,-which(colnames(adult.all.rel)
%in% adult.rare)]
dim(data.adult.abun) #52 samples and 28 genera

#Get the MEDIAN relative abundances of these 28 genera
adult.median.rel.abun <- apply(data.adult.abun,2,median)
length(adult.median.rel.abun) #28 genera

#####################################################################
     Heatmap for sputa in ADULTS with Dendrograms and Metadata
#####################################################################
# To remove sample names from heatmaps
rownames(data.adult.abun) <- rep("",nrow(data.adult.abun))

# To calculate the Bray-Curtis dissimilarity matrix
adult.data.dist <- vegdist(data.adult.abun, method = "bray")
adult.row.clus <- hclust(adult.data.dist, "aver")
dend.samples <- hclust(adult.data.dist, "aver")
plot(dend.samples)
```

```
dev.off()

#You can also add a column dendrogram to cluster the genera that
occur #more often together. Note that this one must be done on the
same dataset #that is used in the heat map (with a reduced number
of genera)

#You have to transpose the dataset to get the most abundant genera
as rows
adult.dist.g.abun <- vegdist(t(data.adult.abun), method = "bray")
adult.abun.col.clus <- hclust(adult.dist.g.abun, "aver")
dend.genera <- hclust(adult.dist.g.abun, "aver")
plot(dend.genera)

#Annotations are added with the ann sublist which takes a data
frame as its input.
adult.ann.dat <- data.frame(Age = sample_data(x.7.0.adult)$Age,
                            FEV = sample_data(x.7.0.adult)$FEV,
                            Nut = sample_data(x.7.0.adult)$IMCcat,
                            Col =
sample_data(x.7.0.adult)$Colonization)

#Heatmap for most abundant genera
graphics.off()
par("mar")
par(mar=c(1,1,1,1))
adult.map.abun.black <-annHeatmap2(as.matrix(data.adult.abun),
col = colorRampPalette(c("lightyellow", "red"), space = "rgb")(66),
breaks = 52, dendrogram = list(Row = list(dendro =
as.dendrogram(adult.row.clus)), Col = list(dendro =
as.dendrogram(adult.abun.col.clus))), legend = 4,
labels = list(Row = list(nrow = 1), Col = list(nrow = 12)),
ann = list(Row = list(data = adult.ann.dat)))
plot(adult.map.abun.black, widths=c(2,10,2), heights=c(2,10,2))

######################################################################
    BARPLOTS to compare mean relative abundance AT PHYLUM LEVEL
       in STOOL samples before and after DHA supplementation
######################################################################
# Prepare a dataframe for stool samples before DHA (Time 1)at
phylum level
x.3.0 <- xTFM
x.3.0.phylum <- tax_glom(x.3.0,taxrank="Phylum")
x.3.0.phylum.stool <- subset_samples(x.3.0.phylum, Sample=="stool")

x.3.0.phylum.stool.DHA.1 <- subset_samples(x.3.0.phylum.stool,
Treatment=="DHA" & Time=="1")
x.3.0.phylum.stool.DHA.1.Abundance <-
data.frame(otu_table(x.3.0.phylum.stool.DHA.1))
x.3.0.phylum.stool.DHA.1.Genera <-
data.frame(tax_table((x.3.0.phylum.stool.DHA.1)))
phylum.stool.DHA.1 <- cbind(x.3.0.phylum.stool.DHA.1.Abundance,
x.3.0.phylum.stool.DHA.1.Genera)
dim(phylum.stool.DHA.1)#25 phyla and 13 variables (7 samples + 6
rank taxa)

phylum.stool.DHA.1 <- phylum.stool.DHA.1[,c(9,1:7)]
```

```
rownames(phylum.stool.DHA.1) <- phylum.stool.DHA.1$Phylum
phylum.stool.DHA.1 <- phylum.stool.DHA.1[,-1]
phylum.stool.DHA.1 <- t(phylum.stool.DHA.1)
dim(phylum.stool.DHA.1)#7 samples and 25 phyla

#To transform raw counts to proportions within a sample:
phylum.stool.DHA.1.rel <-
phylum.stool.DHA.1/rowSums(phylum.stool.DHA.1)

# To determine the mean relative abundance for each column/phylum.
phylum.stool.DHA.1.meanab <- apply(phylum.stool.DHA.1.rel,2,mean)

#To prepare the data
data.phylum.stool.DHA.1 <- as.data.frame(phylum.stool.DHA.1.meanab)
View(data.phylum.stool.DHA.1)
data.phylum.stool.DHA.1 <- cbind(data.phylum.stool.DHA.1,
row.names(data.phylum.stool.DHA.1))
colnames(data.phylum.stool.DHA.1) <-
c("mean.rel.abundance","Phylum")
dim(data.phylum.stool.DHA.1)#25 phyla but some of them without
counts

#To remove rows without counts
data.phylum.stool.DHA.1 <-
data.phylum.stool.DHA.1[apply(data.phylum.stool.DHA.1!=0, 1, all),]

# To prepare a dataframe for stool samples after DHA (Time 2) at
phylum level
x.3.0.phylum.stool.DHA.2 <- subset_samples(x.3.0.phylum.stool,
Treatment=="DHA" & Time=="2")
x.3.0.phylum.stool.DHA.2.Abundance <-
data.frame(otu_table(x.3.0.phylum.stool.DHA.2))
x.3.0.phylum.stool.DHA.2.Genera <-
data.frame(tax_table((x.3.0.phylum.stool.DHA.2)))
phylum.stool.DHA.2 <- cbind(x.3.0.phylum.stool.DHA.2.Abundance,
x.3.0.phylum.stool.DHA.2.Genera)
dim(phylum.stool.DHA.2)#25 phyla and 13 variables (7 samples + 6
rank taxa)

phylum.stool.DHA.2 <- phylum.stool.DHA.2[,c(9,1:7)]
rownames(phylum.stool.DHA.2) <- phylum.stool.DHA.2$Phylum
phylum.stool.DHA.2 <- phylum.stool.DHA.2[,-1]
phylum.stool.DHA.2 <- t(phylum.stool.DHA.2)
dim(phylum.stool.DHA.2)#7 samples and 25 phyla

#To transform raw counts to proportions within a sample:
phylum.stool.DHA.2.rel <-
phylum.stool.DHA.2/rowSums(phylum.stool.DHA.2)

# To determine the median relative abundance for each
column/phylum.
phylum.stool.DHA.2.medianab <-
apply(phylum.stool.DHA.2.rel,2,median)

#To prepare the data
data.phylum.stool.DHA.2 <-
as.data.frame(phylum.stool.DHA.2.medianab)
```

```
data.phylum.stool.DHA.2 <- cbind(data.phylum.stool.DHA.2,
row.names(data.phylum.stool.DHA.2))
colnames(data.phylum.stool.DHA.2) <-
c("median.rel.abundance","Phylum")
dim(data.phylum.stool.DHA.2)#25 phyla but some of then without
counts

#To remove rows without counts
data.phylum.stool.DHA.2 <-
data.phylum.stool.DHA.2[apply(data.phylum.stool.DHA.2!=0, 1, all),]

#To join both dataframes and the comparing variable
phylum.stool.DHA <- rbind(data.phylum.stool.DHA.1,
data.phylum.stool.DHA.2)
phylum.stool.DHA$DHA <- c(rep("Before DHA",11), rep("After DHA",
9))

# Rename phyla with less than 1% abundance
levels(phylum.stool.DHA$Phylum) <-
c(levels(phylum.stool.DHA$Phylum),"Phyla <1% abundance")  #Add the
extra level to your factor
phylum.stool.DHA$Phylum[phylum.stool.DHA$median.rel.abundance <
0.01]<-"Phyla <1% abundance"

# Reorder the chunks
phylum.stool.DHA$Phylum <- factor(phylum.stool.DHA$Phylum,
levels
=c("Firmicutes","Proteobacteria","Actinobacteria","Bacteroidetes",
"Verrucomicrobia","Phyla <1% abundance"))

#Change NA to "Phyla <1% abundance"
phylum.stool.DHA[is.na(phylum.stool.DHA)] <- "Phyla <1% abundance"

#Reorder the bars
phylum.stool.DHA$DHA <- factor(phylum.stool.DHA$DHA, levels =
c("Before DHA","After DHA"))

#Take care of colors
phylum.stool.DHA.colors <-
c("indianred1","darkkhaki","mediumseagreen","lightgreen",
                              "lightblue","orchid1")
#Get the figure
p <- ggplot(phylum.stool.DHA, aes(x=DHA, y=mean.rel.abundance))
p + geom_bar(stat="identity", aes(fill=Phylum)) + theme(axis.text.x
= element_text(hjust = 0.5)) +
  theme(legend.text = element_text(size = 10, face = "italic")) +
  labs(title = "Phylum distribution in stool samples")+
  labs(y="Relative
abundance")+scale_fill_manual(values=phylum.stool.DHA.colors)
ggsave("Phylum.stool.DHA.barplot.pdf")

# Barplots to compare median relative abundance AT GENUS LEVEL in
STOOL samples before and after DHA supplementation were made in a
similar way.
```

```
# I followed the same code, only changing the selection of data, to
plot the median relative abundance at phylum and genus levels for
sputum samples.

# We have seen that there is a trend of decreasing/increasing
abundance of certain phylum/genus after the treatment with DHA.
Thus, there are differences but we need to see which of them are
significant

#####################################################################
  Boxplots for Significant Phylum difference in STOOL treated with
                    DHA between Time 1 and Time 2
#####################################################################
x.4.0<-xTFM
x.4.0 = transform_sample_counts(x.4.0, function(x) ((x/sum(x))))
x.4.0<-tax_glom(x.4.0,taxrank="Phylum")
psmelt.x.4.0.phylum<-psmelt(x.4.0)
psmelt.x.4.0.phylum.stool.DHA<-
subset(psmelt.x.4.0.phylum,psmelt.x.4.0.phylum$sample_Sample
=="stool"&

psmelt.x.4.0.phylum$Treatment =="DHA")
summary(psmelt.x.4.0.phylum.stool.DHA)

# Boxplots for significant phylum related to variable Time (before
and after DHA)
mysignificantPhylumStoolDHA<-vector()
my.p.values.vectorPhylumStoolDHA<-vector()

# Find significant phylum (p<0.05)
mainDir<-"./"
subDir<-"StatisticalTestsPhylumStoolDHA"
if (file.exists(subDir)){
  #setwd(file.path(mainDir, subDir))
} else {
  dir.create(file.path(mainDir, subDir))
  #setwd(file.path(mainDir, subDir))
}

# For every Phylum we will calculate a Wilcoxon Rank-based test
between dicotomic variable Time
for(Phylum in
as.vector(unique(psmelt.x.4.0.phylum.stool.DHA$Phylum))){
mytest<-
wilcox.test(Abundance~Time,data=psmelt.x.4.0.phylum.stool.DHA[psmel
t.x.4.0.phylum.stool.DHA$Phylum==Phylum,])
capture.output(mytest,file=paste(mainDir,"/",subDir,"/",Phylum,"_Tr
eatment_Wilcoxon.txt",sep=""))
my.p.values.vectorPhylumStoolDHA<-
c(my.p.values.vectorPhylumStoolDHA,mytest$p.value)
}

# We need to correct for multiple test error, we use benjamin-
hochberg
my.p.values.vectorPhylumStoolDHA.adj<-
p.adjust(my.p.values.vectorPhylumStoolDHA,method="BH")
```

```
mysignificantPhylumStoolDHA<-
unique(psmelt.x.4.0.phylum.stool.DHA$Phylum)[my.p.values.vectorPhyl
umStoolDHA.adj<0.05]
ps.x.4.0.phylum.stool.DHA<-
psmelt.x.4.0.phylum.stool.DHA[psmelt.x.4.0.phylum.stool.DHA$Phylum
%in% mysignificantPhylumStoolDHA,]

p<-ggplot(ps.x.4.0.phylum.stool.DHA,aes(x=Phylum,y=Abundance))
p # There are not significant differences in phyla before and after
DHA

######################################################################
Boxplots for Significant GENUS difference in STOOL treated with DHA
                      between Time 1 and Time 2
######################################################################
psmelt.x.4.0.genus.stool.DHA<-
subset(psmelt.x.4.0.genus,psmelt.x.4.0.genus$sample_Sample
=="stool"&psmelt.x.4.0.genus$Treatment =="DHA")
summary(psmelt.x.4.0.genus.stool.DHA)

# Boxplots for significant phylum related to variable Time (before
and after DHA)
mysignificantGenusStoolDHA<-vector()
my.p.valuesGenusStoolDHA<-vector()

# Find significant genus (p<0.05)
mainDir<-"./"
subDir<-"StatisticalTestsGenusStoolDHA"
if (file.exists(subDir)){
  #setwd(file.path(mainDir, subDir))
} else {
  dir.create(file.path(mainDir, subDir))
  #setwd(file.path(mainDir, subDir))
}

# For every Genus we will calculate a Wilcoxon Rank-based test
between dicotomic variable Time
for(genus in
as.vector(unique(psmelt.x.4.0.genus.stool.DHA$Genus))){
  mytest<-
wilcox.test(Abundance~Time,data=psmelt.x.4.0.genus.stool.DHA[psmelt
.x.4.0.genus.stool.DHA$Genus==genus,])
capture.output(mytest,file=paste(mainDir,"/",subDir,"/",genus,"_Tre
atment_Wilcoxon.txt",sep=""))
my.p.valuesGenusStoolDHA<-
c(my.p.valuesGenusStoolDHA,mytest$p.value)
}

# We need to correct for multiple test error, we use benjamin-
hochberg
my.p.valuesGenusStoolDHA.adj<-
p.adjust(my.p.valuesGenusStoolDHA,method="BH")
mysignificantGenusStoolDHA<-
unique(psmelt.x.4.0.genus.stool.DHA$Genus)[my.p.values.vectorGenusS
toolDHA.adj<0.05]
```

```
ps.x.4.0.genus.stool.DHA<-
psmelt.x.4.0.genus.stool.DHA[psmelt.x.4.0.genus.stool.DHA$Genus
%in% mysignificantGenusStoolDHA,]

p<-ggplot(ps.x.4.0.genus.stool.DHA,aes(x=Genus,y=Abundance))
p # There are not significant differences in phyla before and after
DHA

# I made a similar analysis to look for significant PHYLA or GENUS
differences in sputa treated with DHA between Time 1 and Time 2.
```