Universidade de Aveiro Departamento de Biologia
2017

Maria Luísa Saragoça Falcão de Brito

Análise metagenómica do microbioma da saliva de pacientes com Doença Pulmonar Obstrutiva Crónica

Metagenomic analysis of saliva microbiome in patients with Chronic Obstructive Pulmonary Disease

Universidade de Aveiro Departamento de Biologia
2017

Maria Luísa Saragoça Falcão de Brito

Análise metagenómica do microbioma da saliva de pacientes com Doença Pulmunonar Obstrutiva Crónica

Metagenomic analysis of saliva microbiome in patients with Chronic Obstructive Pulmonary Disease

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Biologia Molecular e Celular, realizada sob a orientação científica da Doutora Gabriela Maria Ferreira Ribeiro de Moura, Professora Auxiliar do Departamento das Ciências Médicas da Universidade de Aveiro e co-orientação da Doutora Alda Sofia Pires de Dias Marques, Professora Adjunta da Escola Superior de Saúde da Universidade de Aveiro.

Financiado pelo Programa Operacional de Competitividade e Internacionalização COMPETE, através do Fundo Europeu de Desenvolvimento Regional FEDER (POCI-01-0145-FEDER-016701), Fundação para Ciência e Tecnologia (PTDC/DTPPIC/2284/2014) e através do projeto UID/BIM/04501/2013.

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR
Cofinanciado por: PORTUGAL 2020
UNIÃO EUROPEIA
Fundo Europeu de Desenvolvimento Regional

DECLARAÇÃO

Declaro que este relatório é integralmente da minha autoria, estando devidamente referenciadas as fontes e obras consultadas, bem como identificadas de modo claro as citações dessas obras. Não contém, por isso, qualquer tipo de plágio quer de textos publicados, qualquer que seja o meio dessa publicação, incluindo meios eletrónicos, quer de trabalhos académicos.

Dedico este trabalho aos meus avós

o júri

presidente	Professora Doutora Helena Silva Professora Auxiliar do Departamento de Biologia da Universidade de Aveiro
arguente	Doutora Isabel da Silva Henriques Investigadora Auxiliar do Departamento de Biologia da Universidade de Aveiro
orientador	Professora Doutora Gabriela Maria Ferreira Ribeiro de Moura Professora Auxiliar do Departamento de Ciências Médicas da Universidade de Aveiro

agradecimentos

Em primeiro lugar queria agradecer à minha orientadora, a Doutora Gabriela Moura, por todo o acompanhamento e ensinamentos que me deu ao longo deste ano.

Também gostaria de agradecer à Doutora Alda Marques, pois sem ela não teria tido a oportunidade de trabalhar neste tema que tanto me fascina. Bem como às bolseiras de investigação da área da fisioterapia, que procederam a todas as recolhas, por toda a amabilidade e disponibilidade.

À Sofia por tudo o que ensinou e por toda a ajuda no laboratório. E ao Hélder por ter estado sempre pronto para me ajudar durante todo estre processo.

À Andreia, muito obrigado por toda a dedicação, paciência, ajuda e incentivo. Sem dúvida, uma das pessoas essenciais neste processo.

A todos os meus amigos, por todo o encorajamento e por me terem conseguido acalmar nos momentos mais difíceis. Em particular, o Pedro e a Diana, que por estarem no mesmo barco tiveram sempre a palavra certa no momento certo. Um obrigado especial, à Diana com quem compartilhei tantas horas de trabalho, repletas de desabafos, conselhos, sorrisos e lágrimas. Sem ela, tudo teria sido bem mais difícil!

Aos meus pais e irmãs, por toda a paciência e apoio que me deram e por terem sempre acreditado em mim. Em especial, aos meus pais, por terem incentivado a ir sempre mais além e a nunca desistir.

palavras-chave metagenómica, microbioma, DPOC, exacerbação, saliva, biomarcador, bacteria, 16S rRNA, medicina personalizada

resumo

Microbioma é definido como sendo uma comunidade de microrganismos presente num dado ambiente, que engloba todos os microorganismos com seus genes e interações ambientais. O microbioma humano desempenha um papel importante na fisiologia humana e no seu metabolismo, estando associado ao desenvolvimento, nutrição, imunidade e resistência a agentes patogénicos com implicações na saúde e doença.

A doença pulmonar obstrutiva crónica (DPOC) é uma doença pulmonar caracterizada por uma obstrução das vias aéreas persistente progressiva e não reversível. O papel das bactérias como potencial fator patogénico e etiológico na DPOC tem sido tema de debate nos últimos anos. Pensa-se que a colonização dos pulmões por determinadas bactérias, em pacientes com DPOC, é responsável pelo aumento do risco de exacerbações e perda de função pulmonar.

Embora a saliva seja uma das amostras mais facilmente recolhida, são ainda poucos os estudos para caracterizar o microbioma da saliva em pacientes com DPOC, e ainda menos para identificar nele biomarcadores informativos sobre o diagnóstico e progressão desta doença.

O objetivo deste estudo foi implementar a metodologia que permita estudar o microbioma da saliva em pacientes com DPOC, compreender a dinâmica do microbioma da saliva no contexto de uma exacerbação e como o microbioma evolui depois disso.

Para isso, utilizou-se uma abordagem metagenómica utilizando a sequenciação do gene 16S rRNA, para analisar 17 amostras de 7 pacientes com DPOC, recolhidas em 3 momentos diferentes, i.e. em exacerbação, 2 semanas após a exacerbação e após recuperação clínica.

Neste estudo foram encontradas e serão descritas diferenças na composição microbiana das amostras colhidas em tempos diferentes. Verificou-se também uma grande variabilidade nos resultados, com grandes diferenças entre as amostras colhidas de diferentes pacientes.

Estes resultados sugerem que a saliva pode ser uma boa fonte de biomarcadores para a DPOC e poderá representar um avanço para a implementação da medicina personalizada nesta população. No entanto mais estudos com amostras alargadas são ainda necessários. Contudo, mais estudos deverão ser realizados.

keywords metagenomics, microbiome, COPD, exacerbation, saliva, biomarker, bacteria, 16S rRNA, personalized medicine

abstract

Microbiome is a community of microorganisms living in a particular environment that englobes all microorganisms with their genes and environmental interactions. The human microbiome plays a pivotal role in human physiology and metabolism being associated to development, nutrition, immunity, and resistance to pathogens and has recognized implications for health and disease.

Chronic Obstructive Pulmonary Disease (COPD) is a pulmonary disease characterized by persistent and progressive and nonreversible airflow obstruction. The role of bacteria as a potential pathogenic and etiologic factor in COPD has been a topic of debate for many years. It is thought that lung colonization by particular bacterial strains, in patients with COPD, is responsible for the chronic bronchitis phenotype, increased risk of exacerbations, and loss of lung function.

Even though saliva is one of the most easily collectable samples, few studies have been conducted to characterize the saliva microbiome in patients with COPD and even fewer to identify biomarkers that might be informative for disease onset and progression.

The aim of this study was to implement the methodology to study the saliva microbiome in patients suffering with COPD, to understand the dynamics of saliva microbiome in the setting of an exacerbation and how the microbiome evolve after that.

For that a metagenomic approach was carried out, using the sequencing of the 16S rRNA gene, to analyze 17 samples from 7 patients with COPD, collected at 3 different time points, i.e. at exacerbation, 2 weeks after exacerbation, and at clinical full recovery.

In this study, we found microbial shifts in the samples collected at different time points. We also detected high sample variability, especially between samples collected from different individuals.

These results suggest that saliva might me a good source of biomarkers for COPD management and may represent an improvement to the implementation of personalized medicine in this population. However, more and larger studies must be conducted.

Table of contents

List of abbreviations

16S rRNA	16S ribosomal RNA
AECOPD	Acute Exacerbation of Chronic Obstructive Pulmonary Disease
BAL	Bronchoalveolar Lavage
COPD	Chronic Obstructive Pulmonary Disease
ddNTP	dideoxynucleotide triphosphate
dNTP	deoxynucleoside triphosphate
FEV_1	Forced Expirometry Volume in 1 second
GOLD	Global Initiative on Obstructive Lung Disease
iBiMED	Institute for Research in Biomedicine
IGC	Instituto Gulbenkian de Ciência
NGS	Next Generation Sequencing
OTUs	Operational Taxonomic Units
PCoA	Principal Coordinates Analysis
PCR	Polymerase Chain Reaction
rRNA	ribosomal Ribonucleic Acid
SBS	Sequencing by Synthesis
WHO	World Health Organization

List of figures

List of tables

Chapter I – Introduction

1.1. Microbiome

Microbiome is a community of microorganisms living in a particular environment that englobes all the microorganisms with their genes and environmental interactions (Cox et al. 2013). The human microbiome has been defined as “the ecological community of commensal, symbiotic, and pathogenic microorganisms that literally share our body space” (Lederberg & McCray 2001). The human microbiome colonizes a wide range of body niches including skin and mucosae, gut, oral cavity, upper respiratory tract and lung. It is composed by a large variety of microbial communities such as bacteria, archaea, viruses, phage, fungi, and other microbial eukarya, being bacteria the larger fraction and the most well-studied group (Cox et al. 2013).

Today it is known that we have as many microorganisms as we have body cells, around 40 trillion (Sender et al. 2016). Most microbes that are living in our body are commensals and provide us with genetic variation and gene functions that human cells did not evolve on their own (Grice & Segre 2012). Interestingly, the human microbiome plays a pivotal role in human physiology and metabolism being associated to development, nutrition, immunity, and resistance to pathogens with implications for health and disease (Sender et al. 2016; Li et al. 2012). This is why the human microbiome has been widely studied in the past two decades (Sender et al. 2016).

The human microbiome was shown to be a dynamic population that can vary by the introduction or extinction of certain microbial groups or by a change in the population structure - known as dysbiosis (Cho & Blaser 2012). These alterations can be induced by selection through environmental factors, such as host dietary changes, antibiotics administration or exposure to pollutants (Blaser 2006). Dysbiosis disturb the partnership between the host immune system and the microbiota, and may lead to altered immune responses that may underlie several inflammatory disorders, such as inflammatory bowel disease, obesity, diabetes, and colorectal cancer (Degruttola et al. 2016; Round & Mazmanian 2009). Therefore, some authors are currently considering that the interplay between the microbiome and human hosts represent the next frontier in medicine (Mammen & Sethi 2016).

Dysbiosis in the airway microbiome have been associated with multiple pulmonary diseases such as asthma, chronic obstructive pulmonary disease and cystic fibrosis (Marsland et al. 2013). To study the airway microbiome one can collect different types of samples, being the most common the sputum (Wang et al. 2016; Huang et al. 2014; Millares et al. 2014; Molyneaux et al. 2013), followed by others such as bronchiolar alveolar lavage (BAL) (Hilty et al. 2010), lung tissue sample (Sze et al. 2015; Sze et al. 2012), and oropharyngeal swab samples (Liu et al. 2017; Diao et al. 2017; Park et al. 2014).

Even though they all have the goal of informing about the airway microbiome, the results of choosing different kinds of samples can be very divergent (Appendix 1). As we can see on the work of Cabrera-Rubio et al. (2012), where sputum, bronchial aspirate (which represents the upper respiratory tract), BAL and bronchial mucosal biopsy (representing the lower respiratory tract), from stable patients with chronic obstructive pulmonary disease, were collected and compared. The results showed that the sputum had lower diversity than the other sample types. Additionally, the bronchial mucosa and BAL samples showed similar bacterial compositions in contrast to sputum and bronchial aspirate samples. Thus, from this study, it can be concluded that samples collected from the upper respiratory tract are not representative of the lower respiratory tract (Cabrera-Rubio et al. 2012).

On the other hand, there are some other authors that suggest that the community composition exhibits minimal variation inside the airways (Liu et al. 2017; Charlson et al. 2011; Charlson et al. 2010). When compared naso- and oropharyngeal samples and samples from the upper (oral, oro/nasopharyngeal) and lower airways (glottis, BAL, lower airway brush) Charlson and colleagues (2010, 2011), and later Liu et al. (2017) with oropharyngeal and sputum samples, found that despite the differences in the body site, microbiomes in the airways are remarkably similar.

Consistently, the oral microbiome is more similar to the lung microbiome than with the microbiome present in the air (Dickson et al. 2014). This can be explained by the fact that the human respiratory tract is a continuous system, from the nasal and oral cavities to upper and lower airways, without any physical barrier (Dickson & Huffnagle 2015). It is believed that the movement of the microbes occurs regularly

via breathing, microaspiration and mucosal dispersion, and this movement is also bidirectional (Charlson et al. 2011).

A study conducted, with healthy subjects, defined a core saliva microbiome as composed by members of the Firmicutes phylum (genus *Streptococcus*, family *Veillonellaceae*, genus *Granulicatella*), Proteobacteria (genera *Neisseria* and *Haemophilus*), Actinobacteria (genera *Corybacterium*, *Rothia*, and *Actinomyces*), Bacteriodetes (genera *Prevotella*, *Capnocytophaga*, *Porphyromonas*) and Fusobacteria (genus *Fusobacterium*) (Zaura et al. 2009).

When comparing the oral microbiome and BAL samples, of healthy subjects, lung bacterial populations have been found remarkably similar to the oral microbiome, even though at lower concentrations in the oral cavity (Bassis et al. 2015). These results confirm that the bacteria from the upper airway shape the microbial community of lower airways of healthy subjects (Bassis et al. 2015; Morris et al. 2013; Pragman et al. 2012; Charlson et al. 2011), although they are found in lower concentrations (Morris et al. 2013; Charlson et al. 2011), and may have different community composition (Bassis et al. 2015).

Therefore, the type of respiratory specimen should be considered in the interpretation of the results of lung microbiome studies. Nevertheless, the use of saliva to access markers of human health has been already described. Torres et al. (2015), suggested that bacteria abundance profiles in saliva may be useful biomarkers for pancreatic cancer. And this is also true for respiratory diseases, Zemanick et al. (2015), compared saliva samples with samples from the lower respiratory tract of children suffering with cystic fibrosis. They concluded that differences in the lower and upper airways communities were associated with airway inflammation and may be used as biomarkers for disease progression (Zemanick et al. 2015). Thus, these specimen may be informative when dealing with respiratory as well as non-respiratory diseases (Acharya et al. 2017).

1.1.1 Methodologies for studying the microbiome

Historically, cultivation and isolation of bacteria have been the unique method available for identification and characterization of microbes (Grice & Segre 2012). After isolation of a colony, bacterial have been classified by direct observation of bacterial cells, their morphology, their components, biochemical testing, and their ability to grow on different culture media (Grice & Segre 2012; Medini et al. 2008).

However, these methodologies introduce biases, since they select only the microbes that grow in isolation and under specific laboratory conditions, for that reason the complete diversity of microorganisms remain largely unknown when these approaches are used (Grice & Segre 2012). It is estimated that the culturable organisms represent less than 2% of all the organisms within a sample (Wade 2002). In addition, these methodologies are time consuming and expensive (Hugenholtz 2002).

In 1977, Carl Woese and George Fox proposed that it is possible to characterize the tree of life using genetic sequences of the small-subunit ribosomal RNA, such as the 16S ribosomal RNA (rRNA) gene (Woese & Fox 1977). This gene is a highly conserved component of the transcriptional machinery of all bacteria (Cox et al. 2013), and consists of a unique mosaic structure including extreme conserved regions and variable regions with specific evolutionary rates (Bik 2016). More precisely, the gene has a length of 1522 base pairs and is composed by nine hyper-variable regions, labelled V1-V9. These regions have varied throughout bacterial evolution, thus being useful for taxonomic identification and evolutionary dating (Figure 1) (Van de Peer 1996).

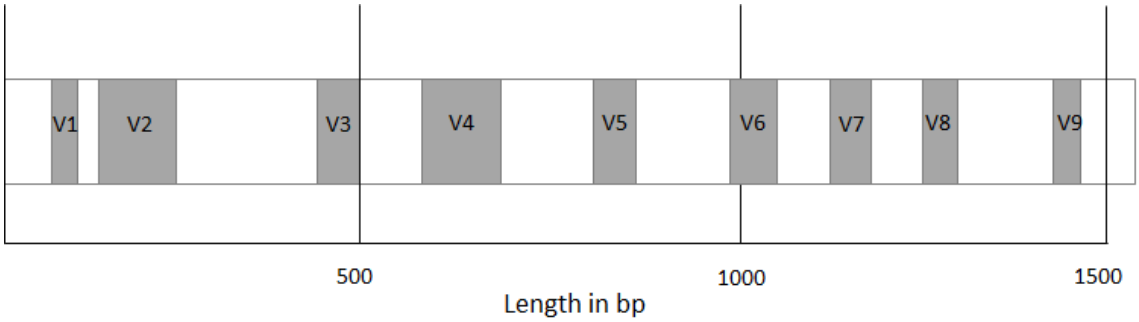


Figure 1 - 16S rRNA gene of *E. coli* showing the nine hypervariable regions (adapted from Cox et al. 2013).

In 1977, Frederick Sanger and colleagues developed the first DNA sequencing method - Sanger Sequencing or First-generation Sequencing (Sanger et al. 1977). This method consists of a “cycle sequencing” reaction, in which cycles of template denaturation, primer annealing and primer extension are performed (Shendure & Ji 2008). Each round of primer extension ends with the incorporation of a fluorescent labeled dideoxynucleotides (ddNTPs), resulting in a mixture of fragments that all have a last nucleotide labelled with a dye which corresponds to the last base inserted (Shendure & Ji 2008). Then all the fragments are separated, based on their length, through a capillary electrophoresis (Schadt et al. 2010). Afterwards, these labelled fragments pass through a detection unit, where the fluorophores are excited by a laser, producing fluorescence emissions of four different colors (based on the four different dNTPs) (Metzker 2005). The information about the last base is used to produce a sequence “trace” that reflects the original DNA sequence (Schadt et al. 2010).

This process results in a read length up to 1000bp, an accuracy per base of 99,999% (Shendure & Ji 2008). Although it has some limitations, such as being time consuming and expensive (Schadt et al. 2010), the sequencing of the 16S rRNA gene with the Sanger technology came to revolutionize the study and classification of microorganisms (Escobar-Zepeda et al. 2015).

Most of the existing understanding of the human microbiome still comes from culture-based approaches using the sequencing of 16S rRNA gene (The NIH HMP Working Group 2009). However, in response to the lower throughput and higher costs provide by the automated Sanger method, in the beginning of the 20th century emerged the Next Generation Sequencing (NGS) technologies. These new sequencing technologies have three major improvements. First, instead of requiring bacterial cloning of DNA fragments they depend on the preparation of NGS libraries in a cell-free system. Second, thousands to millions of sequencing reactions are produced in parallel. Third, the sequencing output is detected with no need for an electrophoresis (Van Dijk et al. 2014). Thus, these technologies have the ability to produce an enormous amount of data at a much lower cost (Metzker 2010).

Among the NGS technologies the most used are the Illumina sequencing by Illumina, the Sequencing by Oligo Ligation Detection (SOLiD) by Life Technologies, and the PacBio systems (Van Dijk et al. 2014).

With the advent of NGS technologies and bioinformatics tools, the determination of the global microbial ecology of a biological sample, and not just the evaluation of those that are able to grow on laboratory growth media, became possible (Mammen & Sethi 2016). These current molecular technologies allow the creation of an entire “community fingerprint”, by identifying members, as well as quantifying and estimating relative bacterial abundance within the community. This made possible to consider the microbial population as a whole “metagenome” by applying a high-throughput sequencing to the entire DNA population of a sample (Han et al. 2012). Globally, these processes are called metagenomics.

1.2 Metagenomics

Metagenomics is defined as “the application of modern genomics techniques to the study of communities of microbial organisms directly in their environments, without the prior need for isolation and cultivation of individual species” (Chen & Pachter 2005). These allow the identification of both community members within a sample and their genetically encoded functional capacity (Han et al. 2012).

Actually, the term can be used for functional and sequence-based analysis of the collective microbial genomes of the microbial community – *shotgun metagenomics sequencing*, or for studies performing a PCR amplification of certain genes, being the 16S rRNA gene the most used – *amplicon sequencing metagenomics* (Oulas et al. 2015). A summary of the two different metagenomics approaches is shown in figure 2.

Shotgun metagenomics tries to answer questions such as: who is within the sample? What are they doing? What are they capable of doing? And how do these microorganisms interact with each other? However, this method requires an enormous informatics power since it produces large amounts of data (Oulas et al. 2015; Sharpton 2014).

Amplicon sequencing metagenomics or marker gene metagenomics is the fastest and more widely used approach to obtain a taxonomic and phylogenetical distribution profile of a sample (Oulas et al. 2015). However, it has some limitations such as the lack of phylogenetic resolution, especially with short read lengths, and the fact that distinct bacterial species contain nearly identical 16S rRNA gene sequences, that leads to widely varying estimates of diversity (Poretsky et al. 2014). In addition, this methodology provides information about the taxonomic composition of the community, but fails to recall the biological functions of these microbiomes within the community (Sharpton 2014).

Nevertheless, mostly for economical reasons, sequencing the 16S rRNA gene is the most used approach to characterize the diversity of microbiota. Under the scope of this method, a community is sampled (e.g. saliva, sputum) and DNA is extracted from all existing cells (Sharpton 2014). Then, specific primers are used to

flank one or two of the hypervariable regions and these regions are amplified through a polymerase chain reaction (PCR) (Huang et al. 2017; Cox et al. 2013).

Then amplicons are sequenced through a NGS technology, which enables the study of many samples at low cost (Metzker 2010). After sequencing, the sequences are clustered into Operational Taxonomic Units (OTUs) – an operational definition to classify, by similarity, groups of DNA sequences as belonging to one species or one group of species (Blaxter et al. 2005). Typically, the similarity between all pairs of sequences is computed as the percentage of common bases in a pairwise sequence alignment – and, in order to belong to the same OTU, a group of sequences must normally share 97% of similarity (Mammen & Sethi 2016; Nguyen et al. 2016). This value was conventionally assumed to represent bacterial species, when using 16S rRNA sequencing (Drancourt et al. 2000).

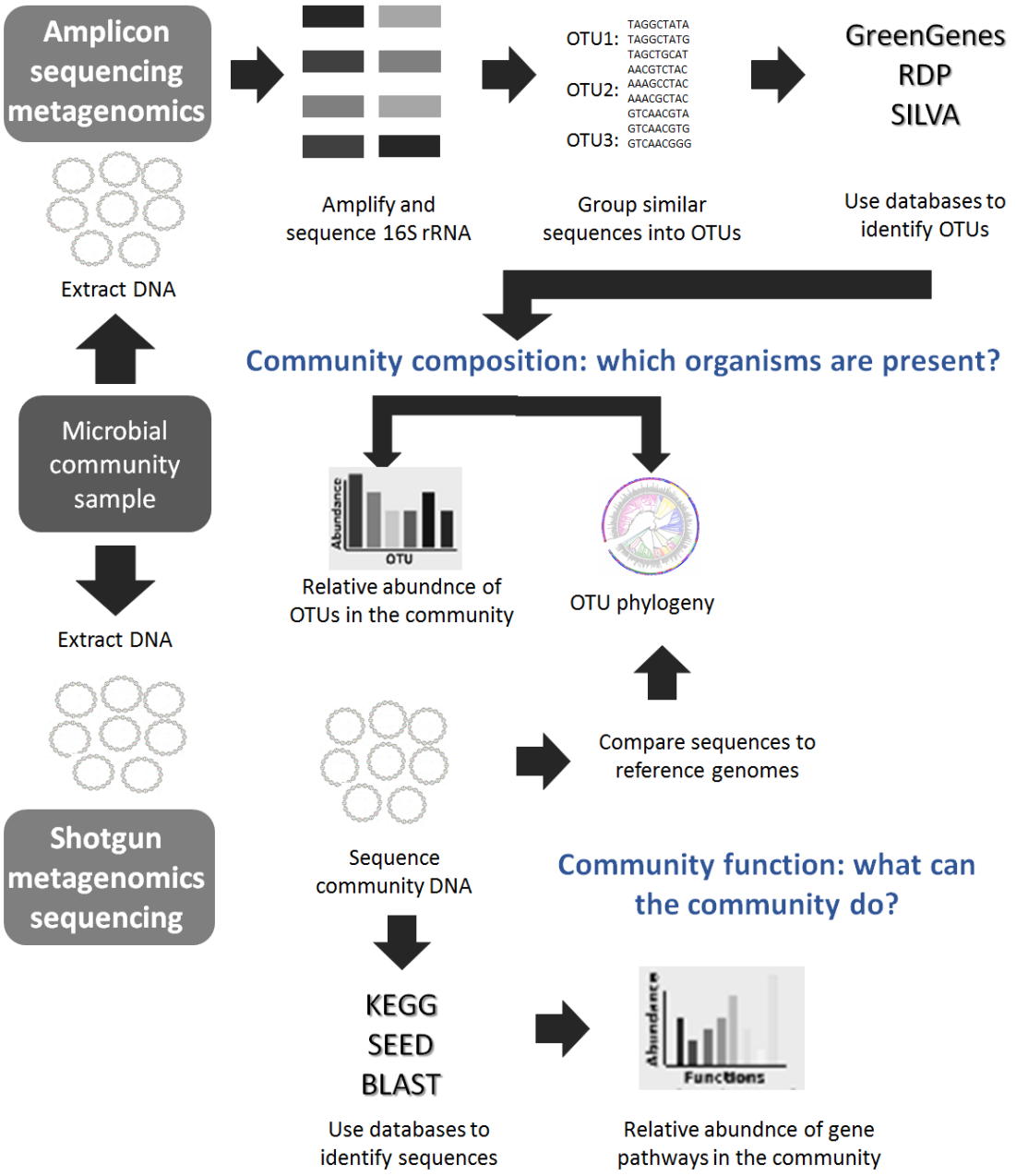


Figure 2 - Metagenomics workflow. First, microbial community DNA is extracted from a sample. Then, two approaches can be done. Amplicon sequencing metagenomics, where the 16S rRNA gene is amplified and sequenced. Similar sequences are then grouped into OTUs, which can be compared to 16S databases to identify their species of origin. The community can be described in terms of which OTUs are present, their relative abundance, and/ or their phylogenetic relationships. The other approach – shotgun metagenomics, comprehends the direct sequencing of the DNA of the community. DNA reads are then compared to reference genomes, so that all genes present are identified and quantified. Using this approach, the functional capabilities of the community can also be determined by comparing the sequences to functional databases (Adapted from: (Morgan & Huttenhower 2012)).

Once OTUs are defined, one can use them to characterize the microbial community, in terms of alpha-diversity – which OTU's are present within the sample or groups of samples i.e. richness; as well as how evenly these OTU's are distributed within the sample or group of samples i.e. diversity. To study the richness of a sample, the most frequently applied measures are the number of OTU's observed, the Chao1 estimator (Chao 1984), and the Abundance-based Coverage (ACE) Estimator (Morgan & Huttenhower 2012). One can also study the beta-diversity, i.e. the similarities, or differences, in the microbial composition between samples. The most common metrics used to study the beta-diversity in metagenomic studies are the Bray-Curtis dissimilarity (Bray & Curtis 1957) and the UniFrac metric (Lozupone & Knight 2005) (Navas-Molina et al. 2013; Morgan & Huttenhower 2012).

1.3 Chronic Obstructive Pulmonary Disease

Chronic Obstructive Pulmonary Disease (COPD) is a pulmonary disease characterized by persistent and progressive airflow obstruction – that is the result of a mixture of airway diseases, such as chronic obstructive bronchiolitis (obstruction of the small airways), emphysema (destruction of alveolus), and chronic bronchitis (mucus hypersecretion) (Barnes et al. 2015), as shown in figure 3. Consequently, patients with COPD suffer with persistent and progressive dyspnea, chronic cough, and chronic sputum production (GOLD 2017).

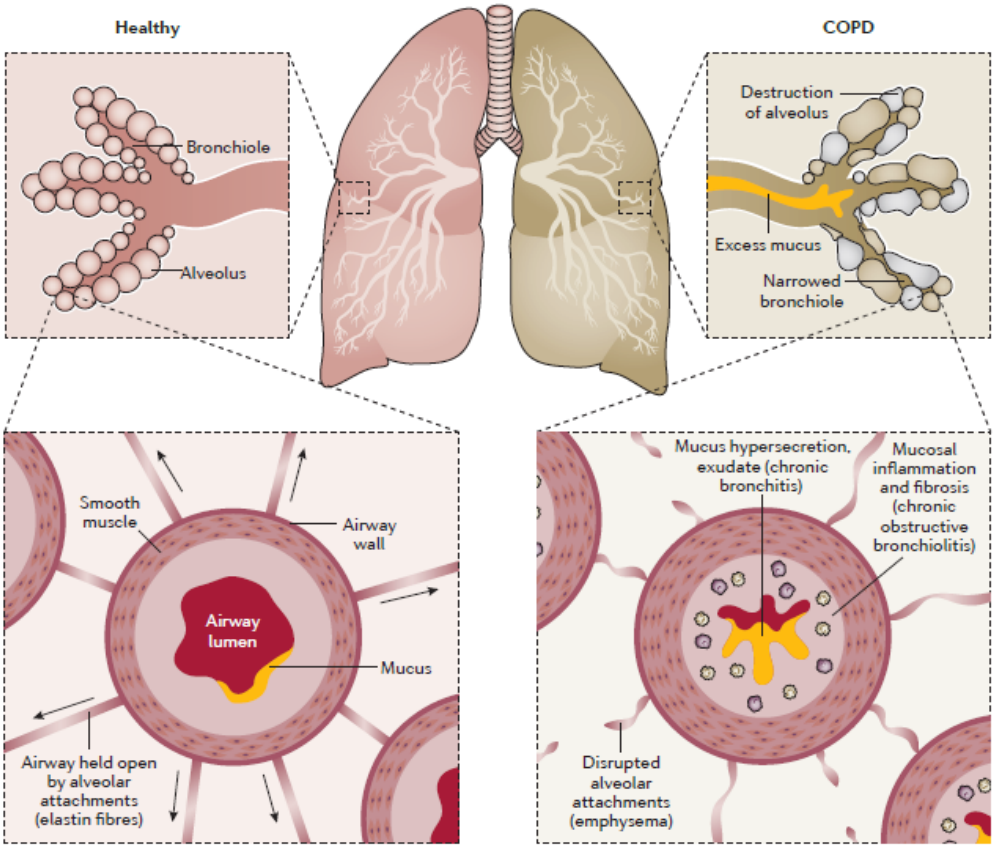


Figure 3 - Comparison between lungs in healthy subjects and in patients with COPD. COPD is characterized by a mucus hypersecretion (chronic bronchitis), a mucosal inflammation and fibrosis (chronic obstructive bronchiolitis), and a destruction of alveolus (emphysema), source: Barnes et al. 2015.

According to Global Initiative on Obstructive Lung Disease (GOLD), the disease can be divided into 4 different stages, based on the airflow limitation, and consequently based on the forced expiratory volume in one second (FEV1) – the volume of air exhaled during the first second (GOLD 2017), as shown in table 1.

Table 1 – Classification of severity of airflow limitation in COPD – The 4 stages of COPD, based on FEV_1. In patients with (FEV1/FVC <70%).

GOLD 1	Mild COPD, $FEV_1 \geq 80\%$ of normal
GOLD 2	Moderate COPD, $80\% > FEV_1 \geq 50\%$ of normal
GOLD 3	Severe COPD, $50\% > FEV_1 \geq 30\%$ of normal
GOLD 4	Very severe COPD, $FEV_1 < 30\%$ of normal

Even though this classification is not clinically accurate, since it does not take into account some other characteristics of patients with COPD, such as the risk of exacerbation, hospitalization, or health status impairment (GOLD 2017), on the microbiome research field, this is still the standard used classification. So in the present study we will also use it.

This disease is one of the most common diseases that affect the lungs and causes significant morbidity and mortality. According to the World Health Organization (WHO), 251 million people worldwide suffer with COPD, resulting in approximately 3 million deaths per year (WHO 2017a), leading to a major public health problem (Nazir & Erbland 2009). In addition it is estimated that in 2030 this will be the third leading cause of death worldwide (WHO 2017a).

The major cause of COPD is tobacco smoking, including passive exposure (WHO 2007). Nevertheless, non-smokers also develop the disease, and so there are other risk factors like outdoor, occupational exposure to air pollution, (including organic and inorganic dusts and chemical agents), history of pulmonary tuberculosis, asthma and respiratory-tract infections during childhood (Salvi & Barnes 2009). There are also genetic risk factors that, together with environmental factors, can influence the susceptibility to the disease or even its evolution once established. One of the well-studied gene associated with COPD is the matrix metalloproteinase 12 gene (MMP12) that has been related to loss of lung function and increased susceptibility to COPD in smokers (Hunninghake et al. 2009).

COPD is frequently punctuated by acute exacerbations (AECOPD). This is an acute event characterized by a worsening of the patient's respiratory symptoms that is beyond normal day-to-day variations and leads to change in the regular medication (GOLD 2017). During respiratory exacerbations there is an increased hyperinflation and gas trapping, with reduced expiratory flow, thus accounting for the increased dyspnea (Parker et al. 2005). Some of the symptoms include an increased dyspnea, cough, and increased sputum volume and/or purulence (GOLD 2017).

These events can lead to considerable morbidity and mortality (Nazir & Erbland 2009). Most of the AECOPD are accompanied by a decrease in lung function, that is usually transient but a significant percentage of patients may not fully recover leading to a further compromised lung function and a growing decline of the patient health status (Nazir & Erbland 2009; Seemungal et al. 2000). Even if the recovery was not fully accomplished, a patient which do not have an exacerbation for 4 to 6 weeks is clinically considered as being recovered (Seemungal et al. 2000).

These acute events are triggered by bacterial or viral infections (which may coexist), environmental pollutants, or unknown factors (GOLD 2017). Most of them are triggered by microbes, being the bacteria more frequently implicated the *Haemophilus influenzae*, *Moraxella catarrhalis*, *Streptococcus pneumoniae* and *Pseudomonas aeruginosa* (Parameswaran & Murphy 2009). Although, there are also non-bacterial exacerbations, it is usually difficult to distinguish between them without a sputum culture. In general, one can say that the bacterial exacerbations are more severe than the non-bacterial ones (Sethi et al. 2008).

To treat the exacerbations the most commonly used medication consists of bronchodilators, corticosteroids, and antibiotics, such as azithromycin and erythromycin (GOLD 2017). However, the use of antibiotics is still controversial (GOLD 2017; Miravitlles & Anzueto 2017). Since several studies suggested that the treatment with antibiotics has no effect in some patients, and therefore the antibiotics side effects, i.e. costs and multi-resistance, should be avoided (Vollenweider et al. 2012).

According to the WHO, a good management of COPD passes through the ability to monitor disease, reduce risk factors, manage stable patients with COPD, and mitigate exacerbation events (WHO 2017b). Although current models to predict these acute events are based on combinations of information about patient history, clinical characteristics, and test results that include biomarkers (e.g. eosinophil counts (Pascoe et al. 2015) and C-reactive protein levels in blood (Peng et al. 2013)). Nevertheless, none of the existing models accomplish the criteria to being used in clinic (Guerra et al. 2017; GOLD 2017). Due to this and since exacerbations increase the decline in lung function, deterioration in health status, hospitalization, and risk of death (GOLD 2017), it would be pivotal to find a model that could be a good predictor of these acute events.

1.4 Microbiome and COPD

Lung microbiome has been studied based on culture-dependent methods although, these methods underrepresent the diversity of the microbial community within a sample (Park et al. 2014). As previously explained, this field had its highest development with the emergence of molecular culture-independent methods (Huang et al. 2017). Therefore, the last 5 years contributed the most of our knowledge in this field, leading to an increased understanding of the roles of microbial communities in health and disease, including COPD (Huang et al. 2017).

It is thought that lung colonization by particular bacterial strains, in patients with COPD, is responsible for the chronic bronchitis phenotype, increased risk of exacerbations, and loss of lung function, which makes microbiome a potential important tool to access COPD progression (Han et al. 2012).

There are some studies addressing the importance of the microbiome, especially the lung microbiome, for the development and worsening of the disease state. Some focus on comparing the microbiome of healthy subjects to patients with COPD (Sze et al. 2015; Park et al. 2014; Sze et al. 2012). Others focus on temporal microbial changes that are observed longitudinally over the course of the disease, as, for example, the differences between the microbiome during the exacerbation and how it evolves after an AECOPD (Wang et al. 2016; Su et al. 2015; Huang et al. 2014).

1.4.1 Microbiome of healthy subjects vs patients with COPD

Microbiome of patients with COPD has been found to be less diverse than from healthy people (Sze et al. 2015; Sze et al. 2012). The abundance of Proteobacteria and Bacteroidetes, in lung and oropharyngeal samples of patients with COPD, has been comparable with those from nonsmokers and healthy smokers but there is an increase of Firmicutes when compared with controls, especially in a more advanced stage of the disease (Park et al. 2014; Sze et al. 2012). In sum, these studies suggest that the lung microbiome of patients with COPD is different from that of healthy people, although the microbial diversity varies based on sample location and type. Additionally, there are other host factors that

are thought to change the microbiome, such as age, medication and disease severity, which have to be taken into consideration (Huang et al. 2017).

1.4.2 Microbiome in AECOPD

The role of bacteria in AECOPD requires a special attention because these acute crises are the main cause of morbidity and mortality in patients with COPD (Nazir & Erbland 2009). Due to that, it has been largely studied and it is believed that bacteria play an important role in the etiology of an exacerbation (Huang et al. 2017).

Through culture-based methods from the airways of patients with COPD, some authors have demonstrated that exacerbations are associated with pathogens such as *Streptococcus pneumoniae*, *Haemophilus influenza*, *Moraxella catarrhalis*, *Pseudomonas aeruginosa*, and *Staphylococcus aureus* (Soler et al. 2007; Hirschmann 2000). On the other hand, since the advent of culture-independent methods, several studies highlighted a higher respiratory microbial complexity during these acute events (Liu et al. 2017; Wang et al. 2016; Millares et al. 2015; Su et al. 2015; Huang et al. 2014; Millares et al. 2014; Huang et al. 2010). A table summarizing this literature findings can be consulted at Appendix 2.

Huang and colleagues analyzed, through 16S rRNA PhyloChip microarray, endotracheal aspirate samples of 8 patients at exacerbation. They identified a core of 75 bacterial taxa present in all samples collected during an acute event of the disease. Additionally, from these some of them, like *Brevundimonas diminuta*, *Arcobacter cryaerphilus* and *Leptospira interrogans*, had never been associated with exacerbation before (Huang et al. 2010).

Millares et al. analyzed, through pyrosequencing of the 16S rRNA, of sputum samples from patients with COPD at the baseline and during exacerbation have shown that during exacerbation bacteria typically associated with these events, such as *Streptococcus*, *Haemophilus*, *Pseudomonas*, *Neisseria* and *Moraxella*, were increased in relative abundance (Millares et al. 2014). Huang and colleagues examined the 16S rRNA gene, using PhyloChip, of sputum samples, before and during an exacerbation and also found that there was an increase of the Proteobacteria phylum during exacerbation (Huang et al. 2014). Also Wang et al.

found an increased proportion of Proteobacteria, with a decrease of microbial diversity during exacerbation, when compared with stable states. They also found a proliferation of *M. catarrhalis* in some individuals, which has been associated with increased airway inflammation during AECOPD (Wang et al. 2016; Parameswaran et al. 2009). These results suggest that even though there is an overall alteration in microbiome during an exacerbation, there are some individuals that appear to be more susceptible to shifts during this period than others (Wang et al. 2016).

Finally, the microbiome as a whole may not be significantly modified by an exacerbation in patients with COPD, as shown by Millares and colleagues (2015). When compared with stable state most of the bacteria do not suffer significant changes in their relative abundances during an AECOPD (Millares et al. 2015).

In addition to the studies that compared the microbiome before and during exacerbation, there are also some that focus on the microbial shifts after an AECOPD. When comparing the microbiome composition during and after an exacerbation, Liu et al. (2017), found, in oropharyngeal and sputum samples, increased levels of *Psychrobacter*, *Lactobacillus*, *Rothia*, *Prevotella*, *Neisseria*, *Streptococcus*, *Haemophilus*, *Actinomyces*, *Leptotrichia*, and *Aspergillus*, at this acute event of the disease.

Some authors have found that after treatment there is a major microbiome shift, and this shift may be influenced by the treatment prescribed (Wang et al. 2016; Huang et al. 2014). This topic will be more deeply analyzed in the next section.

When comparing samples collected at exacerbation and 7 to 16 days post-exacerbation, Su et al. (2015), concluded that despite some common variations in some phyla such as Firmicutes, Proteobacteria and Fusobacteria, the greater variation was found between patients, contributing to patterns almost always personalized to each patient.

Additionally, two different types of AECOPD have been described, the “bacterial” and the “eosinophilic” exacerbation. Bacterial exacerbations are characterized by the presence of bacterial pathogens in the sample, frequently sputum, accessed by a culture, while eosinophilic exacerbations are characterized by the presence of more than 3% of non-squamous cells in the sample (Bafadhel et al. 2011).

When Wang and colleagues studied the “bacterial exacerbation”, they found an increase in Proteobacteria, where in the eosinophilic there was an increase in Firmicutes (Wang et al. 2016). Both phenotypes also responded in different ways to treatment, i.e. the bacterial phenotype responds better to antibiotics, while the eosinophilic responds better to corticosteroids (Gomez & Chanez 2016).

All these results suggest that there is a significant heterogeneity among patients with COPD not just on the microbial diversity during an AECOPD (Su et al. 2015) but also on the changes of the microbiome composition during the stable state versus the exacerbation, and during exacerbation versus after treatment (Wang et al. 2016; Su et al. 2015; Huang et al. 2014).

1.4.3 Microbiome after therapy

The use of different drugs on patients with COPD also interferes with the composition of the lung microbiome (Gomez & Chanez 2016).

Huang and colleagues, (2014), and later Wang et al., (2016), showed that the sputum microbiome evolves differently after an exacerbation depending on the therapy prescribed. Treatment with antibiotics reduces the abundance of certain bacteria, mainly Proteobacteria, while treatment with only corticosteroids leads to an enrichment of these bacteria, as well as Bacteroidetes and Firmicutes members, while induces a decrease in diversity (Wang et al. 2016; Huang et al. 2014). It can be speculated that the treatment prescribed could, over months or years, lead to an alteration in the airway microbiome composition of the patient, that can contribute to a raise in the frequency of exacerbations and a worsening of the disease (Huang et al. 2017).

A study conducted by Slater and colleagues, in 5 patients with moderate-to-severe asthma, on the effects of therapy with azithromycin, demonstrated that therapy with this antibiotic was associated with decreased bacterial richness and altered airway microbiota, as well as a reduction in the amount of *Pseudomonas*, *Haemophilus* and *Staphylococcus* (the three more frequent genera associated with airway disease). Further studies must be conducted to confirm these results (Slater et al. 2014), along with studies comprehending patients with COPD.

1.4.4 The potential role of oral microbiome in COPD

The role of the oral microbiome in pathologies such as COPD is still poorly understood, although it is known that the oral cavity is an important reservoir for pathogens such as *Staphylococcus aureus*, *Pseudomonas aeruginosa*, and some *Enterobacteriaceae* (Amaral et al. 2009). Pragman and colleagues suggested that the lung microbiome appears to reflect microaspiration of oral flora in individuals with moderate and severe COPD suggesting that the lung microbiome is strongly influenced by the oral microbiota (Pragman et al. 2012). These results are in accordance with those from healthy individuals (Bassis et al. 2015; Morris et al. 2013; Charlson et al. 2011).

P. aeruginosa has been detected in both saliva and sputum samples from patients with cystic fibrosis. Furthermore, the clone was the same in both samples, suggesting an ascending or descending passage of bacteria between the oral cavity and lungs (Caldas et al. 2015).

Although there is still a lack of studies comprehending the saliva microbiome and its association with systemic diseases, there are already some studies that confirm that alterations in the salivary microbiome are inherent to some oral and non-oral diseases (Acharya et al. 2017). Adding this to what was mentioned above, further studies evaluating the role of saliva microbiome in COPD must be conducted.

1.5 Aims of the study

In general, the studies presented above show very heterogeneous results suggesting that, despite the fact that this field had a great development in the past years, there is still much to be clarified (Park et al. 2014), and new hypothesis to test. For example, one needs to clarify the true utility of using saliva as a possible biomarker source for monitoring patients with COPD, disease status and progression. Since 1) saliva is one of the most easily collectable human fluid, 2) it has been used in the past to detect non-respiratory pathologies, and 3) has been shown to be a *reservoir* of bacteria traditionally associated with lung microbiome, it seems a very promising vehicle for COPD-related information.

This Master thesis aims to contribute to this effort, by implementing the methodology to study the saliva microbiome in patients suffering from Chronic Obstructive Pulmonary Disease, as a way to understand the dynamics of saliva microbiome in the setting of an exacerbation and how the microbiome evolve after that.

The specific objectives were:

- To carry out a literature survey that would allow to establish major microbiome features to focus in the subsequent study;
- To establish protocols for the saliva collection, conservation, transport and laboratorial treatment to reach good overall sample quality;
- To characterize the saliva microbiome of 7 different patients with COPD, using a culture-independent method, based on NGS sequencing of 16S rRNA gene amplicons and bioinformatics analyses;
- To study the microbiome shifts on a total of 17 samples, that were collected in different time points of the disease, i.e., at the exacerbation onset, two weeks after the exacerbation, and after full clinical recovery (6 weeks and 8 months after exacerbation);
- To conclude, based in the results, if the overall methodology is able to detect microbiome fluctuations relevant for COPD management, to be used in future larger projects.

Chapter II – Methods

2.1 Subjects and sample collection

All the samples were collected from male patients with COPD, which have been enrolled under the scope of “Genial – Marcadores Genéticos e Clinicos na Trajectória da DPOC”, funded by Programa Operacional de Competitividade e Internacionalização - COMPETE, through Fundo Europeu de Desenvolvimento Regional - FEDER (POCI-01-0145-FEDER-016701), Fundação para a Ciência e Tecnologia (PTDC/DTPPIC/2284/2014) and under the project UID/BIM/04501/2013.

Individuals were classified into four distinct stages according to the GOLD guidelines (GOLD 2017). All participants were informed about all stages and conditions of the experiment before they gave their written informed consent, (Appendix 3).

We took advantage of this cohort, since it has been clinically characterized. All the characteristics, including lung function, as well as muscular and functional tests and patient reported outcomes, can be found in Appendix 4.

Sociodemographic and clinical data were collected, including smoking habits, medical antecedents, and treatments (table 2). Participants’ lung function was accessed following the guidelines. Forced Expirometry Volume in 1 second (FEV_1) was registered, as shown in table 2.

Table 2 Clinical characteristics of patients*

Subject	Age (yr)	FEV1 (%)	GOLD	Smoking status (pack-yr)	Treatment for Exacerbation
519	75	61	1	Former (30)	Azithromycin
526	78	31	3	Former (108)	Methylprednisolone
555	76	56	2	Former (60)	Levofloxacin
S29	58	18	4	Former (60)	Azithromycin
E44	73	60	2	Never smoke	None
587	77	84	1	Former (1)	None
616	54	44	3	Former (38)	Azithromycin

*All subjects were male. Age, lung function, GOLD classification and smoking status were collected during the time of enrollment in this study; FEV1, forced expirometry volume in 1s; pack-yr, number of packs of cigarettes smoked per day multiplied by the number of years that the person has smoked.

17 saliva samples were collected for microbiome assessment from 7 different individuals, at different time points. Samples were collected during exacerbation, before the administration of antibiotic therapy, 2 weeks post-therapy and/or at clinical recovery (6 weeks – recovery, and 8 months after exacerbation – recovery_B), as depicted in table 3. The time points of sample collection were chosen based on the fact that after 4 to 6 weeks of an exacerbation a patient is clinically considered to be fully recovered (Woodhead et al. 2011). For some patients, it was also possible to obtain samples collected after 2 weeks of exacerbation. These samples would allow us to deeply characterize the microbial changes after an exacerbation and subsequent treatment. For one patient (subject 526), it was also possible to collect a sample 8 months after the exacerbation, which would allow to compare the microbiome composition in two different time points of clinical recovery.

Table 3 – Time points of sample collection per subject, at exacerbation, 2 weeks after exacerbation, 6 weeks after exacerbation ”Recovery” and 8 months after exacerbation “Recovery_B”

Subject	Exacerbation	Post 2 weeks	Recovery	Recovery_B
519	X	X	X	
587	X	X	X	
526	X		X	X
555	X	X		
E44	X		X	
616	X		X	
S29	X		X	

During sample collection, subjects were asked to spit into sterile falcons (F50). Specific precautions were taken during this procedure to avoid contamination of the samples, as described in (Goode et al. 2014). All subjects were advised to avoid ingestion of any acidic substances, aliments rich in sugar or in caffeine, at least one hour before data collection since these can cause changes in the saliva pH and lead to changes in bacterial composition (Zaura et al. 2017). Patients were also advised to mouthwash with water 10 minutes before the sample collection to minimize the presence of food particles and any modification in saliva pH. Once finished the collection, the falcons were properly labelled, transported in portable

freezers to the Institute for Research in Biomedicine (iBiMED) where they were preserved in freezers at -80ºC, until further processing, which occurred within 8 months interval.

2.2 DNA Extraction and Quantification

Bacterial DNA was extracted from saliva samples using the QIAamp DNA Microbiome kit (QIAGEN, Valencia, CA, USA) according to the manufacture's protocol – Depletion of host DNA (QIAGEN® 2014)

First, host cells were lysed by incubation with Buffer AHL, followed by a centrifugation, which separated the bacterial cells from the host cellular debris (including human nucleic acids). After that, an incubation with Benzonase and Proteinase K degraded the remaining human nucleic acids, avoiding them to be major contaminants of the bacterial DNA extract. Then, efficient lysis of Gram-negative and Gram-positive bacteria was assured, due to an incubation with detergent-containing buffer ATL and a further step of bead-beating. Afterwards the bacterial DNA was purified through adsorption to silica membranes present in columns from the Qiagen kit. DNA-containing silica membranes were washed with two washing buffers to ensure complete removal of residual contaminants. Purified bacterial DNA was eluted, and was ready for subsequent handling.

Before starting all the extractions, several tests were done to optimize the protocol, so that DNA extraction yield amounts were above 50 ng/µl (as given by DeNovix DS-11 FX+ quantification). From these preliminary testing the following modifications resulted:

- Due to the limited quantity of saliva collected (4 of the 17 samples did not have 1 ml of saliva), the protocol was then adapted to process 500 µl of sample, by adjusting the volume of Buffer AHL to the half, as suggested by the kit protocol. A table including all the sample volumes can be found in Appendix 5.
- Instead of an incubation of 30 minutes of the Buffer AHL, samples were incubated for 1 hour, to assure a more efficient lysis of host cells.

- For lysing bacterial cells, a Mixer Mill MM 400 by Retsch® was used. We applied a frequency of 30 Hz for 5 minutes, followed by 5 minutes in ice, and another step of 5 minutes at 30 Hz, based on the protocol described by Yamagishi et al. (2016).
- At the drying step, instead of a centrifugation at 20000 x g for 1 minute, we performed a centrifugation for 2 minutes. This step aims to dry off the membrane before eluting the DNA. By raising the duration of the centrifugation, one decreases the chance of ethanol contamination.
- All samples were eluted using Milli-Q water, since no prolonged storing was predicted for these samples.

DNA was quantified using the DeNovix DS-11 FX+ - a spectrophotometric method. Due to the literature-reported frequent overestimation of DNA quantity that is given by a spectrophotometric method (Simbolo et al. 2013), we also used a fluorometric method - Qubit® dsDNA HS Assay Kits. Only samples with a DNA yield > 10 ng/µl, given by Qubit were used for subsequent analysis. Only one sample was below this threshold and it was left from sequencing, due to the impossibility of doing additional extractions since that sample had a limited quantity of saliva. A table including all the samples from which DNA was extracted and the correspondent concentration can be found in Appendix 5.

Eluted DNA was then stored at -20ºC until further analysis, which happened within 45 days.

2.3 PCR amplification and amplicon sequencing

PCR amplification, library building and fragment sequencing was performed at Instituto Gulbenkian de Ciência (IGC), Oeiras, Portugal. The hypervariable V4 of the 16S rRNA gene was amplified following the protocol described in Walters et al., (2016). The V4 region was amplified with the modified primers 515f (5' – GTGYCAGCMGCCGCGGTAA – 3') and 806r (5' – GGACTACNVGGGTWTCTAAT – 3') (Parada et al. 2016; Apprill et al. 2015). These primers are modified versions of the original ones proposed by Caporaso et al., (2011), which were found to introduce biases against SAR11 clade, and two important environmental Archaea (Walters et al. 2016). PCR was performed in a final volume of 25 μl containing 13 μl of PCR-grade water, 10 μl of 5' Hot master mix, 0,5 μl of forward primer (10 μM), 1,0 μl of reverse primer (10 μM), and 1,0 μl of template DNA. The following PCR conditions were used: initial denaturation for 3 min at 94ºC; followed by 35 cycles of 45 sec at 94ºC, 60 sec at 50ºC, and 90 sec at 72ºC; with a final extension of 10 min at 72ºC, and finally a 4ºC hold.

Amplicons were then used to build Illumina libraries and sequenced using Illumina MiSeq System, as described in the manufacture's protocols, Illumina (2015). The sequencer uses a reversible terminator sequencing by synthesis (SBS) approach - using four fluorescently labelled nucleotides in a massive parallel environment (Illumina 2015b). The overview of this process is synthetized in figure 4. Briefly,

- The protocol started with a library preparation, where specific adapters were added to the 3' and 5' ends (Hodkinson & Grice 2015; Illumina 2015a) (figure 4a).
- The library was then loaded into a flow cell where fragments were captured. Once attached to the flow cell, the single stranded DNA fragments underwent bridge amplification, resulting in a cluster of identical fragments (Lu et al. 2016; Illumina 2015a; McElhoe et al. 2014). When cluster generation finished, the templates were ready for sequencing (figure 4b).

- During each sequencing cycle, labelled deoxynucleoside triphosphate (dNTP) was automatically added to the nucleic chain. The label works as a terminator for polymerization, so after each dNTP incorporation, the fluorescent dye can be imaged to identify the base and then cleaved to allow the incorporation of the next nucleotide (Illumina 2015a). Since each incorporated base is accessed separately, the sequencing errors are significantly reduced (McElhoe et al. 2014) (figure 4c).

The final outputs of the sequencing platform were fastq format files which were then used for bioinformatics sequence analysis (Hodkinson & Grice 2015). The fastq format comprehends the nucleotide sequencing with its corresponding quality scores (Deorowicz & Grabowski 2011).

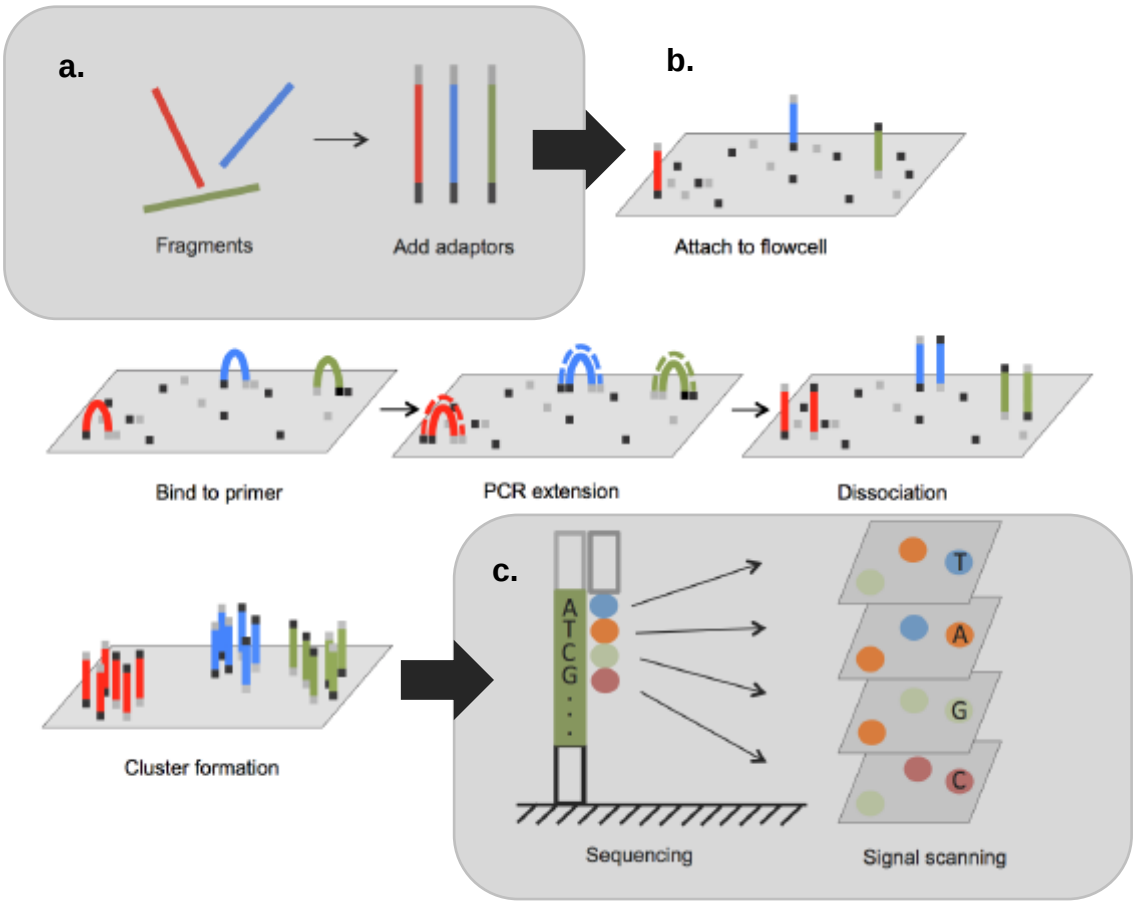


Figure 4 - MiSeq Illumina sequencing workflow – **a.** Library preparation, **b.** Cluster amplification, and **c**. Sequencing, from Lu et al. (2016).

2.4 Sequence analysis

The primary sequence analysis was also performed at IGC. These analyses were done with Quantitative Insights Into Microbial Ecology 1.9.1 (QIIME 1.9.1) (J. Gregory Caporaso et al. 2010), an open source pipeline that can perform standard microbial community analysis, starting with raw sequence data and using statistical analysis and visualization steps (Navas-Molina et al. 2013; Caporaso et al. 2011).

QIIME analysis started with fastq files and a user-generated mapping file, which contained all required information about each sample. This included the sample ID, the state (exacerbation, post 2 week, and recovery – 6 weeks and 8 months after exacerbation) at which the patient was at the time of sample collection, the usage or not of antibiotics for treating the exacerbation, the amount of saliva collected, and amount of DNA extracted.

The analysis required some preprocessing steps. The first was the demultiplexing step which consists of removing all barcodes and primer sequences and assigning reads to each sample, so that only the sequences matching the amplified 16S rRNA gene for each sample remain (Navas-Molina et al. 2013). Default parameters were applied to this step.

The quality-filtering performed next removed the sequence reads that did not reach a minimal quality score. Like most sequencing instruments, Illumina instruments generate a quality score for each nucleotide (Phred scale, (Ewing et al. 1998)), related to the probability of incorrect base calling. To filter reads by quality, QIIME uses Phred score and user-defined parameters, in this case default parameters were used, which were r = 3, p = 75%, q = 3, and n = 0 (r – the maximum number of consecutive low-quality base calls; p – the percentage of consecutive high-quality base calls; q – the minimum Phred quality score; n – the maximum number of ambiguous bases) (Bokulich et al. 2013).

The next part of the QIIME pipeline was clustering the sequence of reads into OTUs. Sequences were clustered together according to a threshold of sequence similarity, set at 97%. This value has been conventionally assumed to separate bacterial species, when using 16S rRNA sequencing (Drancourt et al. 2000). QIIME supports three approaches for OTU picking (*de novo*, closed-reference, and open-reference). The *de novo* approach clusters sequences based on sequence identity,

the closed-reference matches sequences to a database of reference sequences, while the open reference approach combines the *de novo* and the closed-reference approach (Navas-Molina et al. 2013).

Our collaborators at IGC used the open-reference approach. It started with the sequences being compared directly with a reference sequence collection (in this case GreenGenes v13_8 database), as with the closed-reference approach, but then, the sequences which had no hit were clustered into OTUs by the *de novo* method - sequences were compared to each other and then clusters were formed (Navas-Molina et al. 2013). The algorithms used for the OTU picking method were, blast, for the closed reference (Altschul et al. 1990) and uclust (Edgar 2010), for the *de novo* approach.

The next step was to assign the taxonomy to each sequence of the representative set, in other words, this step connected the OTUs to named organisms. Due to the use of a partial *de novo* approach, some clusters were not named according to any reference database, and so, in those cases, the taxonomy must be assigned using a reference dataset (Navas-Molina et al. 2013). In this study GreenGenes (v13.8) database was used. To assign taxonomy against the database, uclust was used.

The sequence alignment, needed for this step was performed by PyNAST (Caporaso et al., 2010) using as reference the GreenGenes core set (DeSantis & Hugenholtz 2006) as explained.

The construction of a phylogenetic tree was the next step after assigning to each cluster the right taxonomical identity. The output was the taxa that were present in each sample, which can be used for diversity analyses and to understand the relationship among the sequences in a single sample. For that, FastTree was used (Price et al. 2009).

Finally, the last step of this pipeline was the construction of the OTU table. The OTU table is a matrix that shows the abundance of each OTU in each sample, and includes also the taxonomic prediction for each OTU. The output was a compact file in Biological Observation Matrix (BIOM) format (Mcdonald et al. 2012).

The workflow about the sequence analysis done in this study is summarized in figure 5.

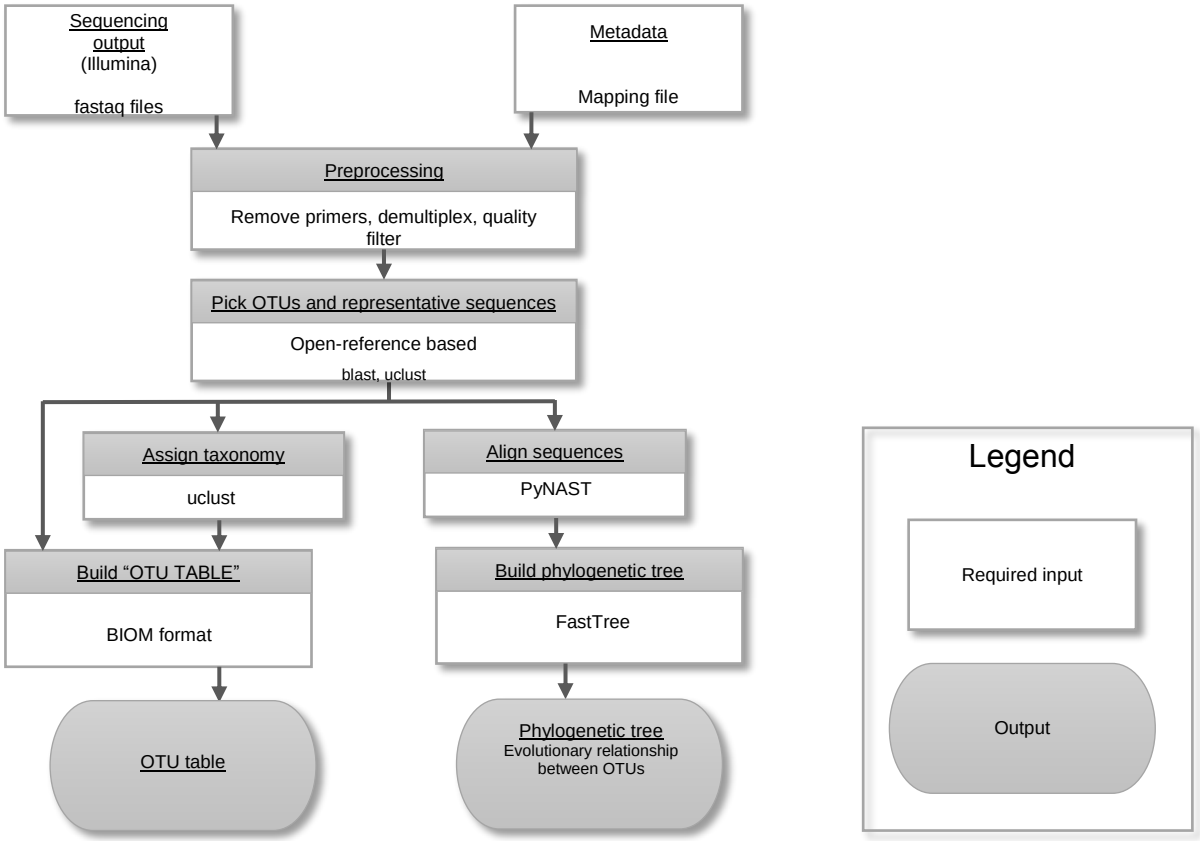


Figure 5 - QIIME workflow overview. This analysis starts with preprocessing the sequencing reads and ends with the OTU table and the phylogenetic tree (output). The white box represents the required input files, while the grey ones represent the output files given by QIIME. The grey and white boxes, show the different steps (in grey), and the methods used for each step (in white) (Adapted from: Navas-Molina et al., 2013).

2.5 Statistical analysis and visualization

Before any analyses, sample depth had to be determined, since it defines the size of the random subset of sequences that were selected for each sample for all subsequent analyses (Navas-Molina et al. 2013). Based on our data, the reasonable rarefaction level was 10342 sequences per sample, which was the number of sequences found in the sample that had less sequences. This normalization step is important because most of the diversity analyses used are very sensitive to the number of sequences used by sample, which has to be homogenized (Kuczynski et al. 2012).

For an overview analysis of the family and genus taxa a phylogenetic tree was created using the webtool *phyloT* (Letunic 2015). PhyloT uses the NCBI taxonomy database to generate a phylogenetic tree based on a list of taxa given by the user.

To study the diversity of a microbial community, one can look at the alpha, and beta-diversities. Alpha diversity is defined as the diversity of organisms in one sample or group of samples, on the other hand, beta-diversity studies the difference in diversities across samples, or groups of samples (Navas-Molina et al. 2013).

When studying the alpha-diversity, we were interested in estimating the community richness i.e. the number of different OTU's within a sample; as well as the community diversity, which englobes the number of OTU's and how these OTU's are distributed (evenness) (Morgan & Huttenhower 2012).

To study the alpha-diversity, we used the observed number of OTUs and the Chao1 estimator (Chao 1984) – the Chao1 estimator is a metric that predicts the total number of OTU's that would be seen with an infinite number of samples. To study the community diversity, we used the Shannon (Shannon & Weaver 1963) and Simpson indexes (Simpson 1949), which are robust estimators of diversity (Haegeman et al. 2013). Simpson index measures the probability of two individuals arbitrarily selected from a sample to belong to the same species, while Shannon index is an entropy measurement which increases with the number of species in the sample (Escobar-Zepeda et al. 2015). The difference between the two statistical distributions is that the Shannon index values more the rare species in the sample,

while Simpson index values more the common species (Krebs 2014). Here, we calculated alpha-diversity for 1) each sampling time point (exacerbation, pos2 and recovery), 2) each patient (samples grouped by patient) and 3) each sample collected per patient.

To access the beta-diversity, the first Bray-Curtis dissimilarity metric (Bray & Curtis 1957) was used, which quantifies the compositional dissimilarity between the groups based on the number of counts per sample. Then to evaluate if these dissimilarities were statistically significant, a permutational multivariate analysis of variance – PERMANOVA (Anderson 2001) (with 999 permutations) was used. This test was used to test the heterogeneity of community structure in a priori groups i.e. age, antibiotic treatment, smoker status, sampling time points, patients).

Other beta-diversity measures were also used, such as the UniFrac (Lozupone & Knight 2005), which uses phylogenetic information to compare samples. For that, all the taxa found in the samples, were placed on a phylogenetic tree, then the amount of unique evolution within each community is measured and compared to another, by calculating the fraction of branch length of the phylogenetic tree that is unique to either one of the pair of communities (Lozupone & Knight 2005; Navas-Molina et al. 2013). We used the unweighted UniFrac, a qualitative measure which has provided results that correlate better with clinical variables than weighted UniFrac (Navas-Molina et al. 2013). The UniFrac measure can be coupled with standard multivariable statistics such as a dendogram (Tyrion 1939) or Principal Coordinates Analysis (PCoA) (Mardia et al. 1979). We used unweighted UniFrac measure combined with a dendogram (Tyrion 1939) to identify potential clusters of samples, as well as a PCoA (Mardia et al. 1979), to visualize phylogenetical dissimilarities between samples.

All the alpha and beta–diversities measures were performed using the R statistical software version 3.4.2 (R Core Team 2015) and the RStudio, version 1.1.383 (RStudio Team 2016), using the packages *vegan* (Oksanen et al. 2017) and *phyloseq* (McMurdie & Holmes 2013).

Considering the results given by the previous tests, STAMP v2.0.0 (Parks et al. 2014) was used to analyze the three time points at which samples were collected,

and to look for statistically significant shifts in the relative abundance of certain taxa, through ANOVA with Games-Howell post hoc testing (Bluman 2007).

There was also a need to visualize the microbiome shifts over time. Samples were grouped by time of collection and plots were built showing the overall taxa relative abundance *per* group of samples. The same was also performed for each subject, but instead of grouping the samples, they were displayed as a way to compare the different time points of sample collection, and presented as plots showing the relative abundances *per* sample.

Chapter III – Results

3.1 Participants

For each subject a single sample was obtained at the onset of the exacerbation and always before new treatments were started ("Exacerbation"). Additionally, one or two samples were collected from the same patient, either "Pos 2", taken 2 weeks after the exacerbation date, "Recovery" – after 6 weeks , or "Recovery_B" – 8 months after exacerbation. All patients were male, mean age of 70,1 years, and most had moderately severe airflow obstruction (GOLD 2 - FEV1 between 50 and 80% (GOLD 2017)). All, except one that never smoked, were former smokers. The medication prescribed, or not, for the treatment of the exacerbation, as well as other characteristics were described earlier on *Chapter II* (see table 2 and 3).

3.2 Sampling and Sequencing

Bacterial DNA was successfully extracted, amplified and sequenced from 17 samples of saliva. After the preprocessing steps, a total of 393 826 reads remained.

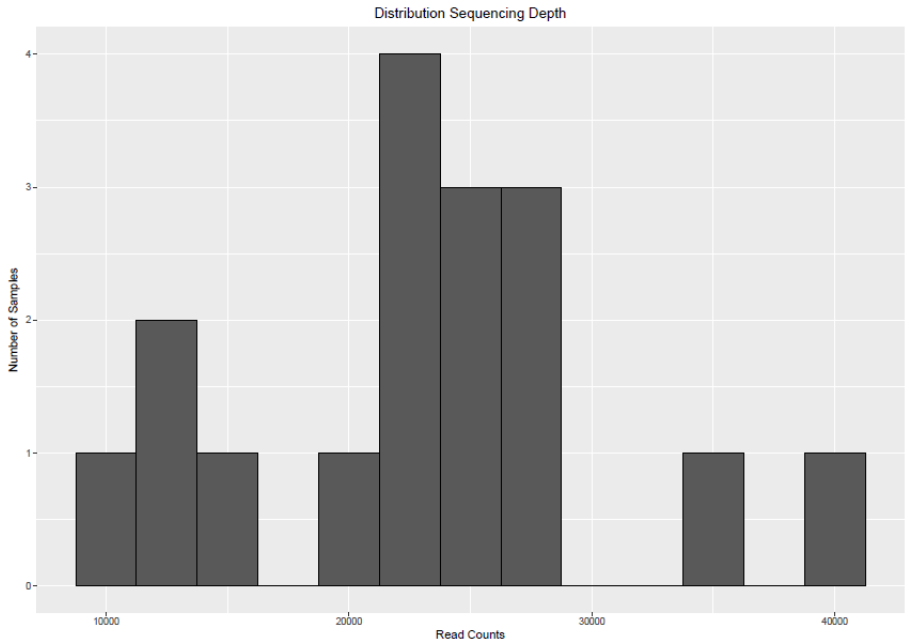


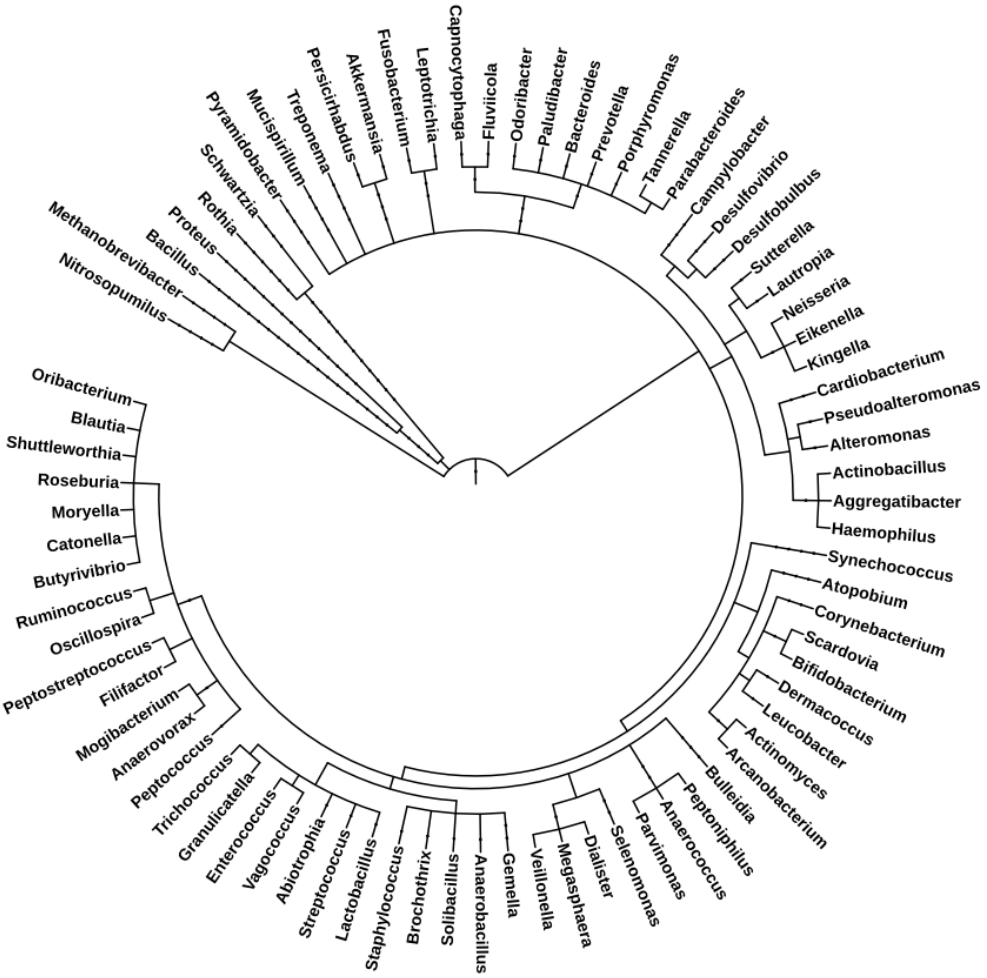
Figure 6 - Distribution of sequencing depth – number of reads per number of samples. Most of our samples had between 2000 and 3000 read counts.

The distribution of reads was between 10 342 and 40 333 per sample, distributed as shown in figure 6.

The bacterial sequences were then normalized to 10 342 randomly chosen reads and clustered by OTUs similarity. The number of OTUs observed at 97% identity was 3848. During the analysis some of the OTUs were found to be “Unassigned” for the Kingdom rank, to avoid further misinterpretations, those OTUs with this classification were removed, leading to final count of 3574 OTUs.

3.3 Overall microbial community

Among all samples 3574 different OTUs were found, 4 belonging to Archaea and 3570 to the Bacteria Kingdom. The later was represented by several species belonging to distinct families (n=70) and genera (n=88), as shown in figure 7.



* The genus *Pseudoramibacter* and BD2-13 strain were also detected in the samples although they are not included in this tree, because *PhyloT* could not recognize them.

Figure 7 - Phylogenetic tree showing the overall bacterial diversity at the genera* level that was detected in our cohort of saliva samples from patients with COPD.

From the total amount of genera detected one can deduce a “core microbiome”, i.e., a group of taxa that were found in all samples collected. This group was formed by *Actinobacillus*, *Akkermansia*, *Bacteroides*, *Fusobacterium*, *Gemella*, *Haemophilus*, *Leptotrichia, Parviromonas*, *Prevotella*, *Rothia*,

Streptococcus, *Veillonella*, members of the family *Bacteroidales*, *Lactobacillalles* and *Lachospiraceae*, and taxa belonging to the Bacilli class.

Table 4 – Relative abundance presented across all samples, relative to phyla and genera levels. Taxa with higher relative abundances are highlighted. Firmicutes and Actinobacteria were the phyla detected with higher relative abundances.

Phyla		**Genera**	
14,3%	**Actinobacteria**	1,4%	*Actinomyces*
		0,2%	*Corynebacterium*
		12,3%	***Rothia***
		0,3%	*Atopobium*
5,1%	**Bacteriodetes**	0,7%	*Porphyromonas*
		1,9%	*Prevotella*
		2,2%	*Capnocytophaga*
		0,1%	Unclassified
0,2%	Cyanobacteria	0,2%	Unclassified
43,7%	**Firmicutes**	0,3%	*Bacillus*
		3,5%	*Gemella*
		0,2%	*Granulicatella*
		0,2%	*Lactobacillus*
		22,2%	***Streptococcus***
		0,2%	*Moryella*
		0,8%	*Oribacterium*
		0,1%	*Filifactor*
		0,1%	*Peptostreptococcus*
		0,1%	*Daliaster*
		0,2%	*Megasphaera*
		0,1%	*Schwartzia*
		0,2%	*Selenomonas*
		11,5%	***Veillonella***
		0,1%	*Mogibacterium*
		0,3%	*Parvimonas*
		0,3%	*Bulleidia*
		3,2%	Unclassified
9,1%	**Fusobacteria**	1,6%	*Fusobacterium*
		7,5%	***Leptotrichia***
27%	**Proteobacteria**	0,3%	*Lautropia*
		0,2%	*Kingella*
		16,1%	***Neisseria***
		0,3%	*Campylobacter*
		0,1%	*Cardiobacterium*
		0,3%	*Actinobacillus*
		0,6%	*Aggregatibacter*
		8,8%	***Haemophilus***
		0,3%	Unclassified
0,1%	Spirochaetes	0,1%	Treponema
0,1%	Synergistetes	0,1%	TG5
0,4%	Unassigned		

As it can be seen in table 4, the most common phyla detected among samples was Firmicutes (43,7%), Proteobacteria (27,0%), Actinobacteria (14,3%), Fusobacteria (9,1%), and Bacteroidetes (5,1%). Of the 88 genera identified, the most abundant were *Streptococcus* (22,2%), *Neisseria* (16,1%), *Rothia* (12,3%), *Veillonella* (11,5%), *Haemophilus* (8,8%), and *Leptotrichia* (7,5%).

Curiously, even though presented at a higher relative abundance, *Neisseria* was not included as a genus presented at the “core microbiome”, as can be seen in figure 8, *Neisseria* were just present in 16 out of the 17 samples (this genus was absent of the exacebertion sample from patient 616).

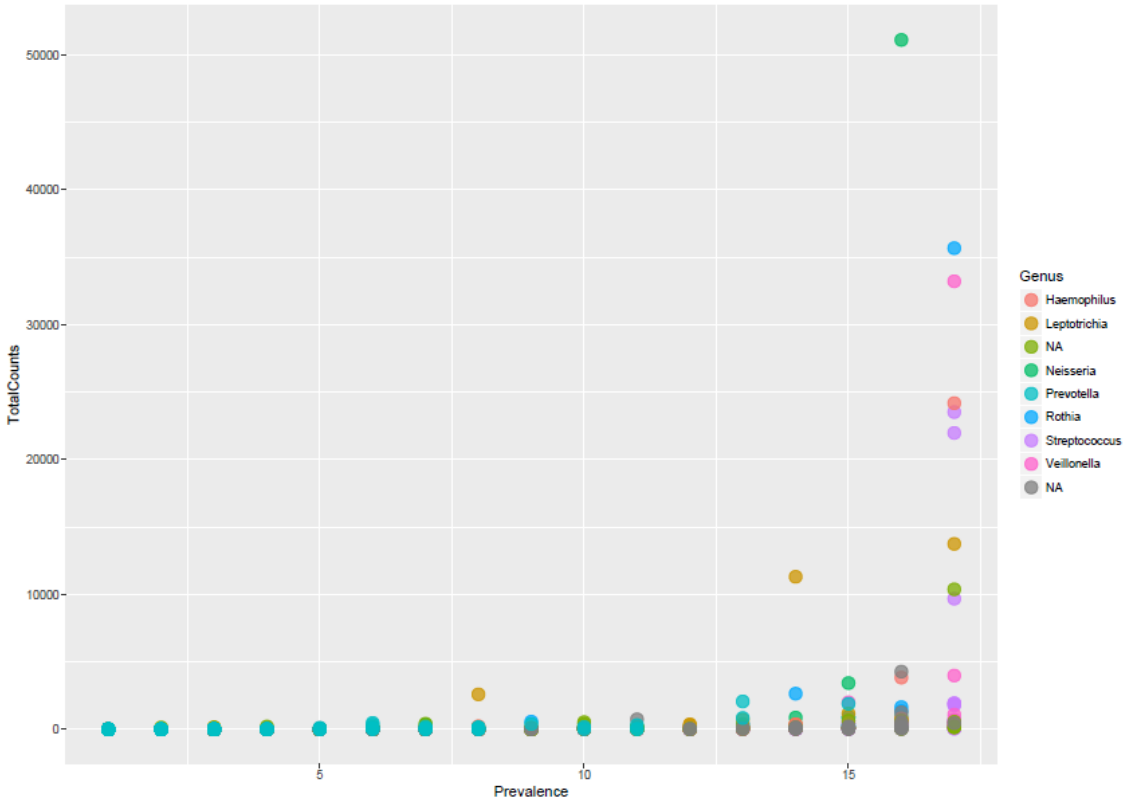


Figure 8 - Prevalence of the most abundant genera, per total of counts, in all samples (n=17). Neisseria had an overall high relative abundance, even though it was absent from one of the samples.

3.3.1 Alpha-diversity

Relatively to the observed bacterial richness (number of different OTU's) found in each sample. The number of OTU's per sample varies between 279 and 1516 taxa, being the average 887 taxa. Overall one can see that the bacterial richness was similar between the samples collected from the same subjects, although the bacterial richness among the different times of collection varies, being impossible to find a pattern. Nevertheless, there are two samples from all the cohort that stand out - the sample collected at exacerbation, from patient 616, that has the lowest richness and the sample collected at recovery (6 weeks after exacerbation) from patient 519 that has the highest richness from all cohort (figure 9).

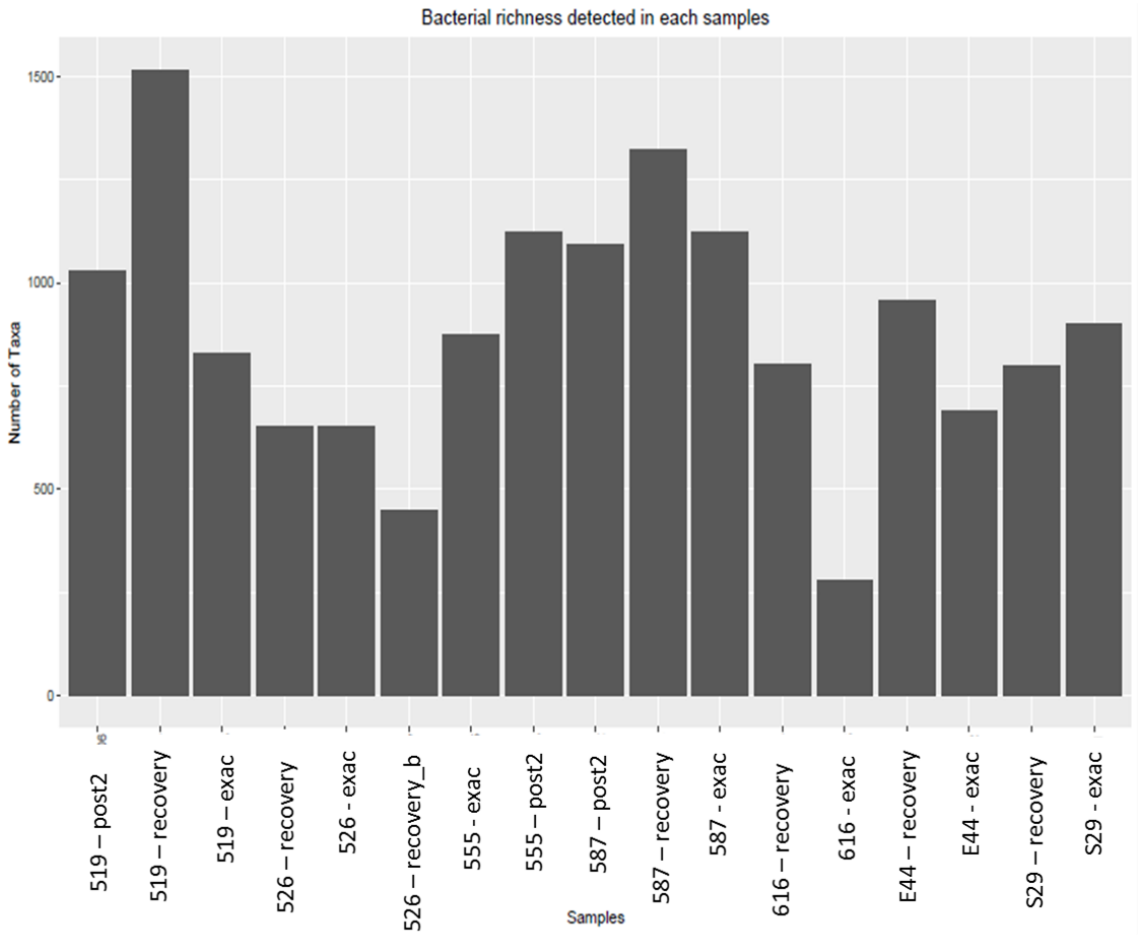


Figure 9 - Bacterial richness (number of taxa per sample) detected in each patient sample. "Exac" – exacerbation; "post2" – 2 weeks after exacerbation, "recovery" – recovery (6 weeks after exacerbation), "recovery_b" – recovery (8 months after exacerbation). From the figure, one can affirm that the sample from patient 616 collected at exacerbation has the lowest richness, while the sample "recovery" from patient 519 has the higher richness from all cohort.

3.3.2 Beta-diversity

To further evaluate the relationship between community composition and clinical patient characteristics, Bray-Curtis dissimilarity metric matrices were built (Bray & Curtis 1957) and then a permutational multivariate analysis of variance (PERMANOVA) was performed (Anderson 2001) in order to identify statistically significant dissimilarities in the groups. No significant dissimilarities (p-value > 0.05) were found between community composition and parameters such as age (PERMANOVA analysis, F = 1.29, R^2 = 0.26, p-value = 0,079), the antibiotic treatment (PERMANOVA analysis, F = 1.38, R^2 = 0.16, p-value = 0.169), regarding the smoker status (PERMANOVA analysis, F = 1.18, R^2 = 0.073, p-value = 0.313), and the GOLD stage (PERMANOVA analysis, F= 1.24, R^2= 0.22, p-value= 0.23). Additionally, the only factor of variability in all the technical procedure, from DNA extraction until data analysis, was the fact that 4 out of the 17 samples analyzed had only 500 μL of saliva, so we also test if this originated any dissimilarity between the two groups, again with no statistically significant differences (PERMANOVA analysis, F = 0.81, R^2 = 0.051, p-value = 0.585).

On the other hand, significant correlations (p-value<0,05) were found between the time points at which samples were collected (PERMANOVA analysis, F = 1.70, R^2 = 0.2813, p-value = 0.043), and among patients (PERMANOVA analysis, F = 2.23, R^2 = 0.573, p-value = 0.001). These results show that that there are significant differences in the community composition of samples when they are grouped by time of collection and by patient.

Furthermore, a cluster dendogram, using the unweighted UniFrac distance metric, was conducted (figure 10). This dendogram shows that samples were clustered based on the patient from which they were collected. Showing higher variability between patients than between disease states. With the exception of patient 616 and the “recovery” sample of patient 519. These results show that the two samples collected for patient 616 do not have a similar phylogenetic composition, as well as the “recovery” sample from patient 519 does not have a similar phylogenetic composition when compared with the other two samples collected from this patient.

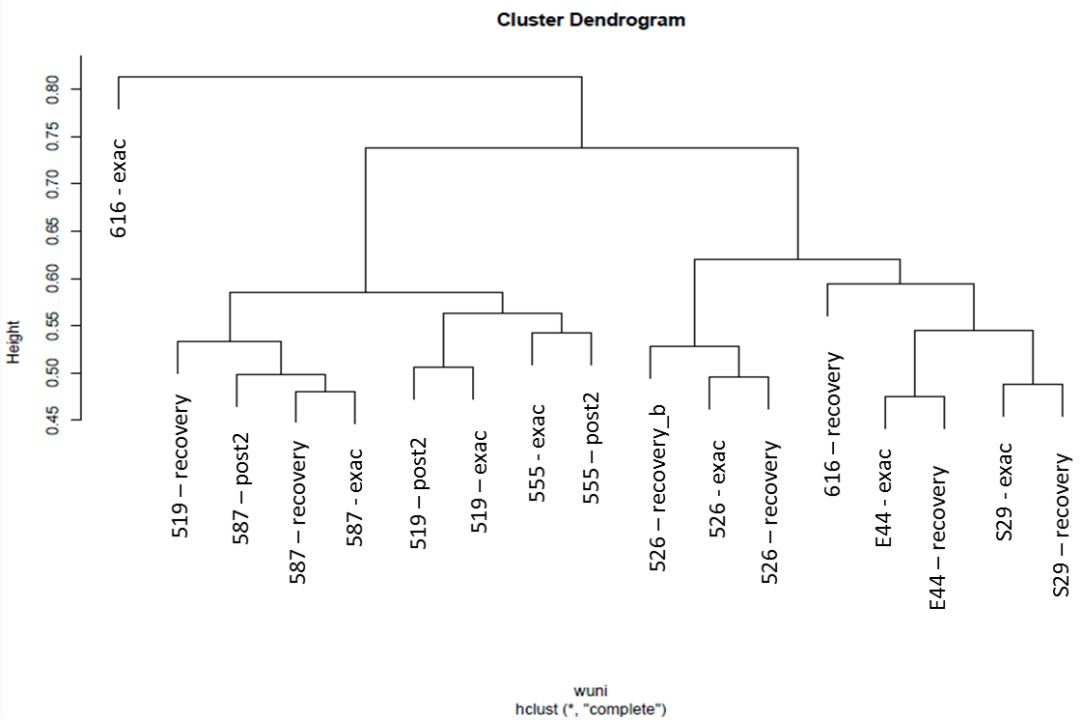


Figure 10 - Cluster dendogram with unweighted UniFrac distance metric of all samples collected. “exac” – exacerbation; “pos2” – two weeks after exacerbation; “recovery – 6 weeks after exacerbation; “recovery_B” – 8 months after exacerbation”. Note that samples became clustered by patient, showing higher variability between patients than between disease states for the same patient.

Additionally, an unweighted UniFrac coupled with Principal Coordinate analysis (PCoA) was performed to validate the previous results (figure 11). The cohort was composed by a very small number of samples (n=17), so changes in the community structure between samples at different time points were not observed. With exception of the fact that the sample collected 8 months after exacerbation (Recovery_B) from the patient 526, and the sample collected at recovery from patient 616, appear to have a community structure more different from the other samples collected.

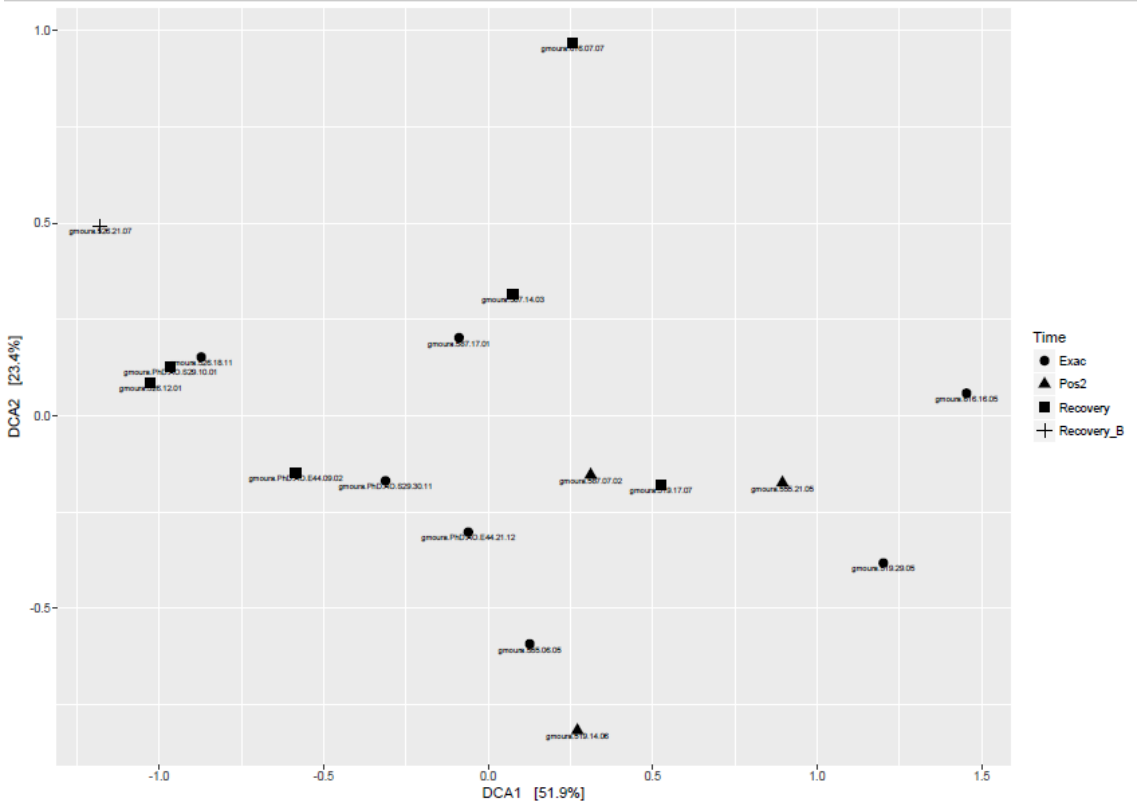


Figure 11 - Principal cordinate analysis (PCoA) showing the distribution per time of sample collection (Exac – exacerbation, Pos2 - after two weeks of exacerbation, Recovery – after 6 weeks of exacerbation, and Recovery_B – 8 months after exacerbation). Note that there is no degree of separation, whith exception for the sample collected at “Recovery_B” and one of the samples at “Recovery”.

3.4 Microbiome over time points

Given the results shown above, alpha-diversity was studied in groups of samples collected at the same time points, measured through richness (number of OTU's observed and Chao1), and diversity (Shannon and Simpson) as shown in figure 12. When comparing alpha-diversity, our results indicated an overall increase in richness and diversity in the samples collected after 2 weeks of exacerbation, contrary to exacerbation and recovery (6 weeks and 8 months after exacerbation). Interestingly, alpha-diversity at AECOPD and recovery was remarkably similar.

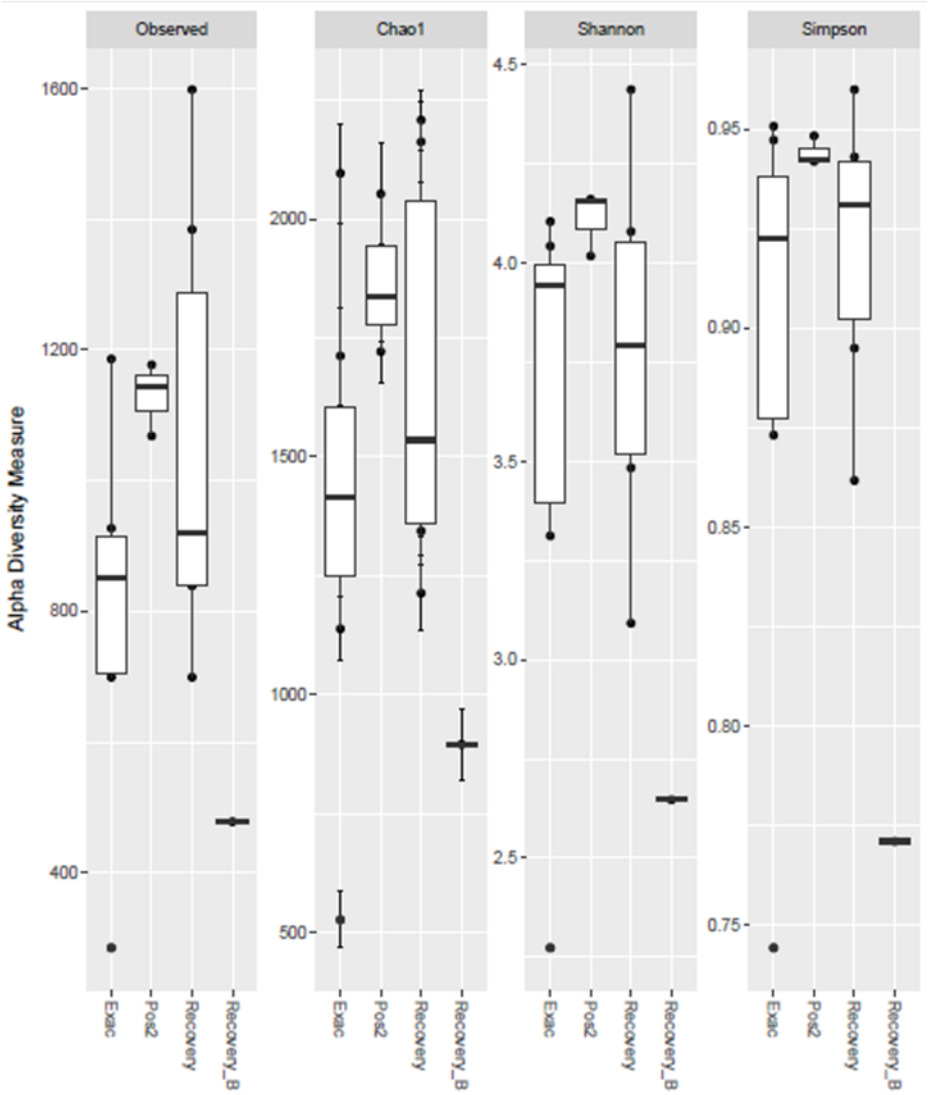


Figure 12 - Alpha diversity comparitions between time points (Exac – exacerbation; Pos2 – two weeks after exacerbation; Recovery – 6 weeks; Recovery_B - 8 months after exacerbation). Samples were agroupped by the time of collection, and bacterial richness (Observed, and Chao1) and diversity (Shannon and Simpson Index) were calculated. All indexes are increased at the middle time point, i.e. two weeks after exacerbarion, when compared to the other three sampling times.

Note that when comparing the bacterial diversity at exacerbation and at recovery, depending on the index used the results are different. For the Shannon index, which values more the rare species within the community (figure 12), one can see that the group of samples collected at exacerbation had more rare species that the samples collected at recovery.

Looking now at the taxonomical composition of samples, at exacerbation, the most common phylum was Firmicutes, followed by Proteobacteria, Actinobacteria, Fusobacteria and Bacteroidetes (see Appendix 6). At genera level, the most abundant genera were *Streptococcus*, *Neisseria*, *Rothia*, *Veillonella*, and *Haemophilus* (figure 13).

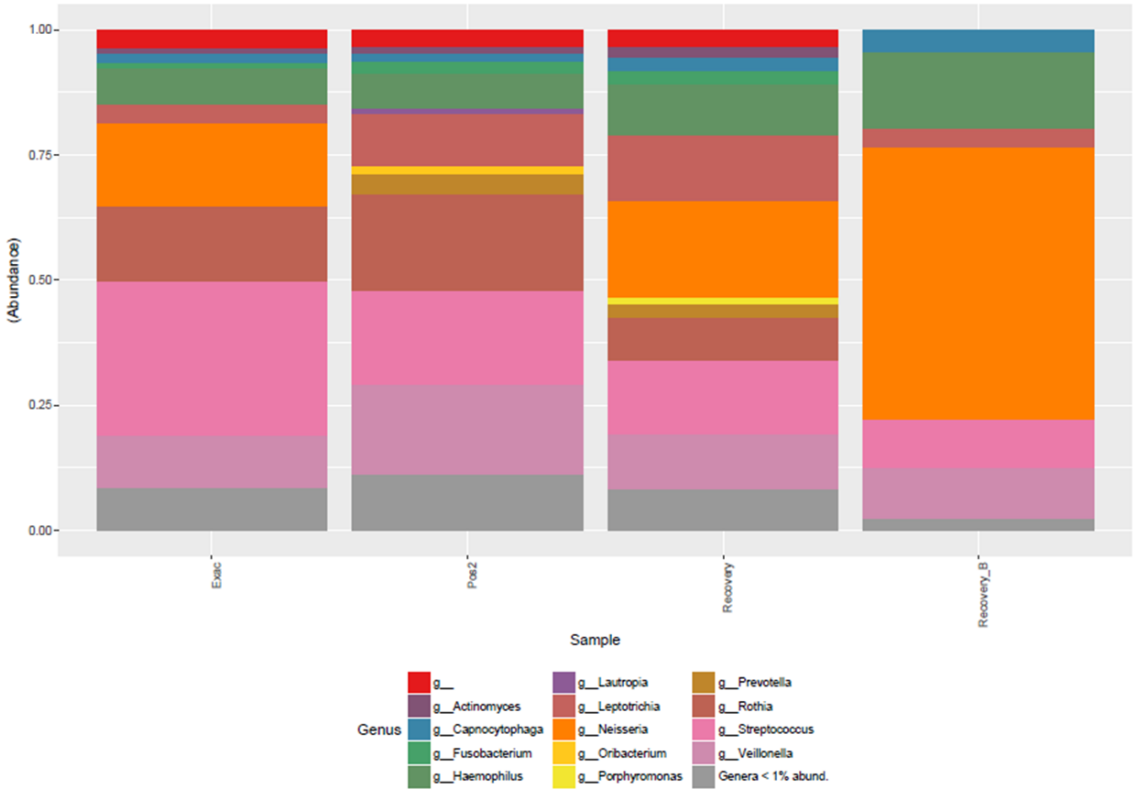


Figure 13 - Samples have been grouped and averaged by time of collection – exacerbation (Exac), 2 weeks after exacerbation (pos2), and 6 weeks (Recovery) and the sample collected 8 months post exacerbation (Recovery_B) -, and taxonomic composition is shown at the genera level. Each column in the plot represents a time of sample collection, and each color in the column represents the relative abundance of the total sample contributed by each taxon group at genera level. The relative abundance of *Streptococcus* is decreasing over the time. The taxonomical composition of "Recovery_B" is remarkably different from the samples collected at "Recovery".

From figure 13, one can see that there was a decrease in abundance of the genus *Streptococcus* over time, and the taxonomical composition of the samples collected at “recovery” (after 6 weeks of exacerbation) and at ”recovery_B” (8 months after exacerbation) are remarkably different. Since the “recovery_B” group was just composed of one sample, caution must be taken when extrapolating results. This topic will be further discussed when comparing the samples collected from subject 526.

In order to compare differences in the relative abundances at different time points we used ANOVA with the Games-Howel post-hoc test, to test if there were differences ($p < 0,05$) in the relative abundances of the most abundant taxa, at phyla and genera level. When comparing the three collection time points (we grouped all the samples collected at recovery, i.e “recovery” and “recovery_B”). There were no significantly differences in the relative abundances at phyla level. However, at genera level there was a decrease in the relative abundance of *Streptococcus* (p-value = 0.029, ANOVA with Games-Howel post- hoc test). There was also a significant shift in the relative abundance of *Rothia*, with an increase of this genus at the samples collected after 2 weeks (p-value = 0.047, ANOVA with Games-Howel post- hoc test). The same trend was observed for *Oribacterium* (p-value = 0.024, ANOVA with Games-Howel post- hoc test) (see appendix 7).

When pairwise comparisons were conducted, at phyla level, there was a significant increase of Fusobacteria at “pos2”, when comparing with the samples collected at exacerbation (t-test, p-value = 5.02e-3), while there was a significant decrease of Actinobacteria in “recovery” comparatively to “pos2” (t-test, p-value = 0.027) (see appendix 8). No statistically significant differences were found at the level of relative abundance, at phyla level, between samples collected at exacerbation and samples collected at recovery.

At the genera level, there was a significant increase of *Oribacterium* at “pos2” when comparing with “exacerbation” (t-test, p-value = 0.021) as well as an increase in *Leptotrichia* for the same time points (t-test, p-value = 0.011). Additionally, when comparing the exacerbation samples with the samples collected at recovery, there was a significantly decrease of the genus *Streptococcus* (t-test, p-value = 0.018), as would be expected by the results from the ANOVA. However, the highest

differences, in terms of relative abundance, were found between the samples collected at “pos2” and at “recovery”, where there were significant decreases of *Rothia* (t-test, p-value = $2.02e^{-3}$), *Oribacterium* (t-test, p-value = 0.032), and *Veillonella* (t-test, p-value = 0.034), while there was a significant increase in the relative abundance of *Neisseria* (t-test, p-value = 0.049) (see appendix 9).

3.5 Microbial composition of individual subjects

Changes in the microbial community of different patients during the time course were explored, based on the results of the PERMANOVA with the Bray-Curtis metric. For that, samples were grouped by patient and alpha-diversity for each patient was accessed, through richness (number of OTU's observed and Chao1 estimator) and diversity (Shannon and Simpson index), as shown in figure 14.

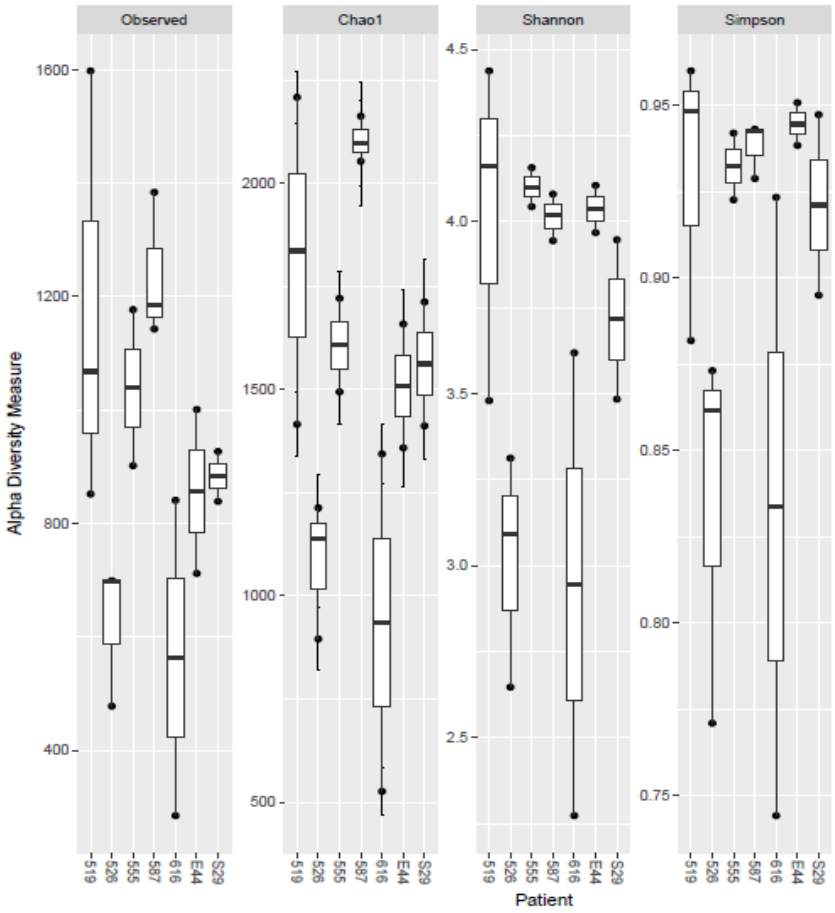


Figure 14 - Alpha diversity comparisons between patients. Samples were groupped by patient clustering all the samples collected at different time points , and bacterial richness (Observed, and Chao1), diversity (Shannon and Simpson Index) were calculated. Samples from patient 587 had the higher richness, while samples from patient 519 had higher diversity. The samples collected from patient 616 had the lowest richness and diversity.

From figure 14, it became clear that samples collected from patient 587 had higher richness, while samples from patient 519 had higher diversity. On the other hand, patient 616 presented the lowest richness and diversity.

Given these results, that show that there was a great inter-subject variation, it became important to analyze every patient individually.

3.5.1 Subject 519

Subject 519 was a 79 years old man with moderate COPD (GOLD stage 2) that was recruited during exacerbation and treated with azithromycin, a large spectrum antibiotic often used for AECOPD treatment (GOLD 2017). Three samples were collected from this subject, exacerbation ("Exacerbation"), after 2 weeks ("Pos2") and after 6 weeks ("Recovery") of exacerbation.

Firmicutes, Actinobacteria, and Proteobacteria were the most abundant phyla. As a curiosity, at exacerbation and before any treatment, a bacteria of the Cyanobacteria phylum was identified (Appendix 6), across the three samples collected from this patient. On the other hand, *Streptococcus*, *Veillonella*, and *Rothia* were the genera with highest relative abundance (figure 15).

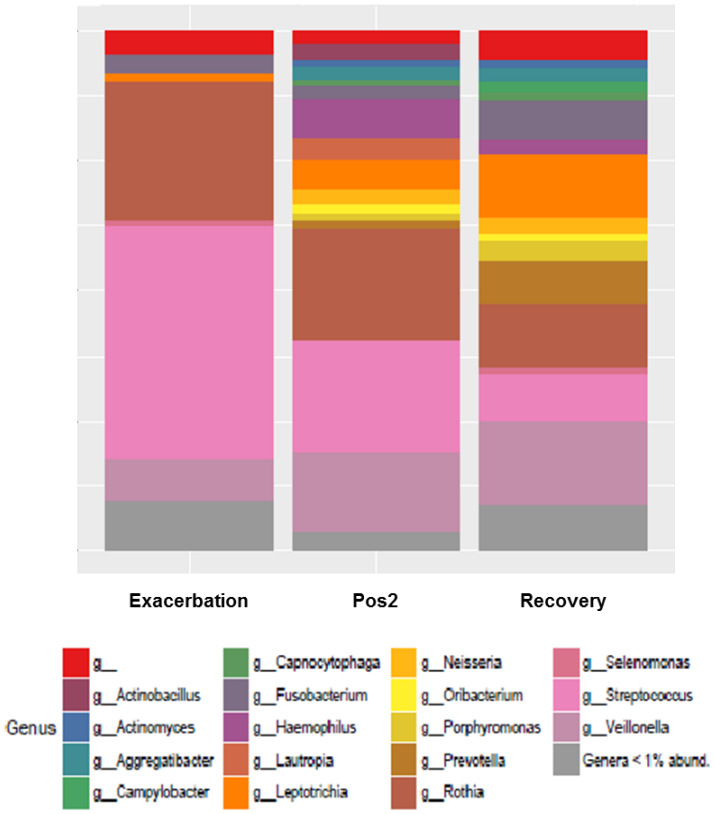
For this patient, "Recovery" was the time point of highest diversity, while the sample collected at exacerbation showed the lowest diversity, as can be seen in table 5.

Table 5 – Bacterial richness (number of OTU's observed and Chao1 estimator) and diversity (Shannon and Simpson indexes) of the samples collected from patient 519. "Exacerbation" – exacerbation, "Pos2" – 2 weeks after exacerbation, "Recovery" – 6 weeks after exacerbation.

	Richness		**Diversity**	
	Observed	**Chao1**	**Shannon**	**Simpson**
519 – Exacerbation	830	1357.58	3.47	0.88
519 – Pos2	1030	1991.57	3.99	0.94
519 - Recovery	1516	2061.63	4.41	0.95

When comparing the relative abundances, at genera level, of the three collecting points, some patterns could be highlighted, such as the relative abundance of *Streptococcus* and *Rothia* that decreased over time, while the relative abundance of *Veillonella*, *Leptotrichia*, *Prevotella*, and *Neisseria* increased over time, even though the last one shows a really small increase at recovery, when comparing with "Pos2".

Curiously, *Haemophilus* was increased at “Pos2”, when comparing with the other two collecting points, whereas the genus *Lautropia* were only present at “Pos2”.



Exacerbation	Pos2	Recovery
(42,7%) *Streptococcus*	**(20,8%) *Streptococcus***	**(8,9%) *Streptococcus***
(26%) *Rothia*	**(20,5%) *Rothia***	**(11,4%) *Rothia***
(8%) *Veillonella*	**(14,7%) *Veillonella***	**(15,5%) *Veillonella***
(3,6%) *Fusobacterium*	(2,4%) *Fusobacterium*	(7,3%) *Fusobacterium*
(1,5%) *Leptotrichia*	**(5,3%) *Leptotrichia***	**(11,9%) *Leptotrichia***
(1%) *Selenomonas*	(0,2%) *Selenomonas*	(1,1%) *Selenomonas*
(0,9%) *Prevotella*	**(1,6%) *Prevotella***	**(7,9%) *Prevotella***
(0,7%) *Oribacterium*	(1,8%) *Oribacterium*	(1,3%) *Oribacterium*
(0,5%) *Campylobacter*	(0,3%) *Campylobacter*	(1,8%) *Campylobacter*
(0,5%) *Neisseria*	**(2,9%) *Neisseria***	**(3%) *Neisseria***
(0,3%) *Actinomyces*	(1,2%) *Actinomyces*	(1,4%) *Actinomyces*
(0,2%) *Actinobacillus*	(3%) *Actinobacillus*	(0,6%) *Actinobacillus*
(0,2%) *Haemophilus*	(7,4%) *Haemophilus*	(2,8%) *Haemophilus*
(0,2%) *Porphyromonas*	(1,1%) *Porphyromonas*	(3,7%) *Porphyromonas*
(0,1%) *Aggregatibacter*	(2,4%) *Aggregatibacter*	(2,7%) *Aggregatibacter*
(0,1%) *Capnocytophaga*	(1,2%) *Capnocytophaga*	(1,5%) *Capnocytophaga*
(0%) *Lautropia*	(4,1%) *Lautropia*	(0%) *Lautropia*

Figure 15 - Samples collect from the subject 519 at different time points – exacerbation (Exacerbation), 2 weeks after exacerbation (Pos2), and 6 weeks post exacerbation at clinical recover (Recovery) - taxonomic composition is shown at the genera level. Plot showing the relative abundances, each column in the plot represents a time of sample collection, and each color in the column represents the relative abundance of each taxon at genera level. *Streptococcus*, *Rothia* and *Veillonella* were the most prominent genera across the three sampling times.

3.5.2. Subject 587

Subject 587 was a 77 years old man, at GOLD 1 of the disease that is considered a very mild stage of COPD. This individual was a former smoker, even though he just smoked for one year. At AECOPD, he did not receive any treatment with antibiotic nor corticosteroid. Besides the exacerbation sample ("Exacerbation"), it was also collected a sample after two weeks of exacerbation ("Pos2") and after 6 weeks ("Recovery").

Firmicutes, Fusobacteria, and Proteobacteria were the phyla found at higher abundances (Appendix 6), while *Streptococcus*, *Veillonella*, *Neisseria*, and *Rothia* were the most common genera across all samples collected from this subject (Figure 16).

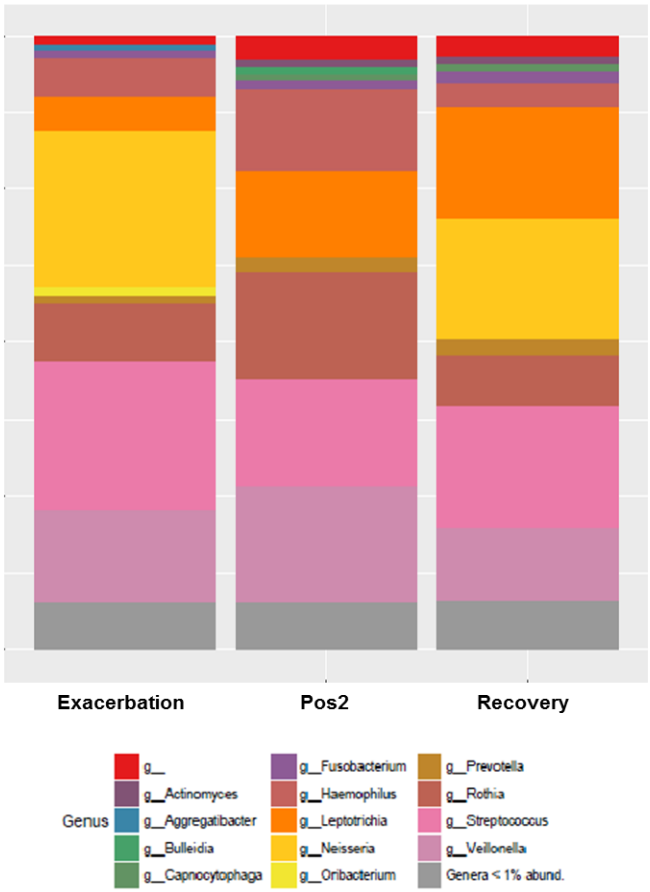
With regard to alpha-diversity over time, the saliva microbiome 2 weeks after the exacerbation was more diverse than at exacerbation although less diverse than at recovery (Table 6).

Table 6 - Bacterial richness (number of OTU's observed and Chao1 estimator) and diversity (Shannon and Simpson indexes) of the samples collected from patient 587. "Exacerbation" – at exacerbation, "Pos2" – 2 weeks after exacerbation, "Recovery" – 6 weeks after exacerbation.

	Richness		**Diversity**	
	Observed	Chao1	Shannon	Simpson
587 – Exacerbation	1122	1983.37	3.92	0.93
587 – Pos2	1091	1991.57	3.99	0.94
587 - Recovery	1323	2075.90	4.06	0.94

Over the time, there was an increase in the relative abundance of *Leptotrichia* and *Prevotella*, a pattern was also found in patient 519. *Neisseria* was present with a really small relative abundance (0,3%) in "Pos2" phase, which also showed a decrease in *Streptococcus*, while *Rothia*, *Veillonella*, and *Haemophilus* increased when compared with "exacerbation" and "Recovery". When comparing "Pos2" with "Recovery", *Veillonella*, *Rothia*, and *Haemophilus* decreased in relative abundance, while *Leptotrichia* increased. On the other hand, *Haemophilus* and *Neisseria*

decreased in relative abundance in “Recovery”, when comparing with “Exacerbation”, as shown in figure 16.



Exacerbation	Pos2	Recovery
(24,9%) *Neisseria*	**(0,3%) *Neisseria***	**(19,3%) *Neisseria***
(23,5%) *Streptococcus*	**(16,7%) *Streptococcus***	**(19,4%) *Streptococcus***
(14,7%) *Veillonella*	**(18,4%) *Veillonella***	**(11,6%) *Veillonella***
(9,2%) *Rothia*	**(17%) *Rothia***	**(8%) *Rothia***
(6,3%) *Haemophilus*	**(12,9%) *Haemophilus***	**(3,9%) *Haemophilus***
(5,4%) *Leptotrichia*	**(13,6%) *Leptotrichia***	**(17,9%) *Leptotrichia***
(1,5%) *Oribacterium*	(0,9%) *Oribacterium*	(0,9%) *Oribacterium*
(1,2%) *Prevotella*	(2,3%) *Prevotella*	(2,6%) *Prevotella*
(1,2%) *Fusobacterium*	(1,4%) *Fusobacterium*	(1,8%) *Fusobacterium*
(1%) *Aggregatibacter*	(0,3%) *Aggregatibacter*	(0,4%) *Aggregatibacter*
(0,9%) *Actinomyces*	(1,2%) *Actinomyces*	(1,2%) *Actinomyces*
(0,5%) *Bulleidia*	(1,2%) *Bulleidia*	(0,6%) *Bulleidia*
(0,4%) *Capnocytophaga*	(1%) *Capnocytophaga*	(1,2%) *Capnocytophaga*

Figure 16 - Samples collected from subject 587 at different time points – exacerbation (Exacerbation), 2 weeks after exacerbation (Pos2), and 6 weeks post exacerbation, after clinical recovery (Recovery) - taxonomic composition is shown at the genera level. Plot showing the relative abundances, each column in the plot represents a time point of sample collection, and each color in the columns represents the relative abundance of each taxon at the genera level. At pos2, there was a decrease in the relative abundance of *Streptococcus* and *Neisseria*, and an increase in *Veillonella*, *Rothia* and *Haemophilus*, when comparing with the other two sampling moments.

3.5.3 Subject 526

Subject 526 was a 78 years old individual, who had severe COPD - stage 3, according to GOLD COPD stages. He was also the heavier smoker individual of the group, who used to smoke 60 cigarettes per day during 36 years. At AECOPD, this subject was treated with Methylprednisolone, a corticosteroid frequently used in patients with COPD (Hunter & King 2001), that is used to suppress the immune system and decrease inflammation (Umberto Meduri et al. 2002). This was the only subject where a sample was collected at recovery, and post recovery, i.e. 8 months after exacerbation ("Recovery_B"). With this time point we aimed to compare the differences between two time points, of the same subject, that are clinically considered as being the same, since after 6 weeks of exacerbation a COPD patient is considered to be completely recovered (Seemungal et al. 2000).

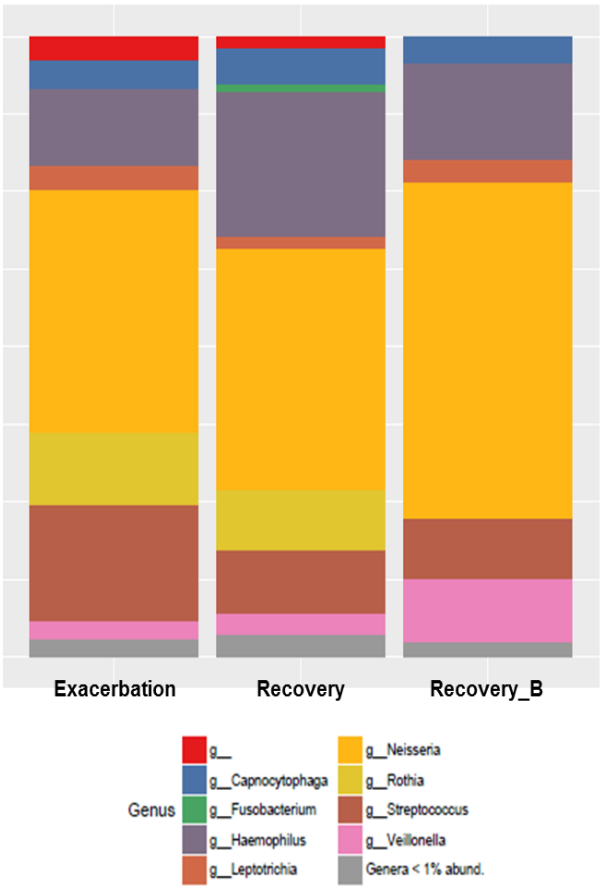
Proteobacteria and Firmicutes were the most prominent phyla (Appendix 6) among all samples of this patient, whereas *Neisseria* and *Haemophilus* were the most abundant genera (Figure 17). With regard to the microbiome dynamics, the sample collected at "exacerbation" had higher diversity than at "recovery" (Table 7), that it was accompanied by an increase in Firmicutes and Proteobacteria over time (Appendix 6).

Table 7 - Bacterial richness (number of OTU's observed and Chao1 estimator) and diversity (Shannon and Simpson indexes) of the samples collected from patient 526. "Exacerbation" – at exacerbation, "Recovery" – 6 weeks, and "Recovery_B" – 8 months after exacerbation.

	Richness		**Diversity**	
	Observed	Chao1	Shannon	Simpson
526 – Exacerbation	652	1076.86	3.29	0.87
526 – Recovery	650	1107.88	3.06	0.86
526 – Recovery_B	448	833.23	2.61	0.77

Furthermore, when comparing the two samples collected at recovery, diversity decreases at “Recovery_B” (Table 7). Additionally, in terms of phylogenetic composition, the sample collected 8 months after exacerbation, appears to be more different from the other samples collected, as it can be seen by the PCoA that was done with the unweighted UniFrac metric (Figure 11). As it can be seen by figure 17, *Haemophilus*, *Neisseria*, and *Rothia* had extremely differences in relative abundances, comparing the two recovery times of collection. However, the relative abundance of *Streptococcus* was identical at “Recovery” (10%) and “Recovery_B” (9,4%), being the only genus that presented identical relative abundances at recovery, i.e. after 6 weeks and after 8 months of exacerbation (Figure 17).

Interestingly the relative abundances of *Neisseria* were exactly the same at the samples collected at exacerbation and 6 weeks after (38%), while it increased at 8 months after exacerbation (52,8%) (Figure 17).



Exacerbation	Recovery	Recovery_B
(38%) *Neisseria*	**(38%) *Neisseria***	**(52,8%) *Neisseria***
(18,2%) *Streptococcus*	**(10%) *Streptococcus***	**(9,4%) *Streptococcus***
(11,9%) *Haemophilus*	**(22,7%) *Haemophilus***	**(15%) *Haemophilus***
(11,3%) *Rothia*	**(9,5%) *Rothia***	**(0,1%) *Rothia***
(4,6%) *Capnocytophaga*	(5,8%) *Capnocytophaga*	(4,1%) Capnocytophaga
(3,9%) *Leptotrichia*	(2,0%) *Leptotrichia*	(3,7%) *Leptotrichia*
(2,8%) *Veillonella*	(3,4%) Veillonella	(9,9%) *Veillonella*
(0,3%) *Fusobacterium*	(1,3%) *Fusobacterium*	(0%) *Fusobacterium*

Figure 17 - Samples collected from the subject 526 at different time points – exacerbation (Exacerbation), 6 weeks post exacerbation (Recovery) and 8 months after exacerbation (Recovery_B). Taxonomic composition is shown at the genera level. Plot with the relative abundances, each column of the plot represents a time point of sample collection, and each color in the columns represents the relative abundance of each taxon at the genera level. The microbial composition of the two recovery sampling times was remarkably different with exception for the relative abundance of *Streptococcus*.

3.5.4 Subject 555

Subject 555, a 76 years old man, had moderate COPD (GOLD 2) and, at exacerbation, was treated with Levofloxacin, a broad-spectrum antibiotic that is frequently used in patients with COPD (Miravitlles & Torres 2004). This subject was the only one from which a sample at "Recovery" was not collected.

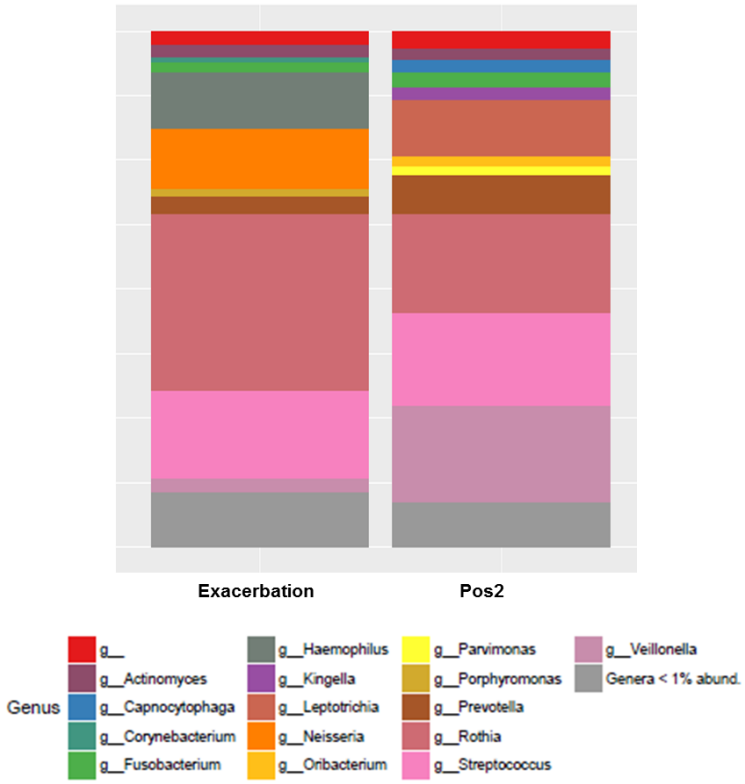
Overall, the two samples collected showed a different diversity, being the "Pos2" sample more diverse than the sample collected at "Exacerbation" (Table 8).

Table 8 - Bacterial richness (number of OTU's observed and Chao1 estimator) and diversity (Shannon and Simpson indexes) of the samples collected from patient 555. "Exacerbation" – at exacerbation, "Pos2" – two weeks after exacerbation.

	Richness		**Diversity**	
	Observed	Chao1	Shannon	Simpson
555 – Exacerbation	874	1435.70	4.02	0.92
555 – Pos2	1121	1635.04	4.13	0.94

Actinobacteria and Firmicutes were the predominant phyla (figure 5 and Appendix 6), whereas, *Rothia*, *Streptococcus*, and *Veillonella* were the genera with the highest relative abundance (Figure 18)

The relative abundances of *Rothia*, *Neisseria*, and *Haemophilus* were higher at the "exacerbation" sample than in the sample collected after 2 weeks, while the opposite trend was observed with *Veillonella*, *Prevotella*, and *Leptotrichia*. The relative abundance of *Streptococcus* did not suffer major changes (16,3% at "Exacerbation", and 17,5% at "Pos2") (Figure 18).



Exacerbation	Pos2
(33%) *Rothia*	**(18,8%) *Rothia***
(16,3%) *Streptococcus*	**(17,5%) *Streptococcus***
(11,4%) *Neisseria*	**(0,1%) *Neisseria***
(10,4%) *Haemophilus*	**(0,1%) *Haemophilus***
(3,2%) *Prevotella*	**(7,4%) *Prevotella***
(2,5%) *Veillonella*	**(18,3%) *Veillonella***
(2,2%) *Actinomyces*	(1,9%) *Actinomyces*
(1,9%) *Fusobacterium*	(2,8%) *Fusobacterium*
(1,4%) *Porphyromonas*	(0,9%) *Porphyromonas*
(1,1%) *Corynebacterium*	(0,6%) *Corynebacterium*
(0,8%) *Capnocytophaga*	(2,4%) *Capnocytophaga*
(0,8%) *Leptotrichia*	**(10,7%) *Leptotrichia***
(0,6%) *Parvimonas*	(1,6%) *Parvimonas*
(0,3%) *Oribacterium*	(1,9%) *Oribacterium*
(0,2%) *Kingella*	(2,4%) *Kingella*

Figure 18 - Samples collected from subject 555 at two different time points – exacerbation (Exac) and 2 weeks after exacerbation (Pos2). Taxonomic composition is shown at the genera level. **a.** Plot showing the relative abundances, each column in the plot represents a time point of sample collection, and each color in the columns represents the relative abundance of each taxon at the genera level. The relative abundances of *Rothia*, *Neisseria*, and *Haemophilus* were higher at the “Exacerbation” sample than in the “Pos2” sample, while the opposite trend was observed with *Veillonella*, *Prevotella*, and *Leptotrichia*.

3.5.5 Subject E44

Subject E44, a 73 years old man, suffered from moderate COPD (GOLD 2) and was the only individual, at this study, that never smoke, nor did receive any treatment with antibiotics or corticosteroids, at AECOPD. Only samples at exacerbation and 6 weeks after exacerbation ("Recovery") were collected.

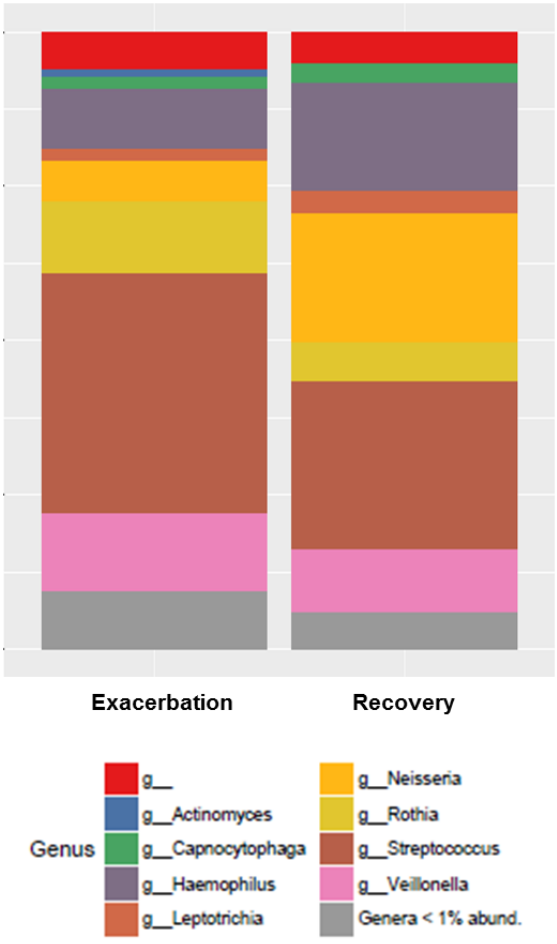
The sample collected at exacerbation was more diverse than the sample collected at recovery, table 9.

Table 9 - Bacterial richness (number of OTU's observed and Chao1 estimator) and diversity (Shannon and Simpson indexes) of the samples collected from patient E44. "Exacerbation" – exacerbation and "Recovery" – 6 weeks after exacerbation.

	Observed	**Chao1**	**Shannon**	**Simpson**
E44 – Exacerbation	688	1309.28	4.08	0.95
E44 – Recovery	955	1584.24	3.94	0.94

Firmicutes, Proteobacteria, and Actinobacteria were the most common phyla (Appendix 6). *Streptococcus*, *Neisseria*, *Veillonella*, and *Haemophilus* were the predominant genera, in the overall composition of the two samples, as shown in figure 19.

No major differences were found between the two moments of sample collection. Nevertheless, small differences could be found, such as the relative abundance of *Streptococcus*, *Neisseria*, and *Rothia* that decreased at "Recovery", while there was an increase in the proportion of *Neisseria* and *Haemophilus,* at the second time of collection (Figure 19).



Exacerbation	Recovery
(37,1%) *Streptococcus*	**(26,2%) *Streptococcus***
(12%) *Veillonella*	**(9,6%) *Veillonella***
(11,1%) *Rothia*	**(6%) *Rothia***
(9,3%) *Haemophilus*	**(16,8%) *Haemophilus***
(6,2%) *Neisseria*	**(20,2%) *Neisseria***
(1,9%) *Capnocytophaga*	(3%) *Capnocytophaga*
(1,8%) *Leptotrichia*	(3,5%) *Leptotrichia*
(1,1%) *Actinomyces*	(0,3%) *Actinomyces*

Figure 19 - Samples collected from subject E44 at two different time points – exacerbation (Exac) and 6 weeks post exacerbation at clinical recovery (Recovery). Taxonomic composition is shown at the genera level. **a.** Plot showing the relative abundances, each column in the plot represents a time point of sample collection, and each color in the columns represents the relative abundance of each taxon at genera level. The relative abundance of *Streptococcus*, *Neisseria*, and *Rothia* that decreased at “Recovery”, while there was an increase in the proportion of *Neisseria* and *Haemophilus*.

3.5.6 Subject 616

Subject 616 was the youngest individual of the group, 54 years old. As with most of the subjects, 616 suffered severe COPD (GOLD 3) and was a former smoker. At exacerbation, he was treated with Azithromycin. Besides the sample collected at exacerbation, a sample at recovery ("Recovery") was also collected.

The sample collected at AECOPD from this patient was the least diverse sample of all, and likewise it was also less diverse than the sample collected at recovery, as it can be seen in table 10.

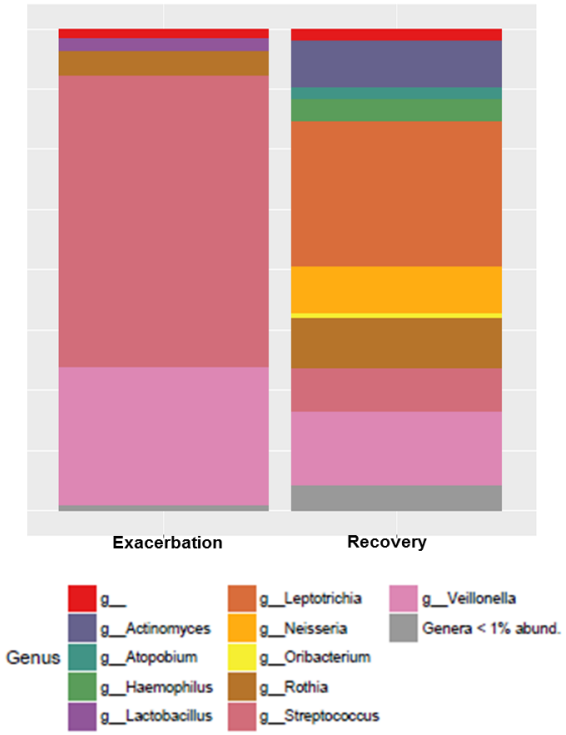
Table 10 - Bacterial richness (number of OTU's observed and Chao1 estimator) and diversity (Shannon and Simpson indexes) of the samples collected from patient 616. "Exacerbation" – at exacerbation and "Recovery" – 6 weeks after exacerbation.

	Richness		**Diversity**	
	Observed	Chao1	Shannon	Simpson
616 – Exacerbation	279	512.68	2.26	0.74
616 – Recovery	802	1275.69	3.69	0.92

Firmicutes and Actinobacteria were the most common phyla (Appendix 6), and *Veillonella*, *Streptococcus*, *Leptotrichia* and *Rothia* were the most abundant genera, among the two times of sample collection.

During exacerbation, two Firmicutes genera – *Veillonella* and *Streptococcus* accounted for around 90% of the total relative abundance. This proportion decreased dramatically during recovery were these two genera represented around 22% of all genera (Figure 20).

During recovery there was an increase in *Rothia* compared to exacerbation. *Actinomyces*, *Atopobium*, *Oribacterium*, *Leptotrichia*, *Neisseria*, and *Haemophilus* were absent during exacerbation, while at recovery they accounted for over 55% of all genera present in the sample (figure 20).



Exacerbation	Recovery
(60,3%) *Streptococcus*	**(8,7%) *Streptococcus***
(28,5%) *Veillonella*	**(14,5%) *Veillonella***
(5,1%) *Rothia*	(9,9%) *Rothia*
(2,5%) *Lactobacillus*	(0%) Lactobacillus
(0%) *Actinomyces*	**(9,2%) *Actinomyces***
(0%) *Atopobium*	**(2,5%) *Atopobium***
(0%) *Oribacterium*	**(1%) *Oribacterium***
(0%) *Leptotrichia*	**(28,9%) *Leptotrichia***
(0%) *Neisseria*	**(9,2%) *Neisseria***
(0%) *Haemophilus*	**(4,3%) *Haemophilus***

Figure 20 - Samples collected from subject 616 at two different time points – exacerbation ("Exac") and 6 weeks post exacerbation, after clinical recovery ("Recovery"). Taxonomic composition is shown at the genera level. **a.** Plot showing the relative abundances, each column in the plot represents a time point of sample collection, and each color in the columns represents the relative abundance of each taxon group at the genera level. *Veillonella* and *Streptococcus* accounted for around 90% of the total relative abundance at "Exacerbation", while at "Recovery" this proportion decreased for around 22% of all genera.

3.5.7 Subject S29

Subject S29, a 58 years old man, was the only subject of the group to be in the most severe stage of the disease (GOLD 4). At exacerbation, he was treated with azithromycin. A sample at exacerbation ("Exacerbation") and 6 weeks after exacerbation ("Recovery") was collected.

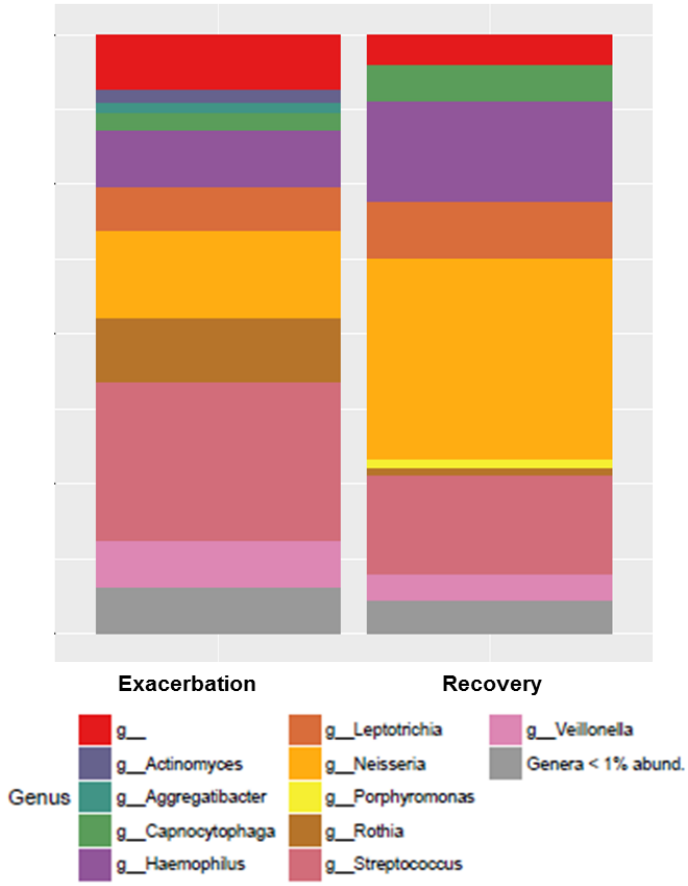
Proteobacteria and Firmicutes were the most abundant phyla, with an increase in the relative abundance of Proteobacteria and a decrease in Firmicutes, over time (Appendix 6). On the other hand, *Streptococcus*, *Neisseria*, and *Haemophilus* were the most common genera, in both samples collected (Figure 21).

Contrary to what was observed in the other individuals, the sample collected from S29 at exacerbation had higher richness and diversity than the one collected at recovery (Table 11).

Table 11 - Bacterial richness (number of OTU's observed and Chao1 estimator) and diversity (Shannon and Simpson indexes) of the samples collected from patient S29. "Exacerbation" – exacerbation and "Recovery" – 6 weeks after exacerbation.

	Observed	**Chao1**	**Shannon**	**Simpson**
S29 – Exacerbation	899	1636.23	3.93	0.95
S29 – Recovery	798	1338.08	3.45	0.89

In terms of community dynamics, the relative abundances of *Streptococcus*, and *Rothia* decreased in recovery, whereas the relative abundance of *Neisseria* and *Haemophilus* increased (Figure 21).



Exacerbation	Recovery
(25,6%) *Streptococcus*	**(16,2%) *Streptococcus***
(14%) *Neisseria*	**(32,7%) *Neisseria***
(10,4%) *Rothia*	**(1,2%) *Rothia***
(9,1%) *Haemophilus*	**(16,4%) *Haemophilus***
(7,4%) *Veillonella*	(4,4%) *Veillonella*
(7,1%) *Leptotrichia*	(9,1%) *Leptotrichia*
(2,8%) *Capnocytophaga*	(5,9%) *Capnocytophaga*
(2,3%) *Actinomyces*	(0,4%) *Actinomyces*
(1,5%) *Aggregatibacter*	(0,1%) *Aggregatibacter*
(0,5%) *Porphyromonas*	(1,3%) *Porphyromonas*

Figure 21 - Samples collect from the subject S29 at two different time points – exacerbation ("Exac") and 6 weeks post exacerbation at clinically recover ("Recovery") - taxonomic composition is shown at the genera level. **a.** Plot showing the relative abundances, each column in the plot represents a time of sample collection, and each color in the column represents the relative abundance of each taxon group at genera level. The relative abundances of *Streptococcus*, and *Rothia* decreased at "Recovery", while the relative abundances of *Neisseria* and *Haemophilus* increased.

Chapter IV – Discussion

4.1. Methodology

One of the aims of this study was to establish protocols for the saliva collection, conservation, transport and laboratorial treatment to reach good overall sample quality. Even though, a recent study from Lim et al. (2017), suggested that the collection method and DNA extraction methods, do not affect the saliva microbiome profiles obtained (Lim et al. 2017), it was pivotal to access the factors of variability that were introduced in the elapse of this study, and determine if and how they influenced our results.

The procedure was all the same for all samples collected, with exception for the volume of saliva collected, 4 out of the 17 samples had 500 μL of saliva, instead of 1 mL. Even though the protocol has been adjusted to correct for this fact, it could have introduced some variability on our results. Based on the results from the PERMANOVA analysis of the Bray-Curtis variances, one cannot say that this was a relevant variation, since no statistically differences were found between the two volumes of sample collected. Even though, no statistically differences were found, we would suggest that in a subsequent study the volume of saliva collected be taken into consideration, since this can be a factor of variability and lead to differences in the obtained results.

Considering the time constrains, it was not possible to relate all the clinical characteristics of the patients (Appendix 4) with the metagenomic data. Thereby, in a future study this characterization should be taken into account in order to do a more detailed study.

4.2. A “core” microbiome in saliva samples from patients with COPD

In this study, the predominant phyla found in all saliva samples were Firmicutes, Proteobacteria, Actinobacteria, Fusobacteria, and Bacteroidetes. This is consistent with the results from sputum (Wang et al. 2016; Su et al. 2015), BAL (Erb-Downward et al. 2011), and oropharynx samples (Park et al. 2014), collected from patients with COPD, although BAL samples seem slightly richer in Fusobacteria.

At genera level, *Streptococcus*, *Neisseria, Rothia, Veillonella*, *Haemophilus*, and *Leptotrichia* were the most abundant. Even though most of the genera detected are in accordance with Wang et al. (2016), Erb-Downward et al. (2011), and Su and colleagues (2015), in samples collected from patients with COPD, in our samples we did not find *Pseudomonas* or *Moraxella*, two genera belonging to the lung microbiome, as identified both by culture dependent (Sethi et al. 2002), and independent methods (Wang et al. 2016; Erb-Downward et al. 2011). However, Su et. al (2015) did not find an increased presence of this genera either, based on a culture independent method, which suggests that there is high variability in genera associated with the COPD microbiome. Therefore other studies are needed to reach a better characterization of this microbiome.

Interestingly, *Porphyromonas*, a genus frequently found in the healthy human oral microbiome at higher relative abundances (Zaura et al. 2009), was found by us as representing only a small percentage of the overall microbiome composition, although being present in 16 out of the 17 samples that were studied.

The “core microbiome” of the lower airway has been described as remarkably similar to the saliva microbiome. This supports the theory of Charlson et al., 2011, later demonstrated by Bassis et al. (2015) and Dickson and colleagues (2015), that defends that the upper airway, especially the oral cavity microbiome, shapes the lower airway, through microaspiration. Therefore, and based on the results from this study, we assume that saliva is a representative sample for the study of the lung microbiome in patients suffering with COPD.

4.3. Microbiome shifts after AECOPD treatment

The most notably overall changes in community composition over time were found in the first samples taken after treatment for exacerbation (pos2). This was also shown in the studies from Huang et al. (2014) and Wang et al. (2016), with an increase in alpha-diversity, using both richness and diversity in samples collected two weeks after exacerbation. Additionally, at our results, it was found a microbial compositional shift towards an increase in the relative abundance of *Veillonella*, *Rothia*, *Oribacterium* and *Leptotrichia* at this collection time. These results are not in agreement with the study from Huang and colleagues, that also suggested that there were higher differences at the first collection time after treatment, although these differences reflect alterations in the relative abundances of genera belonging to the Proteobacteria phyla (Huang et al. 2014).

Likewise large differences were found between the samples collected at "pos2" and at recovery, where there was a significant decrease in the relative abundance of *Rothia*, *Oribacterium*, and *Veillonella*, in contrast with an increase in *Neisseria*.

At recovery, the diversity of the microbial community decreased, becoming more similar to the diversity level at exacerbation. No differences were found at phyla level when comparing the relative abundances from this two collection time points. However, there was a decrease in the relative abundance of the genus *Streptococcus*, which is in accordance with the results from Liu et al. (2017), for sputum and oropharyngeal samples.

Streptococcus is one of the genera frequently found at higher abundances in sputum samples collected during AECOPD, through culture dependent (Hirschmann 2000) and independent methods (Millares et al. 2014). Since there was an overall reduction in the relative abundance of this genera over time perhaps this genus could be used to evaluate if the individual has already recovered from the exacerbation or not. Although, further studies should be done to confirm if this trend is observed in a larger cohort, as well as if there is an overall increase of *Streptococcus* before exacerbation until the onset of this acute event, where this genus would have, supposedly, its higher abundance.

In regard to the two samples from the same subject, patient 526, that were collected at recovery i.e. 6 weeks after exacerbation (Recovery), and 8 months after (Recovery_B), and since these two samples were collected at the same stadium of the disease, one would expect that these would be more similar. However, the sample collected at 8 months after exacerbation had less richness and diversity that the one collected 6 weeks after exacerbation. Likewise, the microbial composition of the two samples are notably different (see Figure 11 and 17), with exception for the relative abundance of the genus *Streptococcus*, that had a similar relative abundance in the two collection time points.

These results suggest that perhaps there are other factors, non-related with the disease that may induce differences in the microbial composition. However, the fact that the *Streptococcus* genus was found at similar relative abundances at the two collection points, reinforces the idea that *Streptococcus* could be used for motorizing the progression of the disease.

From our results, we cannot conclude when the microbiome is considered to be stable, and since our sampling time began at exacerbation, one still needs to collect samples from the whole cycle of the disease, i.e. to start the collection of samples before exacerbation. For that a larger longitudinal study should be done, with a larger cohort and with more sampling time points i.e. once a month until exacerbation, during the acute event, two weeks after exacerbation, 6 weeks after exacerbation and then returning to monthly sampling collections. Studies comprehending pre-exacerbation and post-exacerbation collection time points were already done to study COPD, but using sputum samples (Wang et al. 2016; Huang et al. 2014).

However, from our knowledge, there are no studies like this using saliva, which would allow us to further comprehend the saliva microbiome dynamics at stability (before exacerbation) and which changes occur just before the onset of an exacerbation. Gaining this knowledge would be important to validate saliva microbiome as a potential biomarker for predicting exacerbations in patients with COPD. Furthermore, questions such as how the microbiome evolves after an exacerbation, how it is affected by the prescribed treatment, when does it become similar to the microbiome at stability, or does it never return to the same point.

Nevertheless, we suggest to focus further studies on the dynamics of the *Streptococcus* genus, which from our results seems to be the genus that should be more deeply monitored, in patients suffering from COPD.

4.4. Inter-subject variation

From our results it was evident that, even though there are some statistically significant differences over time, the strongest evidence was that there was an enormous heterogeneity among patients with COPD. These results are in accordance with Su et al. (2015) that showed that even though there are some patterns across patients, the microbiome dynamics were almost personalized to each patient. So it will be important to focus on the microbiome variation in smaller groups of subjects, stratified according to their GOLD grade of disease and other clinical information, to be able to determine how the microbiome truly varies after an acute event of COPD.

In accordance with the overall observation, also the samples collected from patients 519, 587, and 555 had higher diversity and richness at 2 weeks post exacerbation. However, the three patients received different treatments, while 519 and 555 were treated with an antibiotic, patient 587 did not receive any treatment, which may suggest that these differences are not dependent of the treatment administrated.

Additionally, when comparing the samples collected at exacerbation and after 6 weeks, from patients treated with antibiotics – patient 616 and S29 -, on patient 616, it was found an increment in richness and diversity, while on patient S29 the opposite trend was found. Patient 526 was the only individual of the cohort that was treated with a corticosteroid at exacerbation, when comparing the samples collected at exacerbation and the two collected at recovery, it was evident that there was a decrease of diversity over time, while there was an increase in Proteobacteria and Firmicutes. These results are in accordance with what Wang et al. (2016) and Huang et al. (2014) observed in individuals treated only with corticosteroids.

Given our data, one can speculate that treatment for AECOPD overtime, can modify the microbiome community composition of the patients, and therefore contribute to the increase of disease severity and further enhancement of the risk for new exacerbations, as suggested by Huang et al. 2017. To confirm that, however, our approach must be strengthen with more longitudinal studies with

patients receiving different treatments and being followed over longer periods of time.

When grouping the patients by the severity of disease from mild to moderate (GOLD 1 and 2) and severe to very severe (GOLD 3 and 4), one can see that in our cohort, the samples collected from patients in the lower stages of the disease present an overall higher richness and diversity. With exception for patient S29, that was in the very severe stage of the disease and presents also high diversity in his samples. On the other hand, samples collected from patients in the severe stage of the disease – patient 616 and 526 -, had lower diversity and richness. This results disagree with the Pragman et al. (2012), and Garcia-Nuñez and colleagues (2014), which shown that patients in the later stages of the disease has a higher diversity than the ones at the less severe stages. However, our results are in accordance with Erb-Downward et al. (2011). Given the heterogeneity of the results, presented by us and by other authors, more studies comprehending larger cohorts with subjects in different stages of COPD should be performed.

Our results support the idea that despite the fact that there was a core microbiome in the saliva of all patients with COPD, there were also strong inter-individual variations, suggesting that every subject has a unique microbiome that may play a role in the etiology of the disease, leading to different manifestation and progression of the disease among different subjects (Zarco et al. 2012). Therefore, the study of the individual human microbiome may be used to assess the progression of the disease, to evaluate if a certain treatment is working, or not, and how this treatment may compromise the microbiome and consequently the disease progression in a long term. These evidences may suggest that the study of the microbiome should indeed be incorporated into the emergent field of personalized medicine (Sonnenburg & Fischbach 2011).

4.5 Saliva as a potential target for COPD biomarkers

As shown by our results and other studies, COPD is a heterogeneous and complex lung disease. So, the identification of clinical biomarkers would help to diagnose the disease, monitor its progression, evaluate the response to therapy (Shaw et al. 2014) decide the best treatment to apply or even to predict the onset of exacerbation.

Since collecting saliva is a non-invasive and easy process, and, as shown above, it is possible to detect the “core microbiome” in these samples, this methodology becomes highly interesting under this scope. Furthermore, as suggested by Pragman et al. (2012), the study of the oral microbiome may be important to understand the disease progression and exacerbations. This work was then the first step to advance our understanding about the role of saliva microbiome in COPD and to highlight a new potential source of biomarkers for this disease. Nevertheless, much more studies must be done to support this claim, because strong statistical evidence is needed to establish a clinical biomarker based on the saliva microbiome.

However, considering the fact that it was observed an overall decrease in the relative abundance of *Streptococcus* over time, this could become a good biomarker to evaluate the disease progression after an exacerbation. Also Cameron et al. (2016), with sputum samples, and Diao et al. (2017) with oropharyngeal and sputum samples, suggested that changes in the abundance of *Streptococcus* may allow to differentiate between healthy and COPD subjects, as well as to use it as a predictor of COPD progression. Again, further studies with larger cohorts should be done to confirm it.

4.6. Limitations and further perspectives

An important limitation of this study was that the microbiome analyses were carried out on only 17 saliva samples from 7 different patients with COPD, which limits the ability to conduct rigorous statistical analysis. For a longitudinal study, this was an extremely small sample size, therefore extrapolating this results to a broader group of patients should be cautiously done. Also the fact that only samples at exacerbation and post-exacerbation were analyzed, and the heterogeneity of subjects and time points of sample collection per subject, make very difficult to validate the results. Indeed, this study must only be considered a pilot project that managed to implement the methodology so that a valid population study can be conducted, involving a larger number of patients and pre-exacerbation (stable state) samples as well.

An additional limitation of this study, was that only patients suffering with COPD were analyzed, with no healthy or control subjects (subjects that were not at exacerbation). Data from these populations would be pivotal to define the normal saliva microbiota and what is altered in COPD, as well as the changes in microbial composition at different disease phenotypes.

For a future study, it would be interesting to do a deeper analyzes of the saliva samples collected, such as a standard culture, to access the bacterial load in the sample, as well as the presence of typical COPD pathogens (e.g. *Moraxella catarrhalis*, *Staphylococcus aureus*) that are found in other types of samples, e.g. in sputum. This would be interesting to validate the saliva sample as a good sample to study the microbiome in patients with COPD, as well as to differentiate between bacterial and eosinophilic exacerbations. A study like this would allow us to study the microbial differences in the two types of these acute events, and could be interesting for clinical purposes. Since, antibiotics are often prescribed for the treatment of exacerbations, and since there are some non-bacterial exacerbations, if we could distinguish the type of exacerbation based on the microbial composition, a more adjusted and effective treatment could be prescribed.

Another drawback was the fact that this was a metagenomic study based on the 16S rRNA gene amplification alone. A shotgun metagenomics approach would

give much more information, not only about the bacterial composition of the samples, but also about their biological functions, as encoded in the genomes (Sharpton 2014). Ideally, in order to achieve the most complete functional characterization of the microbiome, a combination of shotgun genome sequencing, metatranscriptomics, metabolomics, metaproteomics, and viromics would be necessary (Bikel et al. 2015). A study comprehending all these approaches, although extremely ambitious and costly, would be highly informative about the interplay between saliva microbiome and COPD.

Chapter V – Conclusions

Given the nature of this study, a lot of questions arose from it which should be answered in further studies, as well as some suggestions that may allow us to extend our knowledge in this emerging field. The major limitation of this study was the small number of samples that were analyzed, which limited the ability to conduct rigorous statistical analysis. In further studies, the sample size should take into consideration, as well as the selection of sampling times, patients' characteristics, and the usage of different treatments for exacerbation. In order to do a better characterization of the saliva microbiome and its dynamics in patients suffering with COPD.

Even considering all the mentioned limitations of this study, we could also extract from it some conclusions. The main objective of this study was to check if the saliva microbiome could be used to evaluate some of the clinical outcomes of respiratory diseases, such as COPD. In doing so, we proposed to study the dynamics of the saliva microbiome at exacerbation and how the microbial community evolved after that acute event. Even though this study comprehends a very small group of patients with COPD, the results of this study indicated that saliva microbiome should be taken into consideration to improve our understanding of the pathology itself as well as of the exacerbation event. We also proposed that *Streptococcus* should be considered as a potential biomarker for disease progression. Additionally, we envisage that the study of the saliva microbiome in patients with COPD may be important in the scope of personalized medicine.

References

Acharya, A. et al., 2017. Salivary microbiome in non-oral disease: A summary of evidence and commentary. *Archives of Oral Biology*, 83(May), pp.169–173.

Altschul, S.F. et al., 1990. Basic local alignment search tool. *Journal of molecular biology*, 215(3), pp.403–10.

Amaral, S.M., Cortês, A.D.Q. & Pires, F.R., 2009. Nosocomial pneumonia: importance of the oral environment. *Jornal brasileiro de pneumologia : publicacao oficial da Sociedade Brasileira de Pneumologia e Tisilogia*, 35(11), pp.1116–1124.

Anderson, M.J., 2001. A new method for non parametric multivariate analysis of variance. *Austral ecology*, 26(2001), pp.32–46.

Apprill, A. et al., 2015. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquatic Microbial Ecology*, 75(2), pp.129–137.

Bafadhel, M. et al., 2011. Acute exacerbations of chronic obstructive pulmonary disease: Identification of biologic clusters and their biomarkers. *American Journal of Respiratory and Critical Care Medicine*, 184(6), pp.662–671.

Barnes, P.J. et al., 2015. Chronic obstructive pulmonary disease. *Nature Reviews Disease Primers*, (December), p.15076.

Bassis, C.M. et al., 2015. Analysis of the upper respiratory tract microbiotas as the source of the lung and gastric microbiotas in healthy individuals. *mBio*, 6(2), pp.1–10.

Bik, E.M., 2016. The hoops, hopes, and hypes of human microbiome research. *Yale Journal of Biology and Medicine*, 89, pp.363–373.

Bikel, S. et al., 2015. Combining metagenomics, metatranscriptomics and viromics to explore novel microbial interactions: Towards a systems-level understanding of human microbiome. *Computational and Structural Biotechnology Journal*, 13, pp.390–401.

Blaser, M.J., 2006. Who are we? Indigenous microbes and the ecology of human diseases. *EMBO reports*, 7(10), pp.956–960.

Blaxter, M. et al., 2005. Defining operational taxonomic units using DNA barcode data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1462), pp.1935–1943.

Bluman, A.G., 2007. *Elementary Statistics: A Step by Step Approach* (6th editi., New York: McGraw Hill Higher Education.

Bokulich, N.A. et al., 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature methods*, 10(1), pp.57–59.

Bray, R.J. & Curtis, J.T., 1957. An ordination of the upland forest communities of southern Winsconin. *Ecological Monographs*, 27(4), pp.325–349.

Cabrera-Rubio, R. et al., 2012. Microbiome diversity in the bronchial tracts of patients with chronic obstructive pulmonary disease. *Journal of Clinical Microbiology*, 50(11),

pp.3562–3568.

Caldas, R.R. et al., 2015. Pseudomonas aeruginosa and periodontal pathogens in the oral cavity and lungs of cystic fibrosis patients: A case-control study. *Journal of Clinical Microbiology*, 53(6), pp.1898–1907.

Cameron, S.J.S. et al., 2016. Metagenomic sequencing of the Chronic Obstructive Pulmonary Disease upper bronchial tract microbiome reveals functional changes associated with disease severity. *PLoS ONE*, 11(2), pp.1–16.

Caporaso, J.G. et al., 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences*, 108(Supplement_1), pp.4516–4522.

Caporaso, J.G. et al., 2010. PyNAST: A flexible tool for aligning sequences to a template alignment. *Bioinformatics*, 26(2), pp.266–267.

Caporaso, J.G. et al., 2010. QIIME allows analysis of high-troughput community sequencing data. *Nature methods*, 7(5), pp.335–336.

Chao, A., 1984. Nonparametric estimation of the number of classes in a population. *Scanadinavian Journal of statistics*, 11(4), pp.265–270.

Charlson, E.S. et al., 2010. Disordered microbial communities in the upper respiratory tract of cigarette smokers. *PLoS ONE*, 5(12), pp.1–10.

Charlson, E.S. et al., 2011. Topographical continuity of bacterial populations in the healthy human respiratory tract. *American Journal of Respiratory and Critical Care Medicine*, 184(8), pp.957–963.

Chen, K. & Pachter, L., 2005. Bioinformatics for whole-genome shotgun sequencing of microbial communities. *PLoS Computational Biology*, 1(2), pp.0106–0112.

Cho, I. & Blaser, M.J., 2012. The human microbiome: at the interface of health and disease. *Nature Reviews Genetics*, 13(4), pp.260–270.

Cox, M.J., Cookson, W.O.C.M. & Moffatt, M.F., 2013. Sequencing the human microbiome in health and disease. *Human Molecular Genetics*, 22(R1), pp.88–94.

Decramer, M. & Vestbo, J., 2016. Chronic Obstructive Pulmonary Disease Updated 2016 Global Initiative for Chronic Obstructive Lung Disease. *Gold*, pp.1–44.

Degruttola, A.K. et al., 2016. Current understanding of dysbiosis in disease in human and animal models. *Inflammatory bowel diseases*, 22(5), pp.1137–1150.

Deorowicz, S. & Grabowski, S., 2011. Compression of DNA sequence reads in FASTQ format. *Bioinformatics*, 27(6), pp.860–862.

DeSantis, T. & Hugenholtz, P., 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, 72(7), pp.5069–5072.

Diao, W. et al., 2017. Characterization of throat microbial flora in smokers with or without

COPD. *Internetional Journal of COPD*, 12, pp.1933–1946.

Dickson, R.P. & Huffnagle, G.B., 2015. The lung microbiome: new principles for respiratory bacteriology in health and disease. *PLOS Pathogens*, 11(7), p.e1004923.

Dickson, R.P., Martinez, F.J. & Huffnagle, G.B., 2014. The role of the microbiome in exacerbations of chronic lung diseases. *The Lancet*, 384(9944), pp.691–702.

Van Dijk, E.L. et al., 2014. Ten years of next-generation sequencing technology. *Trends in Genetics*, 30(9), pp.418–426.

Drancourt, M. et al., 2000. 16S ribosomal DNA sequence analysis of a large collection of environmental and clinical unidentifiable bacterial isolates. *Journal of Clinical Microbiology*, 38(10), pp.3623–3630.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19), pp.2460–2461.

Erb-Downward, J.R. et al., 2011. Analysis of the lung microbiome in the “healthy” smoker and in COPD. *PLoS ONE*, 6(2).

Escobar-Zepeda, A., De León, A.V.P. & Sanchez-Flores, A., 2015. The road to metagenomics: From microbiology to DNA sequencing technologies and bioinformatics. *Frontiers in Genetics*, 6(DEC), pp.1–15.

Ewing, B. et al., 1998. Base-calling of automated sequencer traces using Phred. *Genome Research*, (206), pp.175–185.

GOLD, 2017. Global initiative for chronic obstructive lung disease. , 2017 Report.

Gomez, C. & Chanez, P., 2016. The lung microbiome: The perfect culprit for COPD exacerbations? *European Respiratory Journal*, 47(4), pp.1034–1036.

Goode, M.R. et al., 2014. Collection and extraction of saliva DNA for next generation sequencing. *Journal of Visualized Experiments*, (90), pp.1–7.

Grice, E.A. & Segre, J.A., 2012. The human microbiome: our second Genome. *Annual review of genomics and human genetics*, 13(1), pp.151–170.

Guerra, B. et al., 2017. Prediction models for exacerbations in patients with COPD. *European Respiratory Review*, 26(160061), pp.1–19.

Haegeman, B. et al., 2013. Robust estimation of microbial diversity in theory and in practice. *The ISME Journal*, 7(6), pp.1092–1101.

Han, M.K. et al., 2012. Significance of the microbiome in obstructive lung disease. *Thorax*, 67(5), pp.456–63.

Hilty, M. et al., 2010. Disordered microbial communities in asthmatic airways. *PLoS ONE*, 5(1), pp.1–9.

Hirschmann, J. V., 2000. Do bacteria cause exacerbations of COPD? *Chest*, 118(1), pp.193–203.

Hodkinson, B.P. & Grice, E.A., 2015. Next-generation sequencing: a review of technologies and tools for wound microbiome research. *Advances in Wound Care*, 4(1), pp.50–58.

Huang, Y.J. et al., 2010. A persistent and diverse airway microbiota present during chronic obstructive pulmonary disease exacerbations. *OMICS A Journal of Integrative Biology*, 14(1), pp.9–59.

Huang, Y.J. et al., 2014. Airway microbiome dynamics in exacerbations of chronic obstructive pulmonary disease. *Journal of Clinical Microbiology*, 52(8), pp.2813–2823.

Huang, Y.J. et al., 2017. Understanding the role of the microbiome in chronic obstructive pulmonary disease: principles, challenges, and future directions. *Translational Research*, 179, pp.71–83.

Huang, Y.J. et al., 2016. Understanding the role of the microbiome in COPD: Principles, Challenges and Future Directions. *Translational Research*, pp.1–13.

Hugenholtz, P., 2002. Exploring prokaryotic diversity in the genomic era. *Genome biology*, 3(2), p.REVIEWS0003.

Hunninghake, G.M. et al., 2009. MMP12, lung function, and COPD in high-risk populations. *The New England Journal of Medicine*, 361(27), pp.2599–2608.

Hunter, M.H. & King, D.E., 2001. COPD: management of acute exacerbations and chronic stable disease. *American Family Physician*, 64(4), pp.603–612.

Illumina, 2015a. An introduction to next-generation sequencing technology. *Illumina*.

Illumina, 2015b. MiSeq System Guide. *Illumina*, 1(September), pp.17–58.

Krebs, C., 2014. *CHAPTER 13 - Species Diversity Measures*, Boston: Addison-Wesley Educational Publishers, Inc.

Kuczynski, J. et al., 2012. Using QIIME to analyze 16S rRNA gene sequences from microbial communities. *Current protocols in microbiology*, (June), pp.1–28.

Lederberg, B.J. & McCray, A.T., 2001. 'Ome sweet 'omics - a genealogical treasury of words. *The Scientist*, 15(7), p.8.

Letunic, I., 2015. phyloT : Phylogenetic Tree Generator [online].

Li, K. et al., 2012. Analyses of the microbial diversity across the human microbiome. *PLoS ONE*, 7(6), pp.1435–1439.

Lim, Y. et al., 2017. The saliva microbiome profiles are minimally affected by collection method or DNA extraction protocols. *Scientific Reports*, 7(1), pp.1–10.

Liu, H.Y. et al., 2017. Oropharyngeal and sputum microbiomes are similar following exacerbation of Chronic Obstructive Pulmonary Disease. *Frontiers in Microbiology*, 8(JUN), pp.1–11.

Lozupone, C. & Knight, R., 2005. UniFrac : a new phylogenetic method for comparing microbial communities. *Applied and environmental microbiology*, 71(12), pp.8228–8235.

Lu, Y. et al., 2016. Next generation sequencing in aquatic models. In *Next Generation Sequencing - Advances, Applications and Challenges*. pp. 62–79.

Mammen, M.J. & Sethi, S., 2016. COPD and the microbiome. *Respirology*, 21(4), pp.590–599.

Mardia, K. V., Kent, J.T. & Bibby, J., 1979. Multivariable analysis. *London: Academic Press*.

Marsland, B.J., Yadava, K. & Nicod, L.P., 2013. The airway microbiome and disease. *Chest*, 144(2), pp.632–637.

Mcdonald, D. et al., 2012. The Biological Observation Matrix (BIOM) format or : how I learned to stop worrying and love the ome-ome. *GigaScience*, 1(7), pp.1–6.

McElhoe, J.A. et al., 2014. Development and assessment of an optimized next-generation DNA sequencing approach for the mtgenome using the Illumina MiSeq. *Forensic Science International: Genetics*, 13, pp.20–29.

McMurdie, P.J. & Holmes, S., 2013. Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE*, 8(4).

Medini, D. et al., 2008. Microbiology in the post-genomic era. *Nature Reviews Microbiology*, 6(6), pp.419–430.

Metzker, M.L., 2010. Sequencing technologies - the next generation. *Nature reviews genetics*, 11(1), pp.31–46.

Metzker, M.L.M.L.L., 2005. Emerging technologies in DNA sequencing. *Genome research*, 15(12), pp.1767–76.

Millares, L. et al., 2014. Bronchial microbiome of severe COPD patients colonised by Pseudomonas aeruginosa. *European Journal of Clinical Microbiology and Infectious Diseases*, 33(7), pp.1101–1111.

Millares, L. et al., 2015. Functional metagenomics of the bronchial microbiome in COPD. *PLoS ONE*, 10(12), pp.1–13.

Miravitlles, M. & Anzueto, A., 2017. Chronic respiratory infection in patients with Chronic Obstructive Pulmonary Disease: What is the role of antibiotics? *International Journal of Molecular Sciences*, 18(7), pp.1–12.

Miravitlles, M. & Torres, A., 2004. Antibiotics in exacerbations of COPD: Lessons from the past. *European Respiratory Journal*, 24(6), pp.896–897.

Molyneaux, P.L. et al., 2013. Outgrowth of the bacterial airway microbiome after rhinovirus exacerbation of chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care Medicine*, 188(10), pp.1224–1231.

Morgan, X.C. & Huttenhower, C., 2012. Chapter 12: Human Microbiome Analysis. *PLoS Computational Biology*, 8(12), pp.1–14.

Morris, A. et al., 2013. Comparison of the respiratory microbiome in healthy nonsmokers and smokers. *American Journal of Respiratory and Critical Care Medicine*, 187(10), pp.1067–1075.

Navas-Molina, J.A. et al., 2013. *Chapter Nineteen: Advancing Our Understanding of the Human Microbiome Using QIIME* 1st ed., Elsevier Inc.

Nazir, S. a. & Erbland, M.L., 2009. Chronic obstructive pulmonary disease: an update on diagnosis and management issues in older adults. *Drugs & Aging*, 26(10), pp.813–831.

Nguyen, N.-P. et al., 2016. A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity. *npj Biofilms and Microbiomes*, 57(6), pp.10–13.

Oksanen, J. et al., 2017. Package "vegan": Community Ecology Package. *R package version 2.4-4*.

Oulas, A. et al., 2015. Metagenomics: Tools and insights for analyzing next-generation sequencing data derived from biodiversity studies. *Bioinformatics and Biology Insights*, 9, pp.75–88.

Parada, A.E., Needham, D.M. & Fuhrman, J.A., 2016. Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology*, 18(5), pp.1403–1414.

Parameswaran, G.I. et al., 2009. Moraxella catarrhalis acquisition, airway inflammation and protease-antiprotease balance in chronic obstructive pulmonary disease. *BMC Infectious Diseases*, 9(1), p.178.

Parameswaran, G.I. & Murphy, T.F., 2009. Chronic obstructive pulmonary disease: role of bacteria and updated guide to antibacterial selection in the older patient. *Drugs & Aging*, 26(12), p.985–995 11p.

Park, H. et al., 2014. Microbial communities in the upper respiratory tract of patients with asthma and chronic obstructive pulmonary disease. *PloS one*, 9(10), p.e109710.

Parker, C.M. et al., 2005. Physiological changes during symptom recovery from moderate exacerbations of COPD. *European Respiratory Journal*, 26(3), pp.420–428.

Parks, D.H. et al., 2014. STAMP: Statistical analysis of taxonomic and functional profiles. *Bioinformatics*, 30(21), pp.3123–3124.

Pascoe, S. et al., 2015. Blood eosinophil counts, exacerbations, and response to the addition of inhaled fluticasone furoate to vilanterol in patients with chronic obstructive pulmonary disease: A secondary analysis of data from two parallel randomised controlled trials. *The Lancet Respiratory Medicine*, 3(6), pp.435–442.

Van de Peer, Y., 1996. A quantitative map of nucleotide substitution rates in bacterial rRNA. *Nucleic Acids Research*, 24(17), pp.3381–3391.

Peng, C. et al., 2013. C-reactive protein levels predict bacterial exacerbation in patients with chronic obstructive pulmonary disease. *The American Journal of the Medical Sciences*, 345(3), pp.190–194.

Poretsky, R. et al., 2014. Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. *PLoS ONE*, 9(4).

Pragman, A.A. et al., 2012. The lung microbiome in moderate and severe chronic obstructive pulmonary disease. *PLoS ONE*, 7(10).

Price, M.N., Dehal, P.S. & Arkin, A.P., 2009. Fasttree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular Biology and Evolution*, 26(7), pp.1641–1650.

QIAGEN®, 2014. QIAamp® DNA Microbiome Handbook. *Qiagen*.

R Core Team, 2015. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna*.

Round, J.L. & Mazmanian, S.K., 2009. The gut microbiome shapes intestinal immune responses during health and disease. *Nature Reviews Immunology*, 9(5), pp.313–323.

RStudio Team, 2016. RStudio: Integrated Development for R. *RStudio, Inc., Boston, MA*.

Salvi, S.S. & Barnes, P.J., 2009. Chronic obstructive pulmonary disease in non-smokers. *The Lancet*, 374(9691), pp.733–743.

Sanger, F., Nicklen, S. & Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12), pp.5463–5467.

Schadt, E.E., Turner, S. & Kasarskis, A., 2010. A window into third-generation sequencing. *Human molecular genetics*, 19(2), pp.227–240.

Seemungal, T.A.R. et al., 2000. Time course and recovery of exacerbations in patients with chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical care Medicin*, 161, pp.1608–1613.

Sender, R., Fuchs, S. & Milo, R., 2016. Are we really vastly outnumbered? Revisiting the ratio of bacterial to host cells in humans. *Cell*, 164(3), pp.337–340.

Sethi, S. et al., 2002. Bacteria and exacerbations of chronic obstructive pulmonary disease. *The New England journal of medicine*, 347(7), pp.526–527.

Sethi, S. et al., 2008. Inflammatory profile of new bacterial strain exacerbations of chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care Medicine*, 177(5), pp.491–497.

Shannon, C.E. & Weaver, W., 1963. The mathematical theory of communication. *The University of Illinois Press*, 5(1), pp.1–131.

Sharpton, T.J., 2014. An introduction to the analysis of shotgun metagenomic data. *Frontiers in plant science*, 5(June), p.209.

Shaw, J.G. et al., 2014. Biomarkers of progression of chronic obstructive pulmonary disease (COPD). *Journal of Thoracic Disease*, 6(11), pp.1532–1547.

Shendure, J. & Ji, H., 2008. Next-generation DNA sequencing. *Nature Biotechnology*, 26(10), pp.1135–1145.

Simbolo, M. et al., 2013. DNA qualification workflow for next generation sequencing of histopathological samples. *PLoS ONE*, 8(6).

Simpson, E.H., 1949. Measurement of diversity. *Nature*.

Slater, M. et al., 2014. The impact of azithromycin therapy on the airway microbiota in asthma. *Thorax*, 69(7), pp.673–674.

Soler, N. et al., 2007. Bronchoscopic validation of the significance of sputum purulence in severe exacerbations of chronic obstructive pulmonary disease. *Thorax*, 62(1), pp.29–35.

Sonnenburg, J.L. & Fischbach, M.A., 2011. Community health care: therapeutic opportunities in the human microbiome. *Science translational medicine*, 3(78).

Su, J. et al., 2015. Sputum bacterial and fungal dynamics during exacerbations of severe COPD. *PLoS ONE*, 10(7), pp.1–13.

Sze, M.A. et al., 2015. Host response to the lung microbiome in chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care Medicine*, 192(4), pp.438–445.

Sze, M.A. et al., 2012. The lung tissue microbiome in chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care Medicine*, 185(10), pp.1073–1080.

The NIH HMP Working Group, 2009. The NIH Human Microbiome Project. *Genome Research*, 19(12), pp.2317–2323.

Torres, P.J. et al., 2015. Characterization of the salivary microbiome in patients with pancreatic cancer. *PeerJ*, 3, p.e1373.

Tyrion, R.C., 1939. Cluster analysis. *Ann Arbor, MI: Edwards Bros*.

Umberto Meduri, G. et al., 2002. Prolonged methylprednisolone treatment suppresses systemic inflammation in patients with unresolving acute respiratory distress syndrome: Evidence for inadequate endogenous glucocorticoid secretion and inflammation-induced immune cell resistance to glucoc. *American Journal of Respiratory and Critical Care Medicine*, 165(7), pp.983–991.

Vollenweider, D.J. et al., 2012. Antibiotics for exacerbations of chronic obstructive pulmonary disease. *Cochrane database of systematic reviews.*, (12).

Wade, W., 2002. Unculturable bacteria—the uncharacterized organisms that cause oral infections. *Journal of the Royal Society of Medicine*, 95, p.81.

Walters, W. et al., 2016. Improved bacterial 16S rRNA gene (V4 and V4-5) and fungal

internal transcribed spacer marker gene primers for microbial community surveys. *mSystems*, 1(1), pp.e0009-15.

Wang, Z. et al., 2016. Lung microbiome dynamics in COPD exacerbations. *European Respiratory Journal*, 47(4), pp.1082–1092.

WHO, 2007. Chronic respiratory disease. *Causes*, p.1. Available at: http://www.who.int/respiratory/copd/causes/en/ [Accessed January 16, 2017].

WHO, 2017a. Chronic respiratory diseases - Burden of COPD. Available at: http://www.who.int/respiratory/copd/burden/en/ [Accessed November 15, 2017].

WHO, 2017b. Chronic respiratory diseases - COPD management. *WHO*. Available at: http://www.who.int/respiratory/copd/management/en/ [Accessed October 20, 2017].

Woese, C.R. & Fox, G.E., 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America*, 74(11), pp.5088–5090.

Woodhead, M. et al., 2011. Guidelines for the management of adult lower respiratory tract infections. *Clinical microbiology and infection*, 17, p.s6.

Yamagishi, J. et al., 2016. Comparison of boiling and robotics automation method in DNA extraction for metagenomic sequencing of human oral microbes. *PLoS ONE*, 11(4), pp.1–15.

Zarco, M.F., Vess, T.J. & Ginsburg, G.S., 2012. The oral microbiome in health and disease and the potential impact on personalized dental medicine. *Oral Diseases*, 18(2), pp.109–120.

Zaura, E. et al., 2009. Defining the healthy "core microbiome" of oral microbial communities. *BMC microbiology*, 9, p.259.

Zaura, E. et al., 2017. On the ecosystemic network of saliva in healthy young adults. *ISME Journal*, 11(5), pp.1218–1231.

Zemanick, E.T. et al., 2015. Assessment of airway microbiota and inflammation in cystic fibrosis using multiple sampling methods. *Annals of the American Thoracic Society*, 12(2), pp.221–229.

Appendix

1. Table that summarizes published studies that used 16S rRNA gene sequencing to describe the airway microbiome.

Type of sample	Authors	Year	Type and number of patients	Significant findings
Nasopharyngeal and oropharyngeal	Charlson et al.	2010	62 healthy subjects – 29 smokers and 33 nonsmokers	Community composition exhibited minimal body site variation.
BAL samples	Erb-Downward	2011	7 healthy smokers, 3 never smokers and 4 patients with COPD	There is a core pulmonary microbiome that includes: *Pseudomonas*, *Streptococcus*, *Prevotella*, *Fusobacterium*, *Haemophilus*, *Veillonella*, and *Porphyromonas*. Patients with moderate and severe COPD have lower community diversity.
Upper tract: oral and oro-/nasopharyngeal swabs. Lower tract: glottis, BAL, and lower airway protected brush	Charlson et al.	2011	6 healthy subjects	The community composition is indistinguishable in upper and lower tract, although bacteria are present in the lungs of healthy subjects at lower levels compared to the upper respiratory tract.
Sputum, bronchial aspirate, Bronchiolar Alveolar Lavage (BAL) and bronchial mucosal	Cabrera-Rubio et al.	2012	6 patients with COPD	Sputum samples showed significant lower diversity than the other three. BAL and bronchial mucosal samples showed similar compositions in contrast with the other two samples.
Oral and Alveolar	Morris et al.	2013	64 healthy subjects, being 45 nonsmokers and 19 smokers	Most bacteria identified in alveolar lavages were also in the oral microbiome. Although specific bacteria (*Enterobacteriaceae*, *Haemophilus*, and *Methylobacterium*) appear in significant higher abundance in the lungs.
Oral, BAL, nasal	Bassis et al.	2015	28 healthy subjects	The bacterial communities of the lungs were significantly different from those of the nose and mouth. However, the bacterial communities of healthy lungs shared significant membership with the mouth but not with nose.
Oropharyngeal and sputum	Liu et al.	2017	114 samples from 4 patients with COPD, suffering from severe AECOD	Sputum microbiome is similar to the oropharyngeal microbiome.

2. Table that summarizes published studies that analysed the microbiome dynamics in AECOPD.

Authors	Year	Sampling	Significant findings
Huang et al.	2010	Endotracheal samples from 8 different patients at exacerbation	• Patients exhibiting communities with significantly fewer taxa, tended to have more members of Pseudomonadaceae. Patients with richer communities tend to have more members of Clostridiaceae, Lachnospiraceae, Bacillacee, and Peptostreptotaceae. • A core microbiome of 75 taxa, belonging to 27 families, was detected in all patients, including members of the Enterobacteriacee, Campylobacteraceae, and Helicobacteraceae family, as well as potential pathogens such as *A. cryaerophilus*, *B. diminuta*, *L. integrans*, and *P. aeruginosa*.
Huang et al.	2014	Sputum samples from 12 different patients were collected before, at onset of, and after exacerbation (2 weeks and 6 weeks after)	• At exacerbation, predominantly, members of Proteobacteria were enriched, compared with the samples collected at pre-exacerbation. • The major difference in the community composition was found at post treatment, when compared, with pre and at exacerbation. At this time point, the majority of taxa decreased in abundance, especially members of Proteobacteria. • Patients treated with antibiotics showed a reduced abundance of Proteobacteria. Patients treated with steroids and antibiotics showed and increase in Proteobacteria. Treatment with only steroids lead to an enrichment in Proteobacteria and members of Bacteriodetes and Firmicutes.
Millares et al.	2014	Sputum samples from 16 patients with COPD were collected before and at exacerbation	• Increase in abundance of taxa typically associated with COPD, at AECOPD, such as *Streptococcus*, *Pseudomonas*, *Moraxella*, *Haemophilus*, *Neisseria*, *Archromobacter* and *Corynebacterium*, compared with samples collected before.
Millares et al.	2015	Sputum samples from 8 severe patients with COPD collected before and at exacerbation	• No significant differences were found in the community composition among the two collection times.
Su et al.	2015	Sputum samples from 6 severe patients with COPD collected at exacerbation, and 7 to 16 days after exacerbation	• Bacterial community typically found in the patients with COPD: *Acinetobacter*, *Prevotella*, *Neisseria*, *Rothia*, *Lactobacillus*, *Leptotrichia*, *Streptococcus*, *Veillonella*, *Pasteurella*, *Klebsiella* and *Actinomyces*. • Microbial communities showed significant variation among patients, contributing to patterns almost always personalized to each patient.

Wang et al.	2016	Sputum samples from 8 different patients were collected before, at exacerbation, 2 weeks post and 6 weeks post-exacerbation	• At exacerbation, it was found an overall reduced alpha-diversity, with a non-significant increase abundance of Proteobacteria, and decrease in Firmicutes, compared with stable states. • It was found also differences in microbial composition, in eosinophilic and bacterial exacerbations. While bacterial exacerbations showed an overall reduced alpha-diversity and Firmicutes, and an increase of Proteobacteria. While a decreased Proteobacteria: Firmicutes ratio was found in the eosinophilic exacerbations. • Differences were also found when different treatments were applied. When just corticosteroids were use, it was found a decrease microbial alfa- diversity with increase in Proteobacteria and Firmicutes. The opposite trend was found in individuals treated with only antibiotics.
Liu et al.	2017	Oropharyngeal and sputum samples were collected from 4 different patients, at exacerbation and the following days of hospitalization (14-17 days)	• During AECOPD *Psychrobacter*, *Lactobacillus*, *Rothia*, *Prevotella*, *Neisseria*, *Streptococcus*, *Haemophilus*, *Actinomyces*, *Leptotrichia*, and *Aspargillus*, showed high relative abundance.

3. Informed Consent

Termo de Consentimento Livre e Esclarecido

Título do Projeto: "GENIAL – Marcadores genéticos e clínicos na trajetória da DPOC".

Nome do Investigador Principal: Prof. Doutora Alda Sofia Pires de Dias Marques

Por favor leia e assinale com uma cruz (X) os quadrados seguintes.

1. Eu confirmo que percebi a informação que me foi dada e tive a oportunidade de questionar e de me esclarecer. ☐

2. Eu percebo que a minha participação é voluntária e que sou livre de desistir, em qualquer altura, sem dar nenhuma explicação, sem que isso afete qualquer serviço de saúde ou qualquer outro que me é prestado. ☐

3. Eu compreendo que os dados recolhidos durante a investigação são confidenciais e que só os investigadores do projeto da Universidade de Aveiro têm acesso a eles. Portanto, dou autorização para que os mesmos tenham acesso a esses dados. ☐

4. Eu compreendo que os dados recolhidos durante o estudo podem ser utilizados para publicação em Revistas Científicas e usados noutras investigações, sem que haja qualquer quebra de confidencialidade. Portanto, dou autorização para a utilização dos dados para esses fins. ☐

5. Eu concordo então em participar no estudo. ☐

Nome da pessoa	Data	Assinatura
Nome do Investigador(a)	Data	Assinatura

4. List of all tests collected to clinically characterize the patients

Characteristics

- Age (years)
- Gender
- GOLD classification
- Smoking Status
- Other comorbidities

Lung Function Tests

- Predicted FEV_1
- FVC
- FEV_1
- FVC
- FEV_1/FVC

Muscular and Functional Tests

- Maximum Inspiratory Pressure
- Maximum Expiratory Pressure
- Quadriceps muscle strength
- Handgrip
- 5 repetition sit-to-stand
- 1 minute sit-to-stand

Patient Reported Outcomes

- Frequency of exacerbations
- Dyspnea
    - Modified British Medical Research Council questionnaire
    - Borg Scale
- Fatigue
- Anxiety Score
    - Hospital Anxiety and Depression Scale M
- Depression Score
    - COPD Assessment Test M, (IQR)
    - St. George Respiratory Questionnaire
- Symptoms score
- Activity score
- Impact score

5. Volume of saliva and DNA concentration per sample

Table 3. Volume of saliva and DNA concentration, given by Qubit, per sample.

Patient	Stable (pre exacerbation)	Exacerbation	Recovery		
			Post 2 weeks	Post 6 weeks	Post 8 months
519		Volume of Saliva: 1 ml DNA concentration: 19,36 ng/µL	Volume of Saliva: 1 ml DNA concentration: 31,8 ng/µL	Volume of Saliva: 1 ml DNA concentration: 94,8 ng/µL	
526		Volume of Saliva: 1 ml DNA concentration: 42,8 ng/µL	Volume of Saliva: 1 ml Qubit: 81,0 ng/µL		Volume of Saliva: 1 ml DNA concentration: 22,6 ng/µL
555		Volume of Saliva: 500 µL DNA concentration: 63,8 ng/µL	Volume of Saliva: 1 ml DNA concentration: 78,6 ng/µL		
S29		Volume of Saliva: 500 µL DNA concentration: 11,4 ng/µL		Volume of Saliva: 1 ml DNA concentration: 27,2 ng/µL	
E44		Volume of Saliva: 500 µL DNA concentration: 12 ng/µL		Volume of Saliva: 1 ml DNA concentration: 37,6 ng/µL	
587		Volume of Saliva: 500 µL DNA concentration: 12,2 ng/µL	Volume of Saliva: 1 ml DNA concentration: 24,8 ng/µL	Volume of Saliva: 1 ml DNA concentration: 60,6 ng/µL	
616	Volume of Saliva: 1 ml DNA concentration: 2,06 ng/µL	Volume of Saliva: 1 ml DNA concentration: 30,6 ng/µL		Volume of Saliva: 1 ml DNA concentration: 30,2 ng/µL	

6. Taxa plots at phyla level

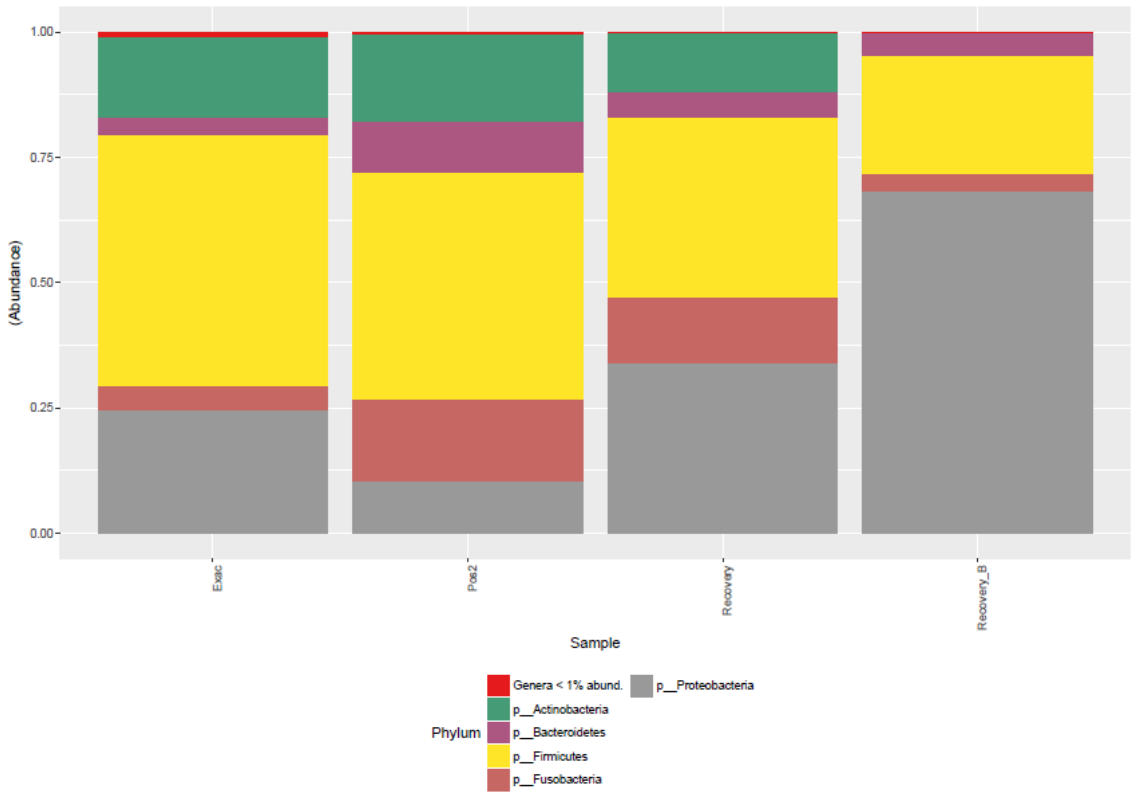


Figure 1. - Samples have been grouped and averaged by time of collection – exacerbation (Exac), 2 weeks after exacerbation (Pos2), and 6 weeks (Recovery) and 8 months) post exacerbation (Recovery_B). Taxonomic composition was shown at phyla level. Each column in the plot represents a time of sample collection, and each color in the columns represents the relative abundance of the total sample divided by different phyla. Across the four sampling times, the most abundant phyla are Firmicutes, Proteobacteria, and Actinobacteria.

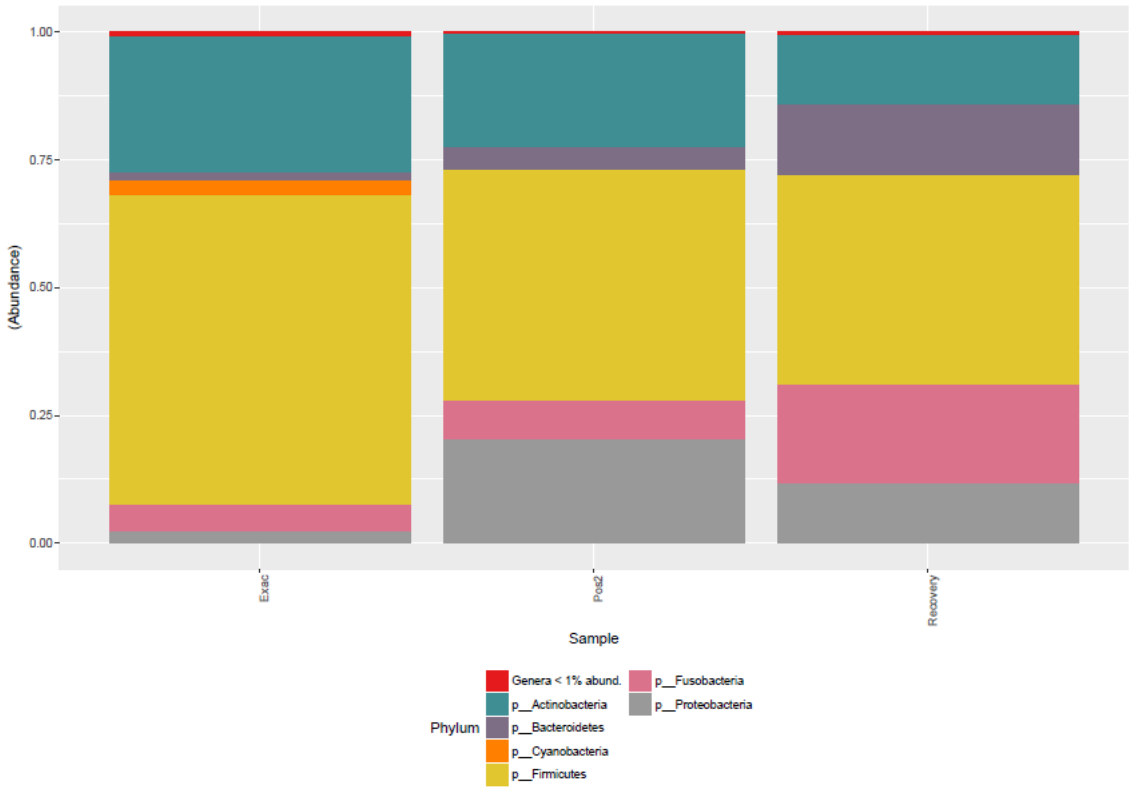


Figure 2. - Samples collect from the subject 519 at different time points – exacerbation (Exac), 2 weeks after exacerbation (Pos2), and 6 weeks post exacerbation after clinical recovery (Recovery). Taxonomic composition is shown at the phyla level. Each column in the plot represents a time point of sample collection, and each color in the columns represents the relative abundance of each taxon group at phyla level. Across the three samples, the most abundant phyla were Firmicutes, Actinobacteria and Fusobacteria. The relative abundance of Firmicutes decreases after “Exac”, and remains constant in “Pos2” and “Recovery”.

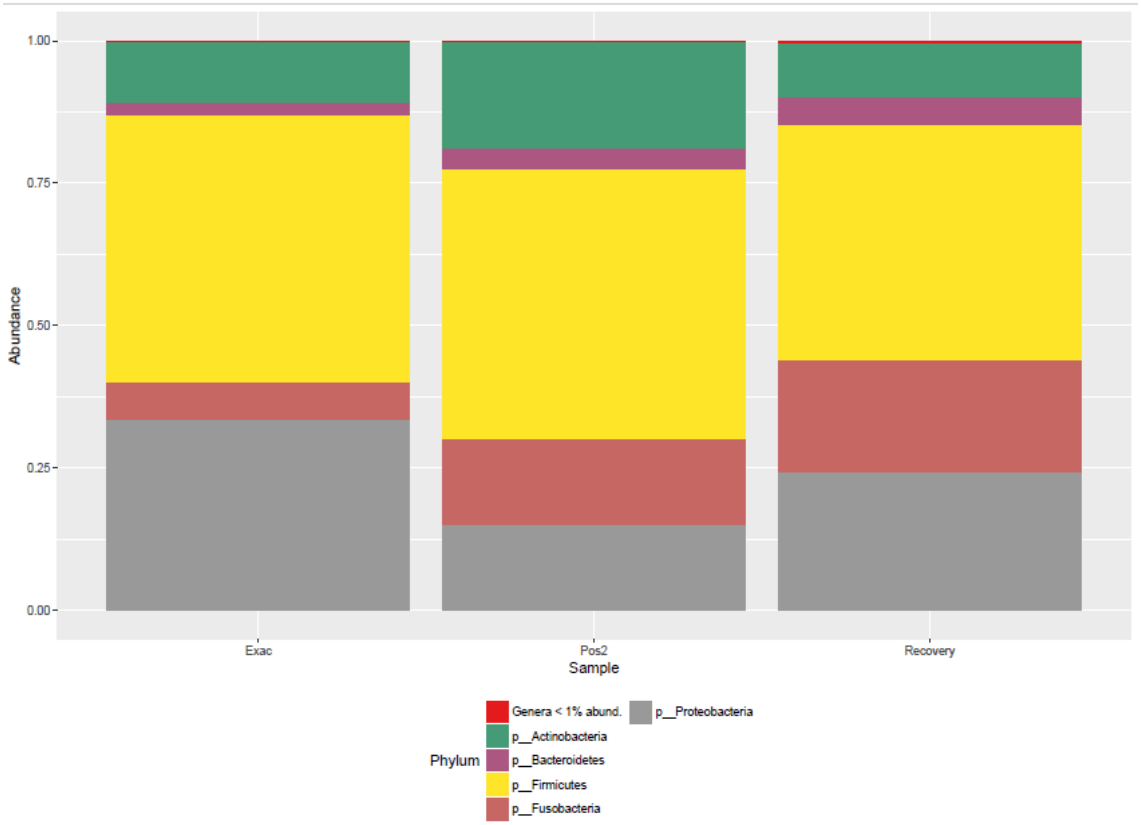


Figure 3. - Samples collect from the subject 587 at different time points – exacerbation (exac), 2 weeks after exacerbation (pos2), and 6 weeks post exacerbation at clinically recover (recovery) - taxonomic composition is shown at phyla level. Each column in the plot represents a time of sample collection, and each color in the column represents the relative abundance of each taxon group at phyla level. The phyla Firmicutes, Proteobacteria, and Fusobacteria had the higher relative abundance across the three times of collection. The relative abundance of Firmicutes is remarkably similar in the three samples.

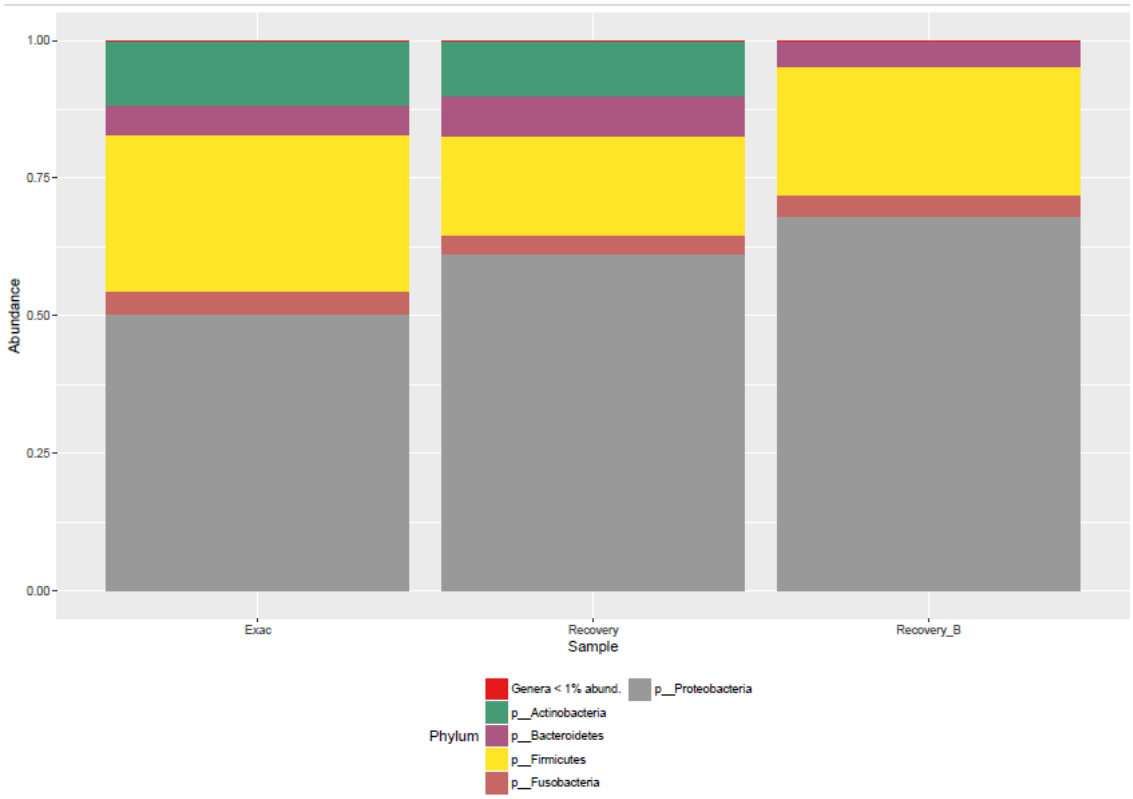


Figure 4. - Samples collect from the subject 526 at different time points – exacerbation (exac), 6 weeks post exacerbation at clinically recover (recovery), and 8 months after exacerbation (recovery_B) - taxonomic composition is shown on phyla level. Each column in the plot represents a time of sample collection, and each colour in the column represents the relative abundance of each taxon group at phyla. Proteobacteria and Firmicutes were the phyla with higher relative abundances. At “Exac” Proteobacteria represents 50% of all the phyla at the sample, and this percentage increases over time. The phyla composition of the two samples collected at recovery is rearkably different.

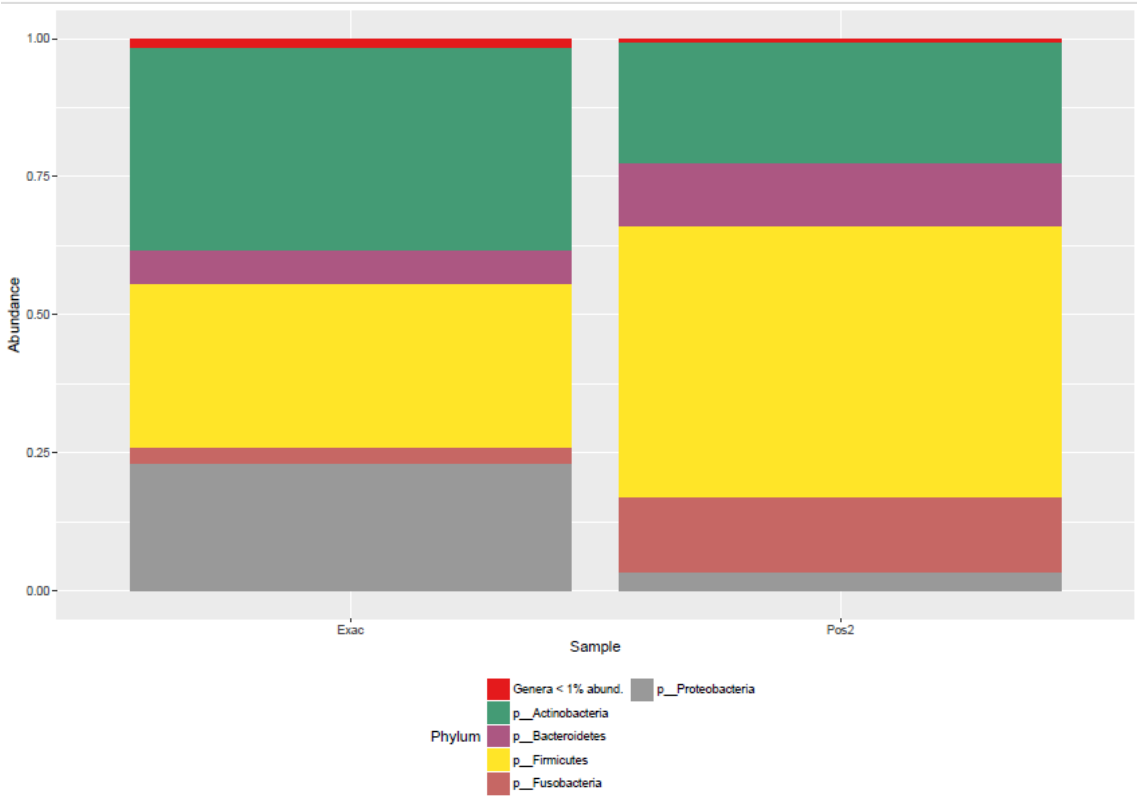


Figure 5. - Samples collect from the subject 555 at two different time points – exacerbation (Exac), 2 weeks after exacerbation (Pos2) - taxonomic composition is shown on phyla level. Each column in the plot represents a time of sample collection, and each color in the column represents the relative abundance of each taxon group at phyla level. Firmicutes, Actinobacteria, and Proteobacteria are the main phyla across the two samples. The relative abundance of Firmicutes increases at “Pos2”, while the relative abundances of Actinobacteria and Proteobacteria decrease over time.

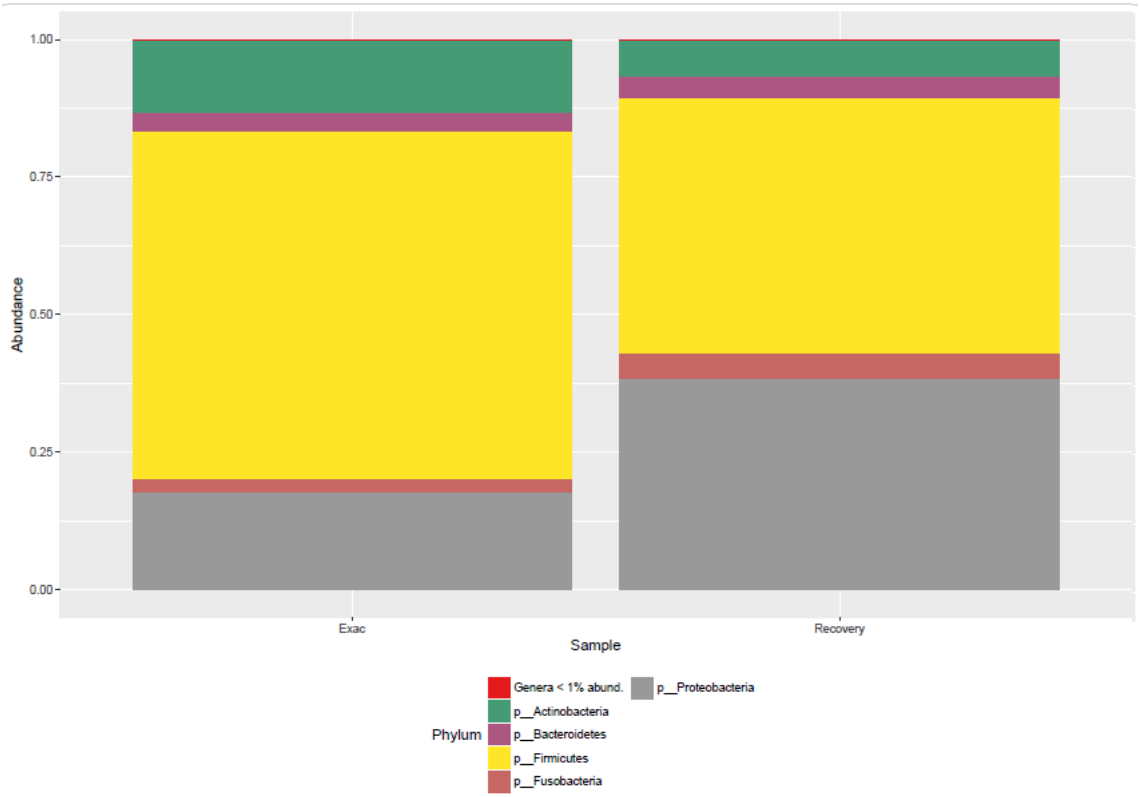


Figure 6. - Samples collect from the subject E44 at two different time points – exacerbation (Exac) and 6 weeks post exacerbation at clinically recovery (Recovery) - taxonomic composition is shown on phyla level. Each column in the plot represents a time of sample collection, and each colour in the column represents the relative abundance of each taxon group at phyla level. The most abundant phyla were Firmicutes and Proteobacteria. The relative abundance of Firmicutes decreases at “recovery”, while the relative abundance of Proteobacteria increases

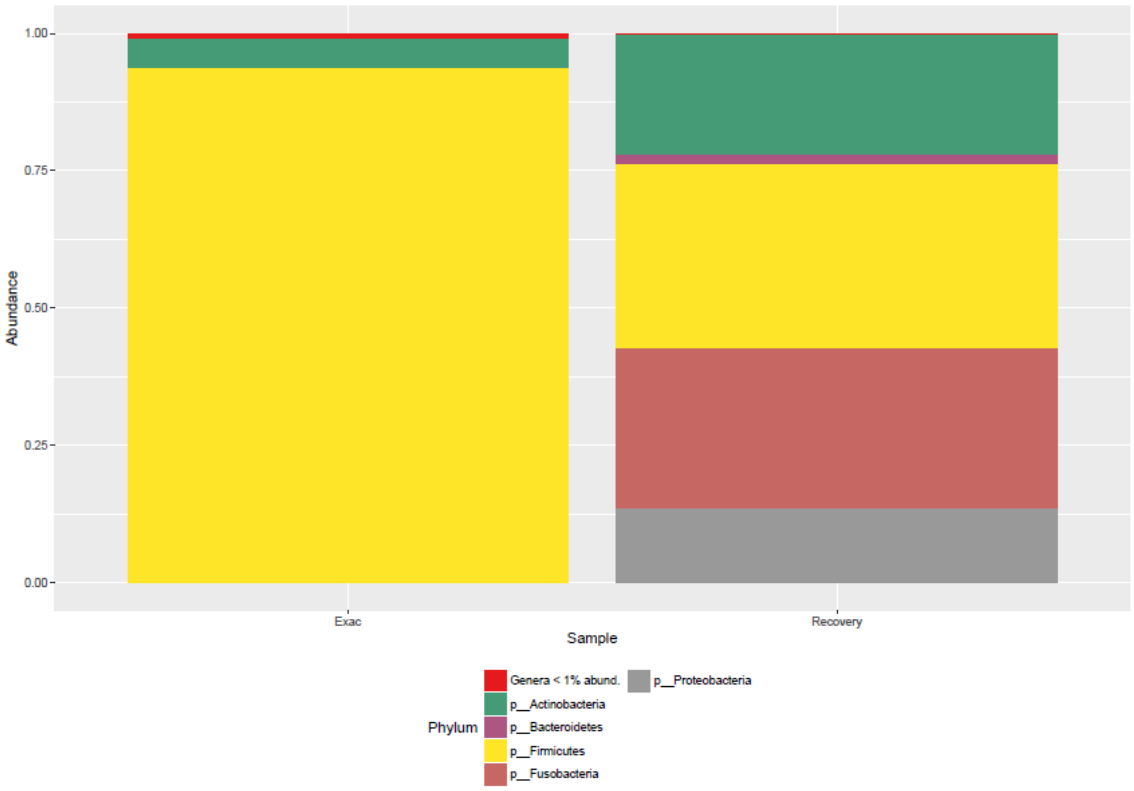


Figure 7. - Samples collect from the subject 616 at two different time points – exacerbation (Exac) and 6 weeks post exacerbation at clinically recover (Recovery) - taxonomic composition is shown on phyla level. Each column in the plot represents a time of sample collection, and each colour in the column represents the relative abundance of each taxon group at phyla level. The phyla with higher relative abundances at the two collection times were Firmicutes, Proteobacteria, and Actinobacteria. At “Exac” Firmicutes is almost the only phyla present in the sample, while at “Recovery” the sample is much more besides Firmicutes has also Proteobacteria, Fusobacteria and Actinobacteria.

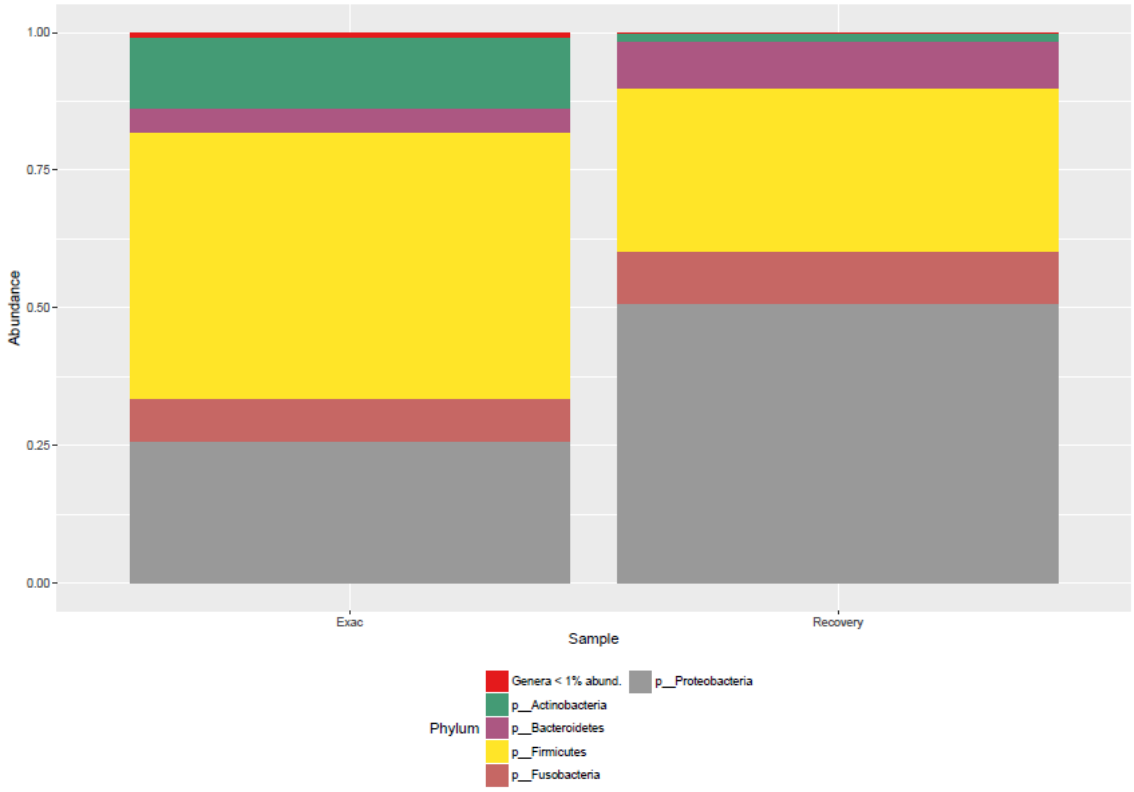


Figure 8. - Samples collect from the subject S29 at two different time points – exacerbation (Exac) and 6 weeks post exacerbation at clinically recover (Recovery) - taxonomic composition is shown at the genera level. Each column in the plot represents a time of sample collection, and each colour in the column represents the relative abundance of each taxon group at genera level. Proteobacteria, Firmicutes, and Fusobacteria had the higher relative abundances at the two sampling times. The relative abundance of Proteobacteria increases over time, while Firmicutes decreases.

7. Statistical analysis per taxon at different collection time points

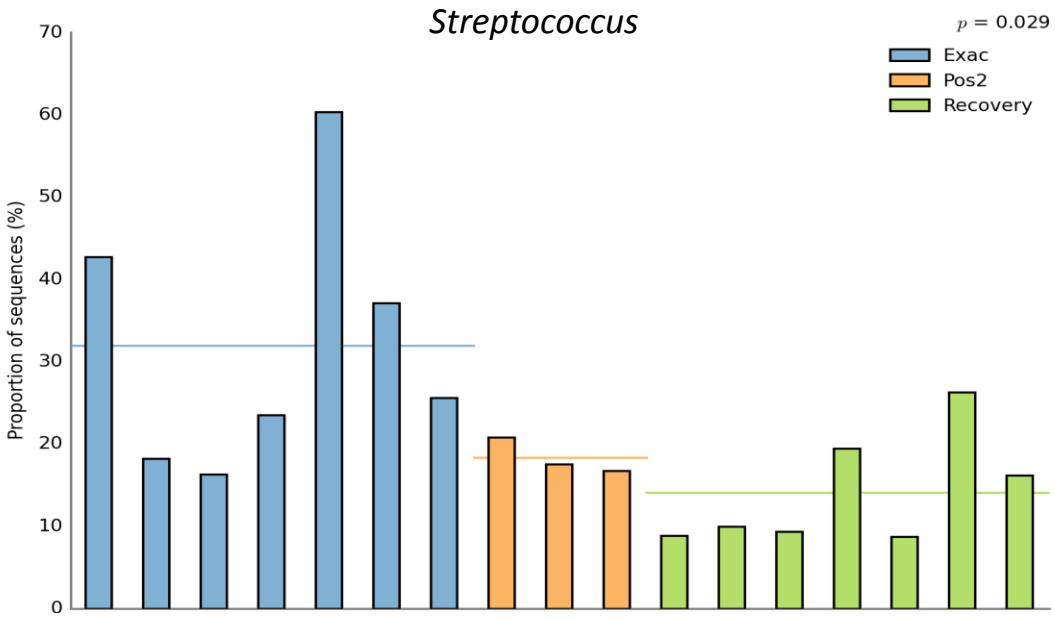


Figure 1. – Bar plot showing the significant microbial composition shifts over time (exacerbation – “Exac”, two weeks after exacerbation “Pos2” and 6 weeks and 8 months after exacerbation – “Recovery”), using ANOVA with the post-hoc test: Games-Howell. Over time there was a decrease in the relative abundance of *Streptococcus*.

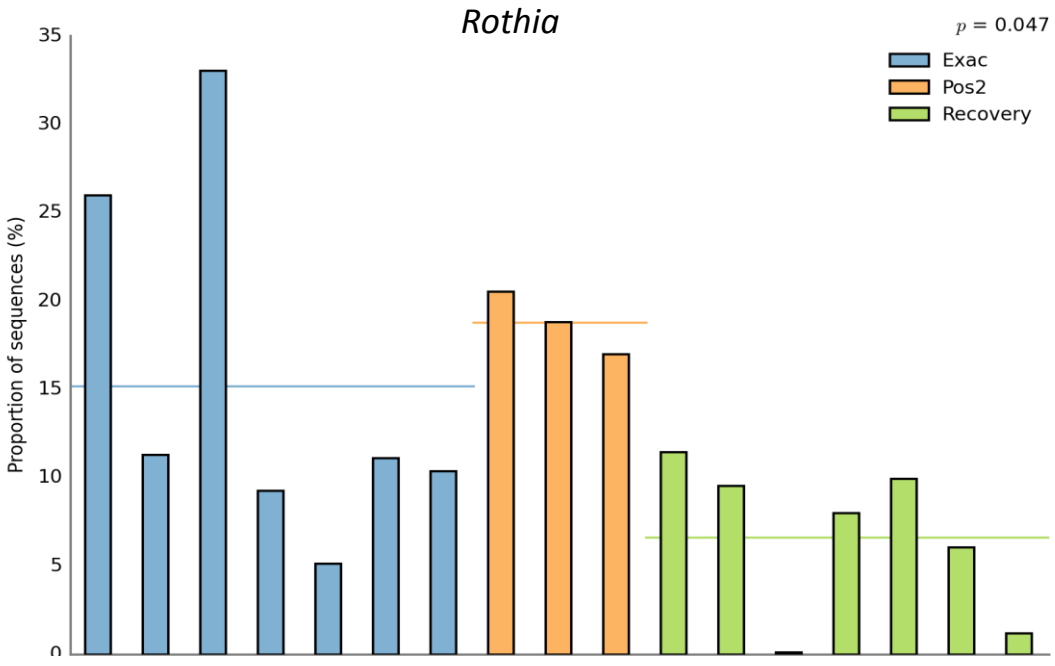


Figure 2. – Bar plot showing the significant microbial composition shifts over time (exacerbation – “Exac”, two weeks after exacerbation “Pos2” and 6 weeks and 8 months after exacerbation – “Recovery”), using ANOVA with the post-hoc test: Games-Howell. There was an increase in the relative abundance of *Rothia* at “Pos2”.

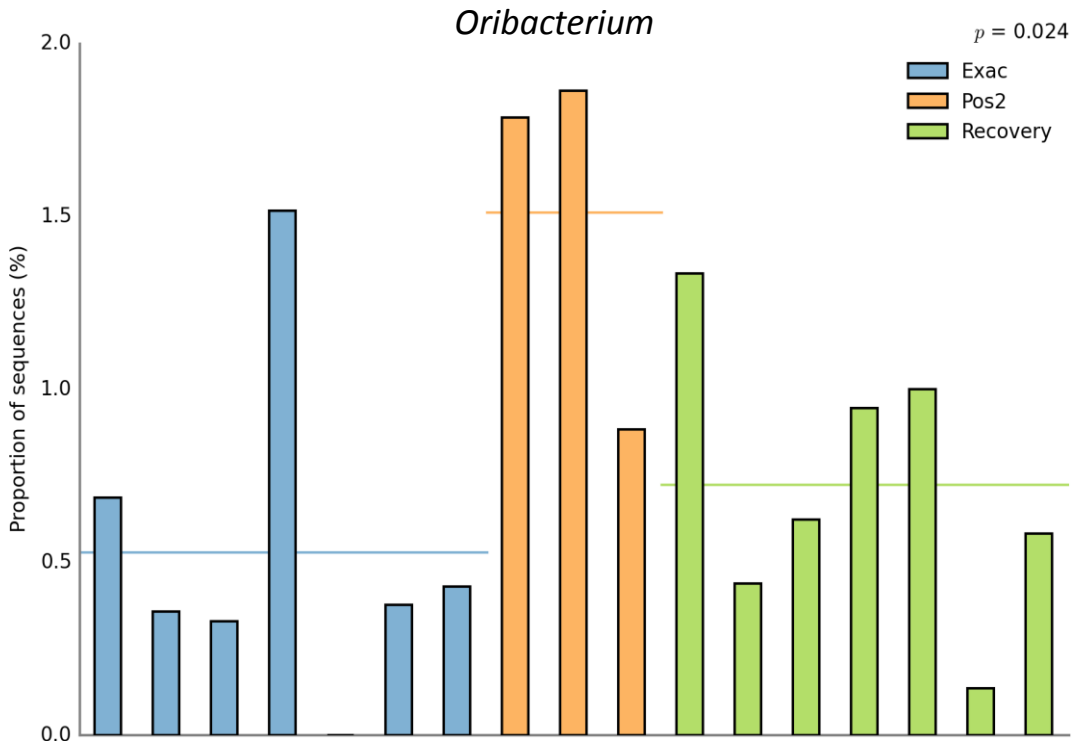


Figure 3. – Bar plot showing the significant microbial composition shifts over time (exacerbation – “Exac”, two weeks after exacerbation “Pos2” and 6 weeks and 8 months after exacerbation – “Recovery”), using ANOVA with the post-hoc test: Games-Howell. There was an increase in the relative abundance of *Oribacterium* at “Pos2”.

8. Statistical analysis per taxon at two different time points (phyla level)

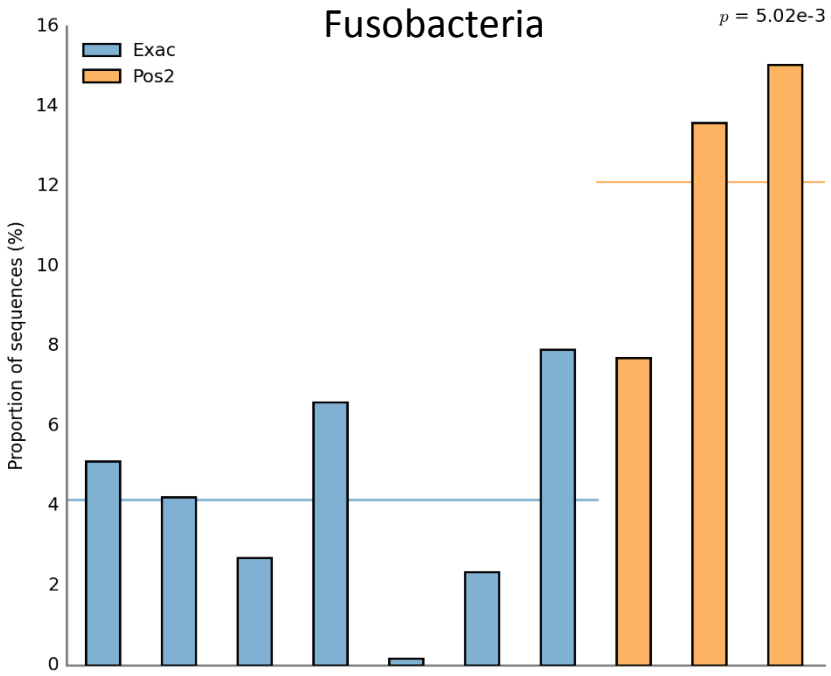


Figure 1. – Bar plot showing pairwise comparisons of the samples collected at exacerbation (“Exac”) and two weeks after exacerbation (“Pos2”), using two sided t-test. At “Pos2” there was an increase in the relative abundance of Proteobacteria.

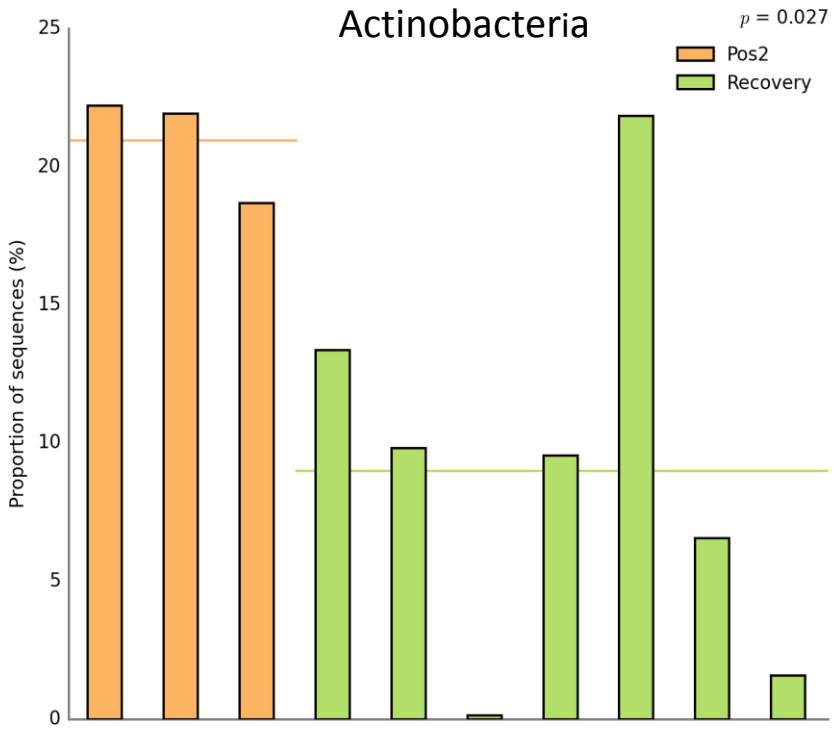


Figure 2 - Bar plot showing pairwise comparisons of the samples collected at two weeks after exacerbation (“Pos2”) and 6 weeks and 8 months after exacerbation (“Recovery”), using two sided t-test. At “Recovery” there was a decrease in the relative abundance of Actinobacteria.

9. Statistical analysis per taxon at two different time points (genera level)

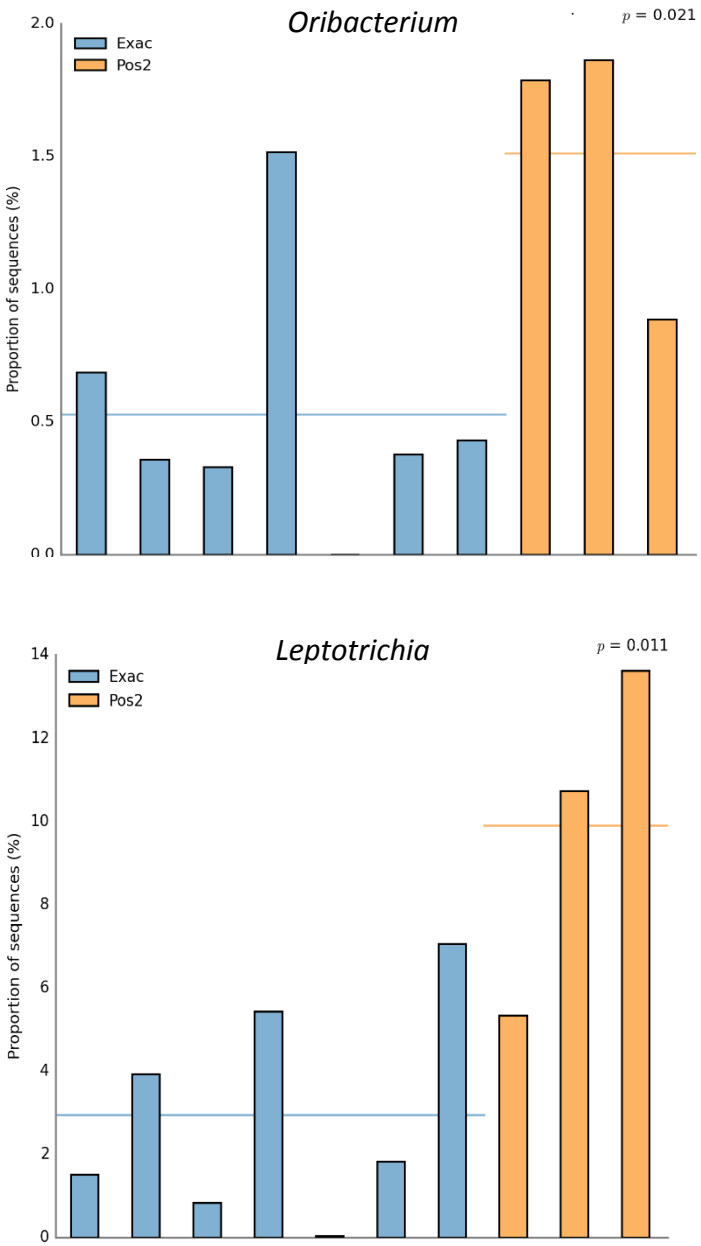


Figure 1 - Bar plot showing pairwise comparisons of the samples collected at exacerbation (“Exac”) and two weeks after exacerbation (“Pos2”), using two sided t-test. At “Pos2” there were increases in the relative abundances of *Oribacterium* and *Leptotrichia*.

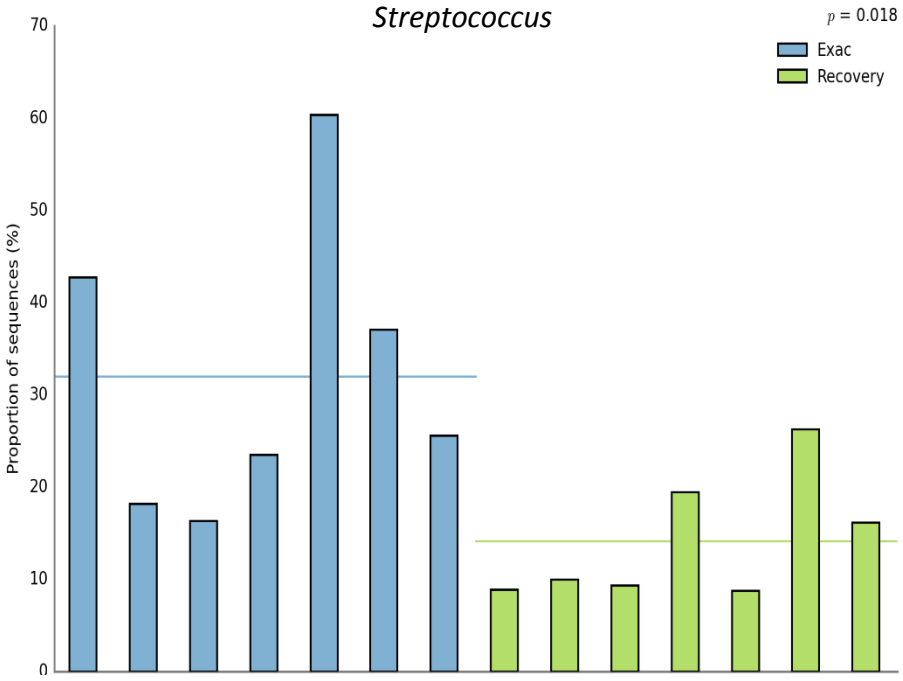


Figure 2 - Bar plot showing pairwise comparisons of the samples collected at exacerbation (“Exac”) and 6 weeks and 8 months after exacerbation (“Recovery”), using two sided t-test. At “Recovery” there was a decrease in the relative abundance of *Streptococcus*.

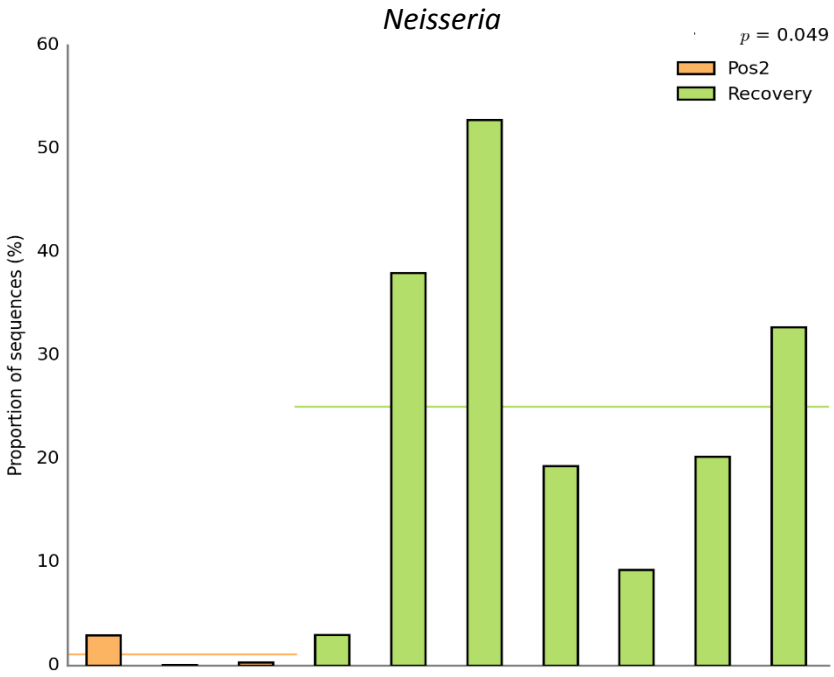


Figure 3 - Bar plot showing pairwise comparisons of the samples collected two weeks after exacerbation (“Pos2”) and 6 weeks and 8 months after exacerbation (“Recovery”), using two sided t-test. At “Recovery” there was an increase in the relative abundance of *Neisseria*.

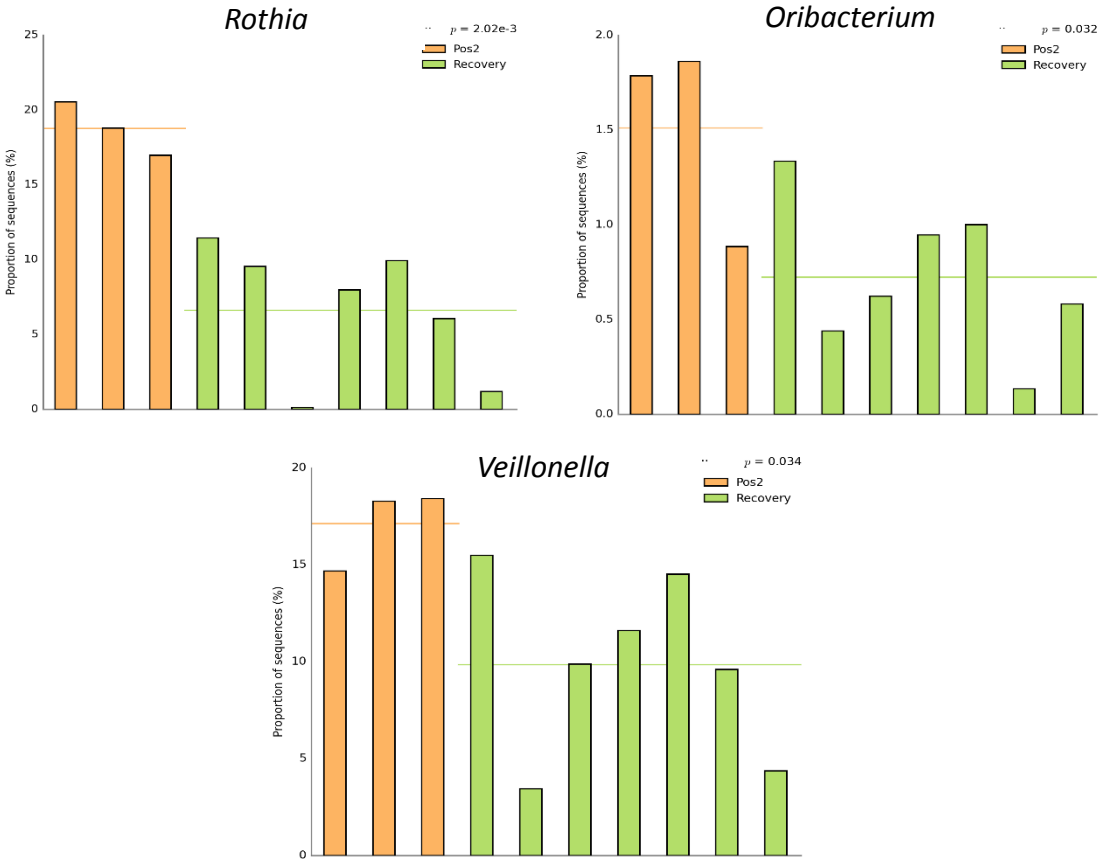


Figure 4 - Bar plot showing pairwise comparisons of the samples collected two weeks after exacerbation ("Pos2") and 6 weeks and 8 months after exacerbation ("Recovery"), using two sided t-test. At "Recovery" there were a decrease in the relative abundances of *Rothia, Oribacterium, and Veillonella.*