Universidade de Aveiro Departamento de Ciências Médicas
2019

Bárbara Cristiana Pereira de Andrade

Efeito da reabilitação respiratória na microbiota pulmonar detetada na saliva de doentes com DPOC.

Effects of pulmonary rehabilitation on airway microbiota detected in the saliva of patients with COPD.

Universidade de Aveiro Departamento de Ciências Médicas
2019

Bárbara Cristiana Pereira de Andrade

Effects of pulmonary rehabilitation on airway microbiota detected in the saliva of patients with COPD.

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Biomedicina Molecular, realizada sob a orientação científica da Doutora Alda Sofia Pires de Dias Marques, Professora Adjunta da Escola Superior de Saúde da Universidade de Aveiro, e sob coorientação científica da Doutora Ana Margarida Domingos Tavares de Sousa, Professora Auxiliar Convidada do Departamento de Ciências Médicas da Universidade de Aveiro

Esta dissertação integrou-se nos projetos PRISMA - Pulmonary Rehabilitation: a response for patients with COPD in an InduStrialized environment and its implication on lung MicrobiotA, financiado pelo International Observatory Hommes-Milieux Estarreja (OHMI Estarreja) e PRIME - Pulmonary Rehabilitation Innovation and Microbiota in Exacerbations of COPD e iBiMED, ambos financiados pelo Programa Operacional de Competitividade e Internacionalização – POCI através do Fundo Europeu de Desenvolvimento Regional - FEDER (POCI-01-0145-FEDER-028806 e POCI-01-0145-FEDER-007628) e Fundação para a Ciência e Tecnologia (PTDC/SAU-SER/28806/2017 e UID/BIM/04501/2019).

o júri

presidente	Prof. Doutora Margarida Sâncio da Cruz Fardilha professora Auxiliar c/ Agregação, Universidade de Aveiro
arguente	Prof. Doutora Isabel da Silva Henriques professora auxiliar do Departamento de Ciências da Vida da Faculdade de Ciências e Tecnologia da Universidade de Coimbra
co-orientadora	Prof. Doutora Ana Margarida Domingos Tavares de Sousa professora auxiliar convidada do Departamento de Ciências Médicas da Universidade de Aveiro

agradecimentos

A concluir a dissertação de mestrado dedico o meu profundo agradecimento:

À Professora Doutora Alda Marques, pela oportunidade de trabalhar na sua equipa, pela sua disponibilidade constante, pelo apoio e orientação.

À Professora Doutora Ana Margarida Sousa, pela coorientação e apoio neste trabalho e pelos conhecimentos transmitidos.

Aos doentes que participaram neste estudo, por toda a paciência, compreensão e disponibilidade. Sem vocês este trabalho não era possível. São o verdadeiro exemplo de que com força e motivação tudo se consegue.

À equipa do Lab3R pela ajuda, pela disponibilidade e pelo apoio ao longo deste percurso.

À Sara Dias que que foi um grande apoio desde o primeiro dia. Obrigada pelo companheirismo, pela amizade e pela enorme ajuda que foste.

Às “migas” e a todos os meus amigos, pelo apoio, pela enorme paciência e pela amizade.

À minha mãe, que nunca me deixou ir abaixo, me apoiou e incentivou desde sempre e nunca me deixou desistir. Um obrigado do tamanho do mundo para ti. Ao meu irmão, que é o melhor companheiro de vida e, como sempre, esteve de mãos dadas comigo em mais uma caminhada. Ao meu pai, pelo apoio e preocupação. A todo o resto da família, pela paciência, por todo o amor, pelo apoio. São a melhor família do mundo!

Ao Roberto, pela compreensão, por acreditar mais em mim que eu e por estar comigo incondicionalmente. Obrigada por tudo o que acrescentas à minha vida.

Às minhas estrelinhas, que estariam tão orgulhosos. Sei que estiveram e estão sempre comigo.

A todos, obrigada do fundo do coração!

palavras-chave

doença pulmonar obstrutiva crónica, DPOC, microbiota, reabilitação respiratória, 16s rRNA.

resumo

A doença pulmonar obstrutiva crónica (DPOC) é a terceira principal causa de mortalidade no mundo. A Reabilitação Respiratória (RR), uma intervenção compreensiva que, inclui várias componentes sendo uma delas o exercício físico, é a terapia mais custo-efetiva para os doentes com DPOC.

O exercício físico aumenta a ventilação e captação de oxigénio, o que provavelmente influencia a microbiota das vias aéreas. No entanto, a forma como esta influência ocorre é ainda pouco compreendida. Este estudo teve como objetivo estudar o impacto da RR na microbiota pulmonar em doentes com DPOC explorando a composição microbiana, alfa e beta diversidade. Foram recolhidos dados sociodemográficos, antropométricos, clínicos e amostras de saliva (uma vez por mês) num grupo de doentes durante um período de ~9 meses (~3 meses antes da RR, 3 meses durante a RR e 3 meses após a RR) e noutro grupo de doentes durante um período de 6 meses (sem RR). A microbiota da saliva foi caracterizada pelo sequenciamento de 16s rRNA e analisada usando o pipeline QIIME2.

Participaram no estudo 25 doentes com DPOC que realizaram RR (19♂, 73±6y, FEV1pp 48±15) e 5 doentes que nunca realizaram RR (5♂, 75±6y, FEV1pp 48±13).

Observou-se um aumento significativo do filo Proteobacteria e do género *Neisseria* do período pré-RR para durante a RR. O LefSe mostrou que a microbiota das amostras do periodo pré-RR em comparação com o período durante a RR são enriquecidas nos géneros *Pseudomonas* e *Shingomonas* e a microbiota das amostras do período RR em comparação com o período pré-RR são enriquecidas em *Neisseria* e *Alloscardovia*. Comparando entre os períodos RR e pós-RR, o LEfSe apontou os géneros *Granulicatella* e *Acinetobacter* como sendo enriquecidas nas amostras do período RR e os géneros *Staphylococcus*, *Selenomonas* e a família Pasteurellaceae como sendo enriquecida no período pós-RP. Comparando o pré-PR com o período pós-PR, LEfSe apontou os géneros *Granulicatella*, *Sphingomonas*, *Pseudomonas* e *Enhydrobacter* como sendo enriquecidas no período pré-PR e a família Pasteurellaceae como sendo enriquecida no período pós-PR.

Não foram observadas diferenças significativas na alfa-diversidade comparando os diferentes períodos. O modelo LME mostrou que o tempo tem um impacto significativo na alfa-diversidade e que a interação entre a RR e passagem no tempo contribui para a dinâmica da microbiota, em métricas não filogenéticas. A PERMANOVA mostrou que a microbiota não converge por período.

Em conclusão, a RR não parece alterar significativamente a estrutura da microbiota (alfa-diversidade), mas sim a sua composição. Em geral, a alfa-diversidade é influenciada pela passagem do tempo e não pela RR. As análises da beta-diversidade mostram que não há convergência da microbiota durante nem apos a RR. Em métricas não filogenéticas, a dissemelhança da microbiota é influenciada pela interação entre a RR e a passagem do tempo.

keywords

chronic obstructive pulmonary disease, COPD, microbiota, pulmonary rehabilitation, 16s rRNA.

abstract

Chronic Obstructive Pulmonary Disease (COPD) is the third leading cause of mortality worldwide. Pulmonary Rehabilitation (PR), a comprehensive intervention that includes several components, one of which is exercise training, is the most cost-effective therapy for patients with COPD.

Exercise training increases ventilation and oxygen uptake, which most likely influences airway microbiota. However, how this influence occurs is still poorly understood. This study aimed to study the impact of RR on pulmonary microbiota in patients with COPD by exploring the microbial composition, alpha and beta diversity. Sociodemographic, anthropometric, clinical and saliva samples (once a month) were collected from a group of patients over a period of ~ 9 months (~ 3 months before RR, 3 months during RR and 3 months after RR) and in another patient group for a period of 6 months (without RR). Saliva microbiota was characterized by 16s rRNA sequencing and analyzed using the QIIME2 pipeline.

Twenty-five patients with COPD who underwent PR (19♂, 73±6y, FEV1pp 48±15) and 5 patients who never had PR (5♂, 75±6y, FEV1pp 48±13) participated in the study.

A significant increase of Proteobacteria phylum and *Neisseria* genus from pre-PR to PR period was observed. LEfSe showed that pre-PR comparing with during PR, samples' microbiotas are enriched in the genera *Pseudomonas* and *Shingomonas* and during PR samples' microbiotas are enriched in *Neisseria* and *Alloscardovia*. When comparing the periods during PR with post-PR, LEfSe pointed *Granulicatella* and *Acinetobacter* as being enriched during PR period and *Staphylococcus*, *Selenomonas* and Pasteurellaceae family as being enriched in the post-PR period. Comparing pre-PR with post-PR periods, LEfSe pointed *Granulicatella*, *Sphingomonas*, *Pseudomonas* and *Enhydrobacter* as being enriched in the pre-PR period and Pasteurellaceae family as being enriched in the post-PR period. No significant differences were observed in alfa diversity when comparing the different periods. LME model showed that time has a significant impact in alfa-diversity and that the interaction between PR and time passage are mainly contributing for microbiota dynamics, in non-phylogenetic metrics. PERMANOVA showed that microbiota does not converge per period.

In conclusion, PR does not seem to significantly alter the structure of microbiota (alpha diversity) but changes composition. In general, alpha diversity is influenced by the passage of time and not by PR. Beta-diversity analyses showed that there is no microbiota convergence during or after PR. In non-phylogenetic metrics, microbiota dissimilarity is impacted by interaction between PR and time passage.

Table of Contents

List of figures

List of tables

List of abbreviations and/or acronyms

5STS	Five times sit-to-stand test
ANCOM	Analysis of composition of microbiomes
ASVs	Amplicon sequence variants
ATS	American Thoracic Society
BAL	Bronchoalveolar lavage
BMI	Body mass index
CAT™	COPD Assessment Test
COPD	Chronic obstructive pulmonary disease
DALY	Disability-Adjusted Life Years
DNA	Deoxyribonucleic acid
ERS	European Respiratory Society
FEV_1	Forced expiratory volume in one second
FEV_{1pp}	Forced expiratory volume in 1 second percent predicted
FVC	Forced vital capacity
GOLD	Global Initiative for Chronic Obstructive Lung Disease
ICF	International Classification of Functioning Disability and Health
LefSe	Linear discriminant effect size
LME	Linear Mixed Effects
mMRC	Modified British Medical Research Council
OTU	Operational taxonomic unit
PCoA	Principle Coordinate Analysis

PERMANOVA	Permutational multivariate analysis of variance
PR	Pulmonary rehabilitation
rRNA	Ribosomal ribonucleic acid
SpO_2	Peripheral capillary oxygen saturation
STS	1-minute sit-to-stand
YLD	Years lived with disability
WHO	World Health Organization

1. Introduction

1.1. Chronic obstructive pulmonary disease

1.1.1. Epidemiology and overall impact

According to the World Health Organization (WHO), chronic obstructive pulmonary disease (COPD) is the third leading cause of death worldwide [1]. The global prevalence of COPD was estimated at about 384 million cases in 2010 [2] and causes about three million deaths a year [2]. In Portugal, the estimated prevalence of COPD is 14,2% for Portuguese over 40 years old [3] and it has been reported to affect 800.000 people [3].

This disease is more prevalent in smokers and ex-smokers than in non-smokers and it increases with the advancing of age [2]. Contrary to what has been previously described, it is now known that the prevalence of COPD is almost equal in men and women, probably due to changes in patterns of smoking habits [2], since these habits have become similar regardless of sex.

COPD represents a significant health, economic and social burden [4]. Factors that most contribute to COPD burden are disease severity, presence of frequent exacerbations of disease and the presence of comorbidities [4]. In the European Union, the cost associated with COPD is 38.6 billion Euros, 56% of the cost of respiratory disease [2]. Hospitalisations due to exacerbations represent the biggest contributor for total COPD economic burden on the healthcare system [4].

COPD is an increasing contributor to disability and mortality around the world [4]. Chronic respiratory diseases account for 6.3% of global Years lived with disability (YLD), with the largest contributor being COPD [4]. Chronic respiratory diseases accounted also for 4.7% of global Disability-Adjusted Life Years (DALY) [4], sum of years of potential life lost because of premature mortality and years of productive life lost due to disability [2].

1.1.2. Definition, diagnosis and clinical manifestation

Chronic obstructive pulmonary disease is defined by the Global Initiative for Chronic Obstructive Lung Disease (GOLD) [2] as "a common, preventable and treatable disease that is characterised by persistent respiratory symptoms and airflow limitation that is due to airway and/or alveolar abnormalities usually caused by significant exposure to noxious particles or gases."

Exposure to noxious inhalants causes chronic inflammation, characteristic of this disease, and that is responsible for structural alterations, narrowing of the small airways and destruction of the pulmonary parenchyma that leads to the loss of alveolar attachments to the small airways and decreased lung elastic recoil [2] (Fig. 1). This leads to reduced ability of the airways to remain open during expiration, airflow limitation and mucociliary dysfunction. [2].

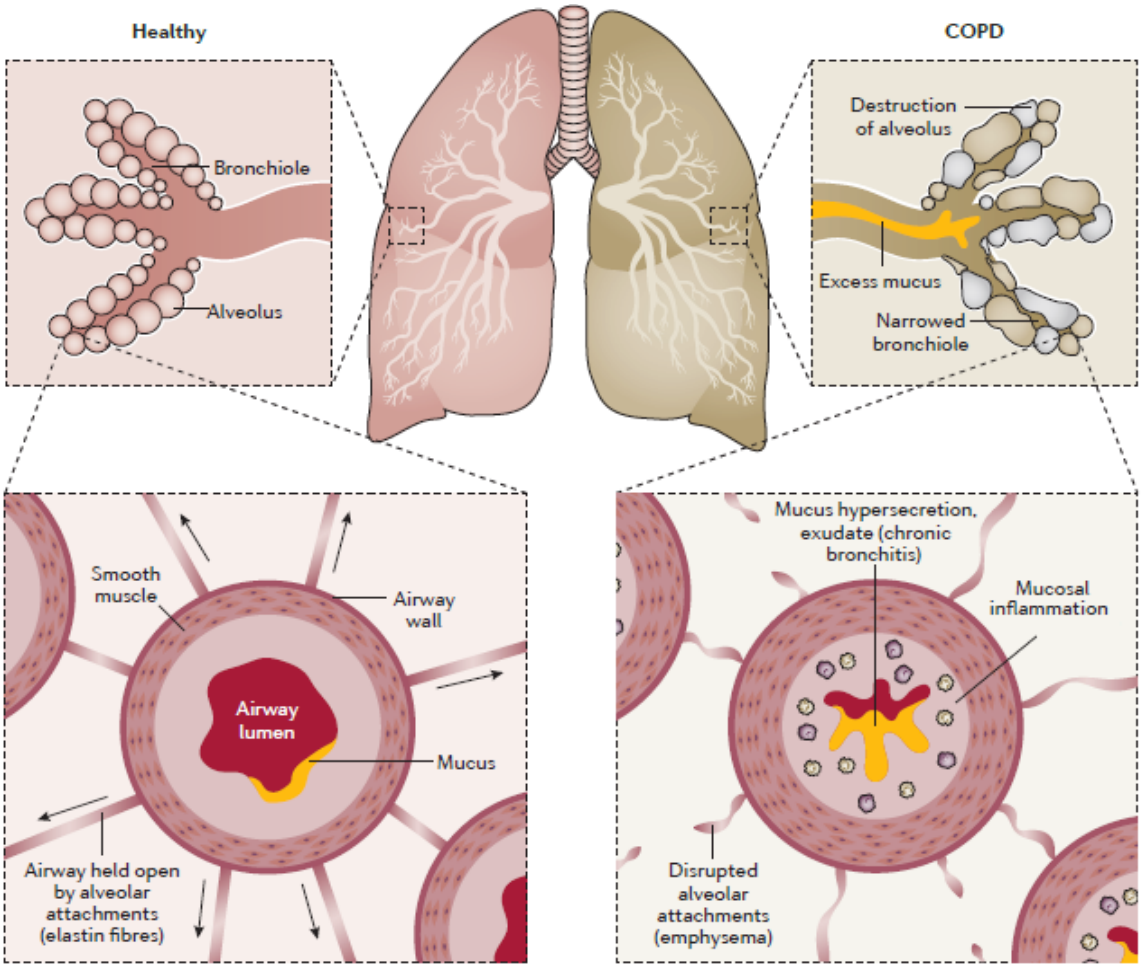

Figure 1. In healthy lungs, the small airways are kept open by alveolar wall attachments. In chronic obstructive pulmonary disease (COPD), the small airways are narrowed through thickening of the

bronchiolar periphery wall by inflammation, disruption of alveolar attachments as a result of emphysema and luminal occlusion by mucus and inflammatory exudate. Adapted by: [5]

Diagnosis of COPD is based on symptoms and risk factors associated with a simple lung function test that evaluate airflow limitation, called spirometry (Fig. 2). Relevant variables measured are forced vital capacity (FVC), the volume of air that can be forced to exhale after a complete inspiration, and forced expiratory volume in one second (FEV_1), the volume of air that can be forced to be exhaled in the first second, after a complete inspiration [6]. In COPD, a ratio FEV_1/FVC, post-bronchodilator, inferior to 0.70 indicates the presence of persistent airflow limitation, and together with characteristic symptoms and significant exposures to risk factors confirms the diagnosis [2]. Based on post-bronchodilator FEV_1, the severity of airflow limitation is classified in four groups GOLD 1 to GOLD 4, corresponding to airflow limitation mild to very severe, respectively.

The disease can then be classified into four groups, A, B, C or D, according to symptom burden and risk of exacerbations (Fig. 2). Symptoms that lead to the suspicion of this disease are chronic and progressive dyspnoea, chronic cough, that can be productive or not, sputum production, wheezing and chest tightness [2].

Dyspnoea is the most common symptom in these patients and is a major cause of disability and anxiety. This symptom is commonly assessed during activities with the Modified British Medical Research Council (mMRC) questionnaire. A comprehensive assessment of symptoms can also be performed using the COPD Assessment Test (CAT™) questionnaire. Risk of exacerbations is assessed based on patient's history of moderate and severe exacerbations (including prior hospitalizations) [2]. Acute exacerbations are defined as an acute worsening of respiratory symptoms that result in additional therapy [2]. These events accelerate lung function decline, reduce physical activity and health-related quality of life and lead to poor prognosis [7]. Viral and bacterial infections are the most common causes of these episodes [7]. A patient who had two or more acute exacerbations in the previous year, or at least one hospital admission related to an acute exacerbation is identified as an individual with high-risk of future events.

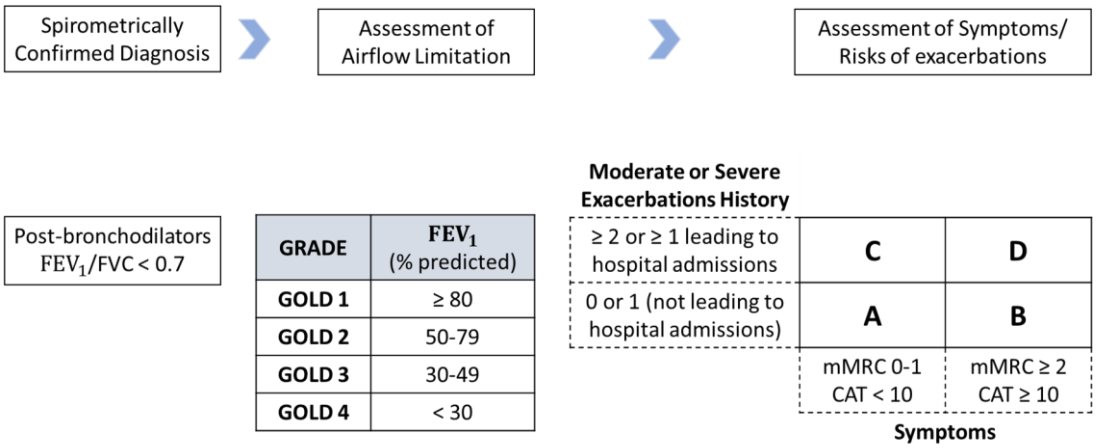

Figure 2. The refined ABCD assessment tool. Adapted from: GOLD [2]

Abbreviations: FEV1: Forced Expiratory Volume in one second; FVC: Forced Vital Capacity; mMRC: Modified British Medical Research Council; CAT: COPD Assessment Test.

1.1.3. Risk factors

COPD is caused by many factors that trigger and maintain inflammation. The principal cause is exposure to tobacco smoke. Tobacco smoking accounts for as much as 90% of COPD risk. However, less than 50% of lifelong smokers will develop COPD [2], which suggests that genetic factors or other environmental conditions may play a role in modifying the individual risk. The genetic risk factor for COPD best described is hereditary α1-antitrypsin deficiency, a glycoprotein member of the serine protease inhibitor family [8]. One of its functions is the protection of the lung against proteolytic damage of neutrophil elastase. Deficiency of this glycoprotein increases the risk of emphysema and, consequently, COPD [8].

Occupational exposures to dust, chemicals and vapours, when sufficiently intense or prolonged, can cause COPD [2]. Also, high levels of indoor air pollution caused by the use of biomass fuel and high levels of urban air pollution, called outdoor air pollution, are an important risk factor for the development of COPD [2].

Furthermore, early life environmental factors occurring during gestation, birth, childhood and adolescence can increase risk of COPD, like smoking mothers, frequent respiratory infections and asthma in childhood and bronchial hyperreactivity gases [7] . On the other hand, the socioeconomic status is inversely associated with the risk of

developing COPD [2], since it strongly influences nutrition, exposures to pollutants, infections, among others.

These risk factors play a role in many disease domains, such as on the extra-pulmonary manifestations and inflammatory process characteristic of the COPD as well as on the dysbiosis of the microbiota from the respiratory tract and lungs [9].

1.1.4. Treatment of COPD

COPD is not a curable disease, but it is treatable. Treatment can be pharmacologic and non-pharmacologic, aiming to improve symptoms, health-related quality-of-life, exercise tolerance, muscle strength, daily functioning and diminish the risk of exacerbations. Pharmacological treatment is commonly characterised by the use of inhaled bronchodilators and corticosteroids, varying accordingly to the individualised assessment of symptoms and exacerbation risk [2]. Antibiotic therapy is also a pharmacological treatment in COPD used to treat acute exacerbations of the disease or, in some cases, used as a prophylactic measure to reduce the risk of exacerbations [2]. Non-pharmacologic treatment includes lifestyles changes, like smoke cessation and avoidance of other risk factors, adopting a healthy diet, influenza vaccination, oxygen therapy, ventilatory support, physical activity and pulmonary rehabilitation (PR) [2].

1.1.4.1. Pulmonary rehabilitation and its impacts on airway microbiota

Pulmonary rehabilitation is defined by the American Thoracic Society (ATS) and the European Respiratory Society (ERS) as: “a comprehensive intervention based on a thorough patient assessment followed by patient-tailored therapies that include, but are not limited to, exercise training, education and behaviour change, designed to improve the physical and psychological condition of people with chronic respiratory disease and promote the long-term adherence to health-enhancing behaviours” [10].

PR is a GRADE A evidence for the management of COPD being the most cost-effective treatment strategy [2]. It is beneficial for patients with COPD by reducing symptoms (such as dyspnoea, anxiety, depression), exacerbations and hospital admissions, improve patients' exercise tolerance, muscle strength, knowledge and

awareness of the disease and health-related quality of life [2], [10]. This comprehensive intervention implemented by an interdisciplinary team, should be individualized to the needs of each patients, based on disease severity, complexity, and comorbidities [10]. Based on the evidence, PR should include: structured and supervised exercise training, smoking cessation, nutrition counselling, and self-management education [2], therefore, adopting a healthy lifestyle to reduce the impact of the disease.

Among the many disease-related aspects influenced by PR, the airway microbiota is a likely candidate. It is known that bacterial composition is affected by regional growth conditions, such as oxygen tension, blood flow, nutrient availability, local pH, temperature, effector inflammatory cell disposition, local microbial competition, and host epithelial cell interactions [9], [11]. Some of these factors are also affected by PR. Specifically, exercise training, one of the core components of pulmonary rehabilitation, increases the mobilised air flow or pulmonary ventilation and consequently, oxygen uptake [12]; improves cardiovascular function [10] hence increases the blood flow; and reduces dynamic hyperinflation [13]. Together, these factors contribute to improvements in gas change and consequently, have an impact in pH and temperature of the lung [10]. Exercise also has an influence on inflammation, depending on the intensity and duration this effect may be positive or negative [14], [15].

Pulmonary rehabilitation also includes smoking cessation, airway clearance techniques that aim to clear sputum from the lungs and breathing control. This may also in some way alter the airway and lung environment, conditions that influence bacterial growth.

These factors individually or jointly will most certainly influence the airway microbiota but how this process occurs during pulmonary rehabilitation is still unknown.

This dissertation will be focused on investigating microbiota modulations induced by PR.

1.2. Microbiota

1.2.1. What is the microbiota/ microbiome?

The terms microbiome and microbiota are used indiscriminately even if their scientific definition is dissimilar. Microbiota is defined as the total collection of microorganisms of a specific region, like specific host tissues or organs, or a time period [16]; whereas "microbiome", describes all genomes of microorganisms in a particular ecosystem in which they interact, such as an individual [16], [17].

These microorganisms consist not just of bacteria, but numerous species of viruses, bacteriophages and fungi [18]. In the human body, the number of microbes is of the same order as the number of human cells [19]. Microorganisms, along with their genes, proteins and metabolites, are in constant interaction with the host, which is best described as a superorganism [20].

1.2.2. Role of microbiota in chronic obstructive pulmonary disease

Until recently, the lung environment was believed to be sterile, and that it was only colonised by microbes during infection [21]. However, next-generation sequencing technology has revealed that even the healthy lung does have a microbiota [21].

Currently, microbiota is suspected to play a substantial role in healthy and diseased lungs, similarly to what happens in other organs, like the gut [11]. In health, processes such as colonization resistance, epithelial integrity, and immunoregulation are important processes performed by the pulmonary microbiota [11], [22]. Disruption of the composition or overall numbers of "normal" microbiota, known as dysbiosis, specifically in the lung may contribute to the development and progression of lung disorders, like COPD [11], [22]. Airway microbiota of patients with COPD have been reported to be distinguish from the healthy individuals [5]. The lung healthy microbiota seems to be dominated by phyla Firmicutes and Bacteroidetes, the prominent genera include *Streptococcus*, *Veillonella, Prevotella, Fusobacterium* and *Haemophilus* genera [23]–[26]. In COPD, the most abundant phyla in the lung are Firmicutes, Proteobacteria and

Bacteroidetes and *Veillonella*, *Haemophilus*, *Streptococcus*, *Prevotella* and *Moraxella* representing the five most abundant genera[27], [28].

In COPD, dysbiosis of the lung microbiota composition is associated with inflammatory responses, pathological changes of local immune responses [29] and increased risk of exacerbations. Further microbial dysbiosis has been identified as the cause of about half of the total reported acute exacerbations of COPD [30]. The "Vicious Circle Hypothesis" (Fig. 3) is a theory supported by these observations and explains how these factors interact for the progression of the disease [31], [32]. It proposes that insults, like smoking, can disturb innate defence mechanisms such as mucociliary clearance, and this allow specific pathogenic bacteria that normally enter and exit the lower respiratory tract, to persist and proliferate, altering the bacterial composition [31], [32]. This unbalanced microbiota can induce the inflammatory process in COPD, which by impairing the protease-antiprotease balance in lung causes progressive airflow obstruction and possibly airway epithelial damage [31], [32]**.** On the other hand, abrupt changes in the lung microbiota, for example by introduction of a new strain of a respiratory pathogen, can drive an acute exacerbation of COPD, that result in increases in airway and systemic inflammation, and consequently leads to worsening of the respiratory symptoms [31], [32]. This induces further damage to innate lung defence mechanisms, allowing further bacterial colonization and thus establishing the vicious cycle (Fig. 3)[31], [32].

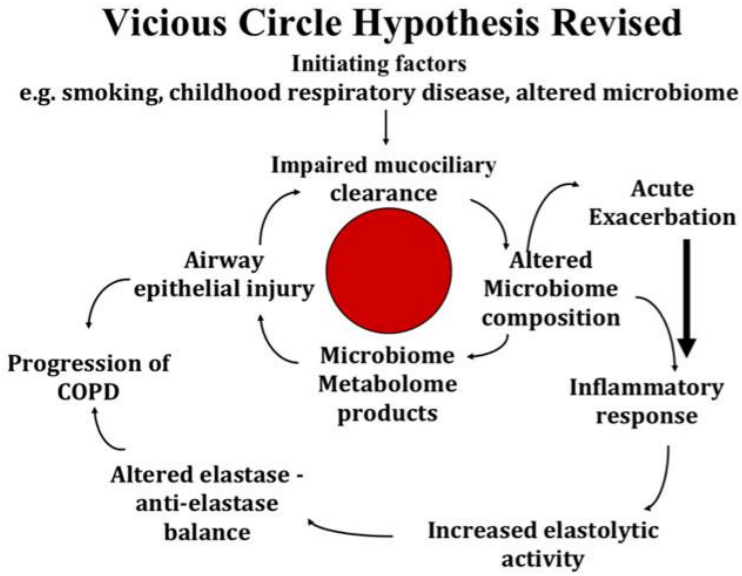

Figure 3. Vicious circle hypothesis revised. The initial factors impair innate defence; microbial colonization perpetuates a cyclical sequence of events that contributes to the persistent inflammation and microbiome changes that are characteristic of COPD. From: Mammen and Sethi, 2016. [32]

The bacterial colonization is also associated with increased cough and sputum and accelerated loss of lung function [30]. The "British Hypothesis" supports this association. It suggests that recurrent bronchial infections are the reason why only a few smokers developed progressive airway obstruction and the others did not and propose that persistent or recurrent bacterial infections in the airways are in part responsible for chronic cough and mucus hypersecretion [33] which accelerates lung function decline [34], [35].

1.2.3. Representativity of pulmonary microbiota in salivary microbiota

Migration of microbiota to the lower respiratory tract is promoted by microaspiration of secretions from the upper airways, a normal process that occurs especially during sleep, by inhalation of bacteria and direct mucosal dispersion [9], [11], [32]. Microbial migration is counterbalanced by mechanisms of the respiratory system designed for microbial elimination: cough, mucociliary clearance and innate and adaptive host defences [11]. In COPD, this balance is impaired because patients have a reduced laryngotracheal mechanosensitivity [36] and a poor coordination of breathing and swallowing [37]. Therefore, they are more prone to microaspiration, which is aggravated

by impaired mucociliary function present in COPD. This results in a decrease airway clearance, i.e., inability to eliminate the aspirate[38], which creates favourable conditions for certain bacteria to grow [11]. Conversely, since healthy lungs are inhospitable for bacterial growth, the bacterial composition is mainly determined by the dynamic balance between its immigration and elimination [11], [39], [40].

Several studies show an overlap between the oral and pulmonary microbiota both in healthy and in patients with COPD. Analyses of bronchoalveolar lavage (BAL) fluid, oral wash, oropharyngeal and nasopharyngeal swabs have showed that lung microbiota is indistinguishable in community composition from upper airway microbiota [25], [32], [41], [42], showing that microbial immigration from the oral cavity appears to be the significant source of the lung microbiota. However, the proportion of oral bacteria in the pulmonary microbiotas is higher in COPD than in healthy pulmonary microbiota. Although the bacterial concentration in the lungs is lower than in the mouth, the upper airway microbiota may be predictive of the lung tissue microbiota. The most common genera in both BAL and oral wash, that is in lungs and mouth, were *Prevotella*, *Streptococcus*, *Fusobacterium*, *Neisseria* and *Veillonella* [25], [41], [43].

These studies support that the oral microbiota is representative of the pulmonary microbiota. In fact, this is not unexpected, since migration from the upper airways is the most likely source of pulmonary bacteria; the mouth is particularly rich in microbes and it is contiguous to lungs [26].

Furthermore, saliva, as a biological sample, has a great advantage since its collection is a simple, relatively quick and easy procedure, does not cause discomfort and is non-invasive and painless to patients. This is particularly important when studying a population living with a chronic health condition.

2. Methods

2.1. Ethics

Ethical approval was previously obtained from Centro Hospitalar do Baixo Vouga (which includes Estarreja Hospital) and Agrupamento dos Centros de Saúde do Baixo Vouga, which includes all primary healthcare centres (PHCC) in Estarreja. Approval from the National Data Protection Committee was also obtained (Annex I).

2.2. Design, participants and data collection

This dissertation consisted in a longitudinal study. Patients with COPD were first identified by a pulmonologist, who provided a brief explanation about the purposes of the study. Only interested participants were approached by the research team. Patients were eligible if they were diagnosed with COPD according to GOLD criteria (1); were in a stable phase of the disease, i.e., no acute exacerbations in the previous month and were able to give informed consent. Patients were excluded if they presented: history of an acute cardiac/respiratory condition within the previous month; significant cardiac, musculoskeletal or neuromuscular diseases that impaired their participation in the PR programme; signs of cognitive impairment and history of neoplastic or immunological disease. Written inform consent was obtained from all participants.

All participants were offered 12-weeks of community-based pulmonary rehabilitation, (2 weekly sessions), with exercise training twice a week and psychoeducational sessions once every other week. Each session lasted approximately 60 minutes. Patients were followed for 9 months (including ~3 months before PR, 3 months during PR and 3 months after PR). The intervention was delivered by a multidisciplinary team of healthcare professionals. These participants composed the “intervention group”, since it was possible to assess patients, not enrolled in pulmonary rehabilitation, and followed them for 6 months, thus composing the “control group” of the study. During these periods, sociodemographic, anthropometric, clinical data and saliva samples (once a month) were collected.

Sociodemographic (i.e., age and gender), anthropometric (i.e., BMI) and general clinical data (i.e., smoking status, medication, comorbidities, number of acute exacerbations in the previous year, number of hospital admissions due to acute exacerbations) were first collected with a structured questionnaire based on the International Classification of Functioning, Disability and Health (ICF) checklist [44]. Further, clinical data were then collected. Lung function was assessed with spirometry according to the international recommendations [45] and forced expiratory volume in 1 second percentage predicted (FEV_{1pp}) was used to determine the GOLD grade. Peripheral capillary oxygen saturation (SpO_2) was measured with a pulse oximeter. Impact of the disease was assessed with the COPD assessment test (CAT) [46] and activity limitations due to dyspnoea with the modified Medical Research Council questionnaire [47]. These two questionnaires together with the number of acute exacerbations in the previous year and number of hospital admissions due to acute exacerbations were used to determine the GOLD ABCD. Clinical data were collected three months before the PR programme, immediately before and after the intervention, (i.e., the intervention lasted 3 months) and at three months after the intervention. During the clinical assessment, patients were asked to collect their saliva to a cup. Saliva samples were used to assess airway microbiota profiles at baseline, during and after intervention, and understand the short-mid-term impact of PR on lung microbiota. All saliva samples were transferred to microtubes and then frozen at -80ºC (Fig. 4).

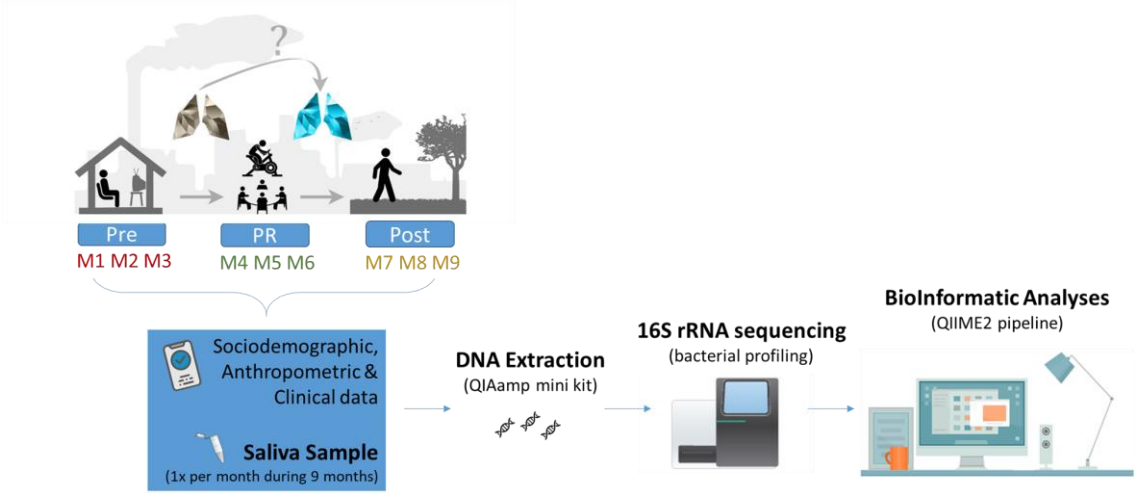

Figure 4. Flowgram of the methodology followed in data collection. PR-pulmonary rehabilitation.

2.3. Data analysis

Descriptive statistics were used to characterise the sample. For the purpose of this dissertation we will focus on the analysis saliva's microbiota.

2.3.1. Samples analysis

2.3.1.1. DNA extraction

Prior to manipulation, samples were unfrozen at room temperature. DNA from saliva samples was extracted with the QIAamp DNA mini kit (Qiagen, Hilden, Germany) following the manufacturer's instructions, with slight modifications: initial sample volume was doubled, buffers and Qiagen protease volumes were adjusted and elution volumes were reduced to a quarter of which is recommended (full protocol is available in appendix 1). Three negative controls of DNA extraction procedure were performed, where saliva was replaced by 1x phosphate-buffered saline.

2.3.1.2. DNA quantification

Quantity and quality of extracted DNA were assessed by absorbance method in Denovix DS-11 FX spectrophotometer, with OD260/280 and OD260/230 ratios. 1µL of milli Q-water, elution buffer, was used as a blank. OD260/280 and OD260/230 are indicators of sample quality. Pure DNA usually present an OD260/280 of 1.8 and an OD260/230 of 2.0-2.2 [48]. Altered values of OD260/280 and OD260/230 may indicate contamination by residual phenol, guanidine, or other reagents used in extraction procedure, or that the sample has very low DNA concentrations [49] .

2.3.1.3. 16s rRNA sequencing

16S rRNA gene amplification and sequencing was carried out at the Gene Expression Unit from Instituto Gulbenkian de Ciência, following the service protocol. Briefly, for each sample, the hypervariable V4 region of 16S rRNA gene was amplified, using universal pair of primers F515 (5'-CACGGTCGKCGGCGCCATT-3') / R806 (5'-GGACTACHVGGGTWTCTAAT-3'). Samples were then pair-end-sequenced on an Illumina MiSeq Benchtop Sequencer, following Illumina recommendations.

2.3.1.4. Microbiota and Statistical Analysis

QIIME2 2019.4 [50] was used to perform microbiota bioinformatic analyses, since this pipeline allows reliable and reproducible results due to the availability of a common framework for microbiome data analysis. First, demultiplexed 16S paired-end sequences were imported, quality filtered, with a Phred quality threshold of 33, and mapped to their native samples. Subsequently, the forward and reverse reads were joined (q2-vsearch), submitted to Q-score base filtering [51], chimera removing and 16S-denoising with Deblur [52] to retain the true biological sequences present in the total data (reads without minimum of 60% identity similarity to sequences from 85%-OTU-GreenGenes-13_8 database were discarded). Results from previous steps were next summarized in a feature table. Taxonomy assignment of amplicon sequence variants (ASVs) was performed with q2-feature-classifier plugin [53], through classify-sklearn method with pre-trained Naïve Bayes classifier against 99%-OTU-GreenGenes-13_8 reference sequence[54] (sequences trimmed to only include 250bp of V4 region, bound by the F515/R806 primer pair). Mitochondria and chloroplasts were filtered out from resulting data with q2-taxa plugin through filter-table method.

q2-phylogeny plugin was next employed to produce a MAFFT alignment [55] of ASVs which was consequently used to construct a rooted phylogeny with FastTree2 [56] for subsequent applications. Alpha diversity was calculated through Shannon's diversity index, observed OTUs, Pielou's Eveness and Faith's Phylogenetic Diversity index metrics, as implemented in QIIME2. Beta diversity metrics like Unweighted-Unifrac, Weighted-Unifrac, Jaccard distance and Bray-Curtis dissimilarity were also calculated. Spatial dissimilarities between bacterial communities of different groups were assessed with Principle Coordinate Analysis (PCoA), plotted with the different beta diversity metrics' distance matrixes. PCoA plots is a visualisation in 3-dimension of distance matrix that results from the calculation of beta diversity indices for all combinations of pairs of samples [57], in other words, it is a representation in 3-dimension of the samples in terms of dissimilarity. Therefore, the closer two points (two samples) are, the more similar they are. Alpha diversity metrics, Beta diversity metrics and PCoA were estimated using q2-

diversity plugin after rarefaction of samples (subsample without replacement) to 5314 sequences per sample.

Friedman test with Dunn's correction was used to compare differences in bacterial ratios and differences in alpha diversity metrics between periods and different months, to compare differences of relative abundances of phyla and genera, differences of alfa diversity and differences of beta diversity in cases of one month interval. Wilcoxon was used to compare differences of relative abundances of phyla and genera, differences of alfa diversity and differences of beta diversity in cases of two months difference. Permutational multivariate analysis of variance (PERMANOVA) [58] was used with the different beta diversity metrics' distance matrixes to quantify differences in the community composition.

Longitudinal data analysis was performed with q2-Longitudinal plugin, that incorporates multiple methods for analysis of longitudinal and paired-sample data. Of these, Linear Mixed Effects (LME) models and some data transformations, such as, first distances and difference/distance to baseline were performed [59]. LME models examine the relationship between one or more independent variables (fixed effects) and a single response variable, where observations are made across dependent samples. Time and patient exposure (if participant was intervention or control) were used as fixed effects. First distance identifies the beta diversity distance between successive samples from the same subject. It's important because allows us to analyse longitudinal changes in beta diversity using linear-mixed-effects model that cannot operate directly on a distance matrix. The first differences and first distances to baseline calculates differences/distances from a static point (e.g., baseline or a time point when a treatment is administered): $\Delta Y_t = Y_t - Y_{baseline}$. Calculating baseline differences/distances can help tease apart noisy longitudinal data to reveal underlying trends in individual subjects or highlight significant experimental factors related to changes in diversity or other dependent variables [59].

Analysis of composition of microbiomes (ANCOM) [60] and Linear discriminant effect size (LefSe) analysis [61] were performed to identify differentially abundant OTUs between groups of samples formed according to periods, Pre-PR, PR and Post. LefSe is an

algorithm for high-dimensional biomarker discovery that uses linear discriminant analysis to estimate the effect size of each taxon that is differentially represented in each group of samples, not accounting for the compositional nature of the microbiota [60]–[62]. ANCOM uses a log ratio analysis to make point estimates of the variance and mean of taxa taking into consideration the data compositional nature [60]. These analyses were conducted with a filtered feature table produced with q2-feature-table plugin. Low abundance OTUs (<10 sequences summed across all samples) and OTUs that show up in only one or a few samples (<5) were filtered out from the feature table. The feature table was then collapsed at different taxonomic levels, 2 (phylum) to 6 (genus), for the analyses. ANCOM was performed in R with ANCOM 2.0 script (available at the developer webpage [63]) with taxa-wise multiple correction and a W cut-off of significance of 0.7. (both recommended by the developer, based on simulation data). LefSe was performed in the online version [64] with default parameters.

Alfa diversity analysis

In this specific context, alpha diversity is defined as the diversity of taxa of an individual's microbiota, and is associated with the richness and evenness of taxa within the airway environment [65]. Richness is a count of the number of different taxa observed in the community without take into account their frequencies [57]. Evenness refers to the equitability of the taxa frequencies in a community [57], in other words, how close in numbers each taxa in a community are. Alfa-diversity measures differ in the weight they give to each of these components. Commonly used alpha diversity measures include Observed OTUs, Shannon's diversity index, Evenness and Faith's Phylogenetic Diversity. Observed OTUs [66] is equal to richness and therefore provides no weight to the evenness component, Shannon's diversity index [67] equally weights richness and evenness, Evenness or Pielou's Evenness [68], as the name implies, is a measure of community evenness and the Faith's Phylogenetic Diversity [69], a phylogenetic measure of alfa diversity, is a qualitative measure of community richness that incorporates phylogenetic difference between features. These differences in weighting and units explain differences often observed in results from each measure.

Beta diversity analysis

Beta-diversity is the difference in diversities across samples or communities [65]. A beta diversity index is calculated for each pair of samples and represents either a similarity or a distance between the two samples [57] . The calculation of beta diversity indexes for all combinations of sample pairs results in a distance matrix. The distance matrix then can be visualized with ordination methods such as Principal Coordinate Analysis (PCoA) [65]. There are many phylogenetic and non-phylogenetic beta diversity metrics, such as weighted Unifrac and unweighted Unifrac, as an example of phylogenetic metrics, and Jaccard distance and Bray Curtis dissimilarity, as an example of non-phylogenetic metrics. Unifrac measures the amount of unique evolution within each community with respect to another by calculating the fraction of branch length of the phylogenetic tree that is unique to either one of a pair of communities [65] . There are several variants of Unifrac, including weighted and unweighted Unifrac. The weighted Unifrac metric [70] is weighted by the relative abundance of OTUs and it incorporates these abundances when calculating shared/unshared branch lengths to calculate distance, whereas unweighted Unifrac [71] only consider the absence/presence of the OTUs [65]. So the impact of low-abundance features in weighted Unifrac is diminished and so it is useful for examining differences in community structure [71]. Bray Curtis dissimilarity [72] is based on abundance of OTUs. It assigns the differences in microbial abundance between two samples varying in values of 0 to 1. 0 means both samples share the same OTUs at exact the same abundances and 1 means both samples have complete different OTUs and/or abundances. Jaccard distance [73] is based on presence or absence of OTUs. It attaches the differences in microbial composition between two samples also varying in values of 0 to 1. 0 means both samples share exact the same OTUs and 1 means both samples have no OTUs in common.

3. Results and Discussion

3.1. Participants' characteristics

Thirty patients with COPD were enrolled in this study, 25 in the intervention group (72.5±6.5 years old; FEV1pp= 48.2 ±14.7; BMI= 26.3±4.6 kg/m2; 19 (86.4%) male), and 5 in the control group (75.4±5.9 years old; FEV1pp= 48.2±13.1; BMI= 25.2±2.3 kg/m2; 5 (100%) male). Detail cohort characterisation is summarised in table 1.

Table 1. Sociodemographic, anthropometric and clinical characteristics of study participants at baseline

Characteristics	**Intervention (n=25)**	**Control (n=5)**
Age (years) at enrolment, mean±SD	72.5±6.5	75.4±5.9
Male sex, n (%)	19 (86.4%)	5 (100%)
Body mass index (BMI), mean±SD	26.3±4.6	25.2±2.3
GOLD Grade, n (%)		
I	1 (4%)	0
II	10 (40%)	2 (40%)
III	1 (4%)	3 (60%)
IV	6 (24%)	0
GOLD ABCD assessment, n (%)		
Group A	8 (32%)	1 (20%)
Group B	10 (40%)	0
Group C	13 (52%)	2 (40%)
Group D	6 (24%)	2 (40%)
SpO_2, mean±SD	94.7±1.8	95.4±2.2
FEV_1pp, mean±SD	48.2±14.7	48.2±13.1
Number of exacerbations in the year before enrolment, n (%)		
0-1	18 (72%)	3 (60%)
≥2 or 1 with hospital admission	28 (7%)	2 (40%)
Hospital admissions, due to COPD, in the year before enrolment, n (%)		
0	23 (92%)	5 (100%)
1	2 (8%)	0
CAT Scores, n (%)		
Higher Impact (≥10)	19 (76%)	2 (40%)
Lower Impact (<10)	6 (24%)	3 (60%)

n (%): number of individuals in each group plus the corresponding percentage. mean±SD: mean±standard deviation. BMI: Body Mass Index; GOLD Grade: I – Mild; II – Moderate; III– Severe; IV– Very Severe; GOLD ABCD assessment: Group A: 0-1 exacerbations not leading to hospital admission and CAT scores <10; Group B: 0-1 exacerbations not leading to hospital admission and CAT scores ≥10; Group C: ≥2 or 1 exacerbation with hospital admission and CAT scores <10; Group C: ≥2 or 1 exacerbation with hospital admission and CAT scores ≥10; SpO_2: peripheral capillary oxygen saturation, FEV_1pp: forced expiratory volume in 1 second percent predicted; CAT: COPD assessment Test.

3.2. Does pulmonary rehabilitation lead to changes in the salivary microbiota?

In the following sections we explored the possibility of PR influencing microbiota analysing the data in three different ways: i) in each period (Pre-PR, PR and post-PR) and month ; ii) short term variation (over one or two months) and iii) in a temporal series.

3.2.1. Comparison of microbiota composition, alfa and beta diversity, determined in discrete time points pre, during and post pulmonary rehabilitation

3.2.1.1. Microbiota Composition

We analysed the microbiota composition in terms of the taxonomic levels, phyla and genera. We compared the means of the top 5 most represented phyla and the means of the most represented genera of each phyla by period: pre-PR, PR and post-PR. A similar analysis was done by month, considering M3 (the last month pre-PR) and M4, M5 and M6 (the 3 months of PR). A significant increase from pre-PR to PR was observed in the phylum Proteobacteria (Friedman test P=0.0263, Dunn's multiple comparisons, *P*=0.0487)(Fig. 5A and 6) and in the genus *Neisseria* (Friedman test P=0.0003, Dunn's multiple comparisons, *P*=0.0002)(Fig. 5A and 6). No other significant differences were found (Fig. S1 and S2).

To detect significant differentially abundant bacterial groups, at taxonomic levels phyla and genera, between periods (pre-PR to PR, PR to post-PR and pre-PR to post-PR) we used two methods: LEfSe, and ANCOM. With LEfSe no significant differences were found at the level of phyla between periods. At genus level, comparing pre-PR with PR period, LEfSe pointed *Pseudomonas* and *Sphingomonas* as being enriched in pre-PR and *Neisseria and Alloscardovia* as being enriched during PR (Fig. 5 B). When comparing PR with post-PR period, LEfSe pointed *Granulicatella* and *Acinetobacter* as being enriched in PR period and *Staphylococcus*, *Selenomonas* and Pasteurellaceae family as being enriched in post-PR period. Comparing pre-PR with post-PR period, LEfSe pointed *Granulicatella*, *Sphingomonas*, *Pseudomonas* and *Enhydrobacter* as being enriched in pre-PR period and

Pasteurellaceae family as being enriched in post-PR period. ANCOM did not reveal any abundant genera between periods.

In sum, the composition of microbiota changes upon PR, with an increase of Proteobacteria phylum, particularly, *Neisseria* genus. Pre-PR samples comparing to PR samples were enriched in genera *Pseudomonas* and *Shingomonas* both Proteobacteria, *Alloscardovia* genus, from phylum Actinobacteria was also augmented in PR samples. Between PR and post-PR period, LEfSe showed that PR samples were enriched in *Granulicatella* and *Acinetobacter*, from phyla Firmicutes and Proteobacteria, respectively, and post-PR samples were enriched in *Staphylococcus* and *Selenomonas* both from phylum Firmicutes and family Pasteurellaceae from Proteobacteria phylum. In comparison between pre-PR with post-PR period, LEfSe pointed *Granulicatella*, from phylum Firmicutes, and *Sphingomonas*, *Pseudomonas* and *Enhydrobacter*, from Proteobacteria phylum, as being enriched in pre-PR period and family Pasteurellaceae as being enriched in post-PR period. The compositional approach performed with ANCOM did not show differential abundant OTUs between periods.

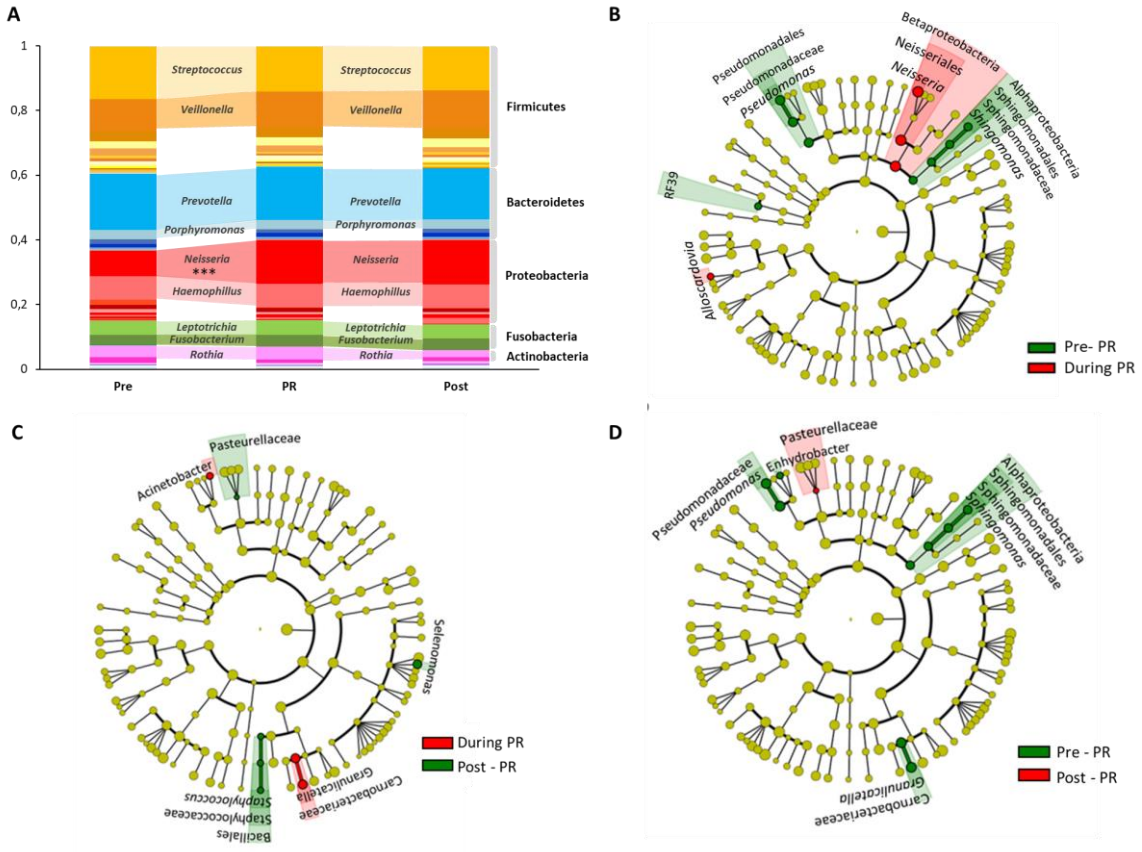

Figure 5. Composition of the microbiota during the three periods considered. **A)** Mean frequency of phyla and genera of bacteria present in saliva before (Pre), during (PR) and post-PR (Post). **B)** Cladogram showing significant differences (Linear Discriminant Analysis, effect size analysis) in microbiota composition between the different periods. Pre-PR is enriched in the genera *Pseudomonas* and *Sphingomonas* and PR is enriched in *Neisseria* and *Alloscardovia*. **C)** Cladogram showing significant differences (Linear Discriminant Analysis, effect size analysis) in microbiota composition between the different periods. PR is enriched in the genera pointed *Granulicatella* and *Acinetobacter* and post-PR is enriched in *Staphylococcus*, *Selenomonas* and family Pasteurellaceae. **D)** Cladogram showing significant differences (Linear Discriminant Analysis, effect size analysis) in microbiota composition between the different periods. Pre-PR is enriched in the genera *Granulicatella*, *Sphingomonas*, *Pseudomonas* and *Enhydrobacter* and post-PR is enriched in family Pasteurellaceae. **P*<0.05, ***P*<0.01, ****P<0.001. PR-pulmonary rehabilitation.*

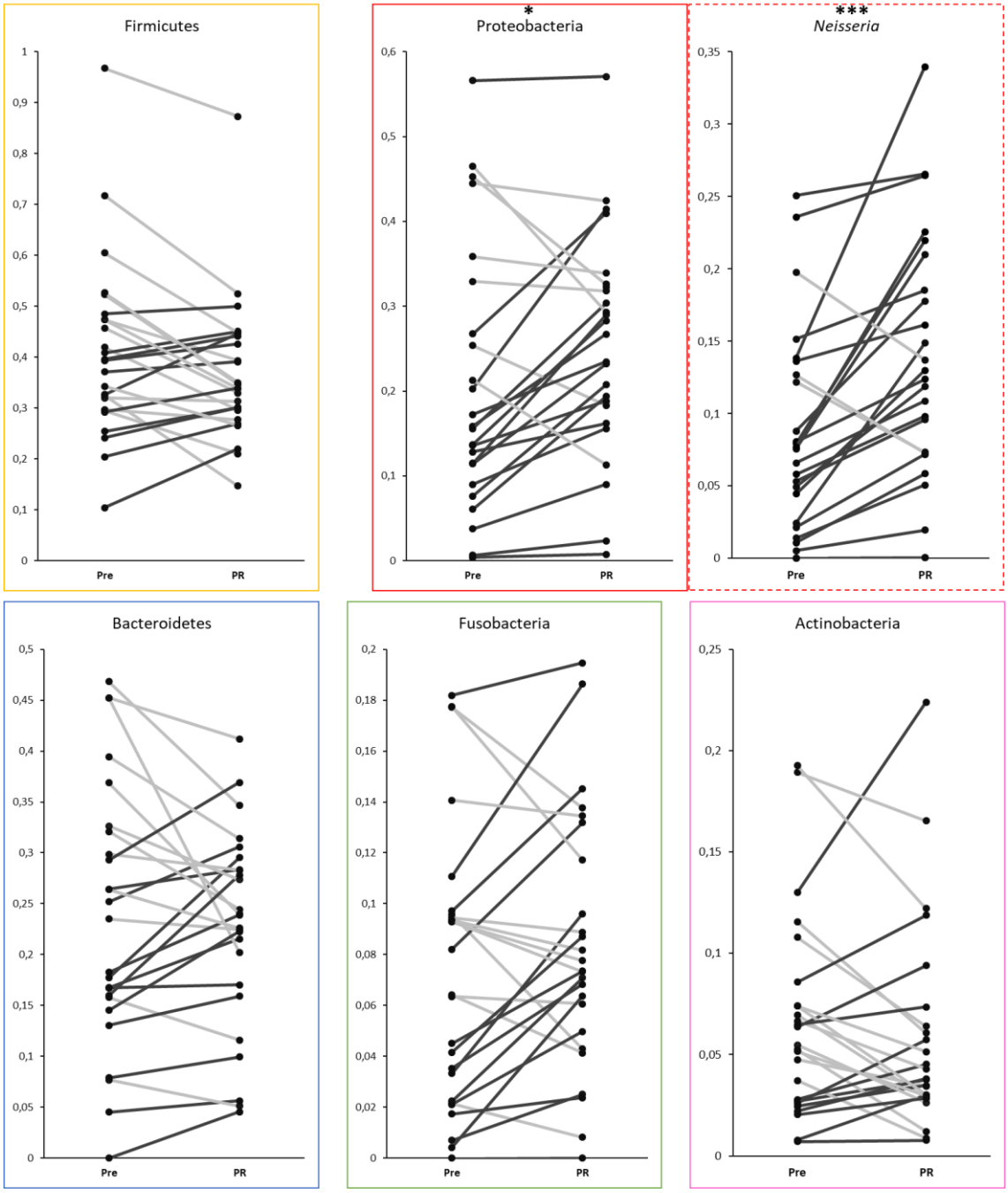

Figure 6. The composition of the microbiota changes upon PR. Changes in frequencies *of Firmicutes, Bacteroidetes, Proteobacteria, Fusobacteria and Actinobacteria* phyla and in genus *Neisseria* from Pre-PR to PR. Lines in black represent an increase from Pre-PR to PR and in grey a decrease. Only Proteobacteria phylum (Wilcoxon test *P*= 0.0255) and *Neisseria* genus (Wilcoxon test *P*= 0.0012) showed a significant change upon PR. **P*<0.05, ***P*<0.01, *****P<0.001*

Since PR is known to have a beneficial effect in patients with COPD, in terms of improvement of several clinical parameters, we expected that PR was going to modulate

microbiota composition to ensembles analogous to less severe state ensembles, with enriched Bacteroidetes [74]. Our results refuted this hypothesis, since we observed an increase in Proteobacteria, specifically of *Neisseria* genus. Increases in Proteobacteria relative abundance have been linked to more advanced stages of COPD [75]–[77] and several inflammatory diseases like inflammatory bowel disease (IBD) [78], [79], due to chronic inflammation. In this specific case, the chronic inflammatory processes affect primarily the small airways and alveoli [80]. Proteobacteria are well adapted to resist and even grow in stress situations, due to the aerobic respiration [81]. Several studies have shown that moderate to vigorous exercise training can increase inflammation and oxidative stress, particularly in people not used to exercise [15], [82]–[84]. This response is due to muscle and connective tissue damage caused by exercise training [82]–[84]. It has also been shown that the beneficial effects of PR appear to be independent of inflammation [85]. So, it is expected that Proteobacteria relative abundance might increase in comparison to steady state. However, the most commonly mentioned genera of Proteobacteria phylum are *Haemophilus* and *Moraxella*, not *Neisseria* [77]. This might be explained by the use of saliva samples instead of sputum or BAL samples like in other studies [25], [77].

Pseudomonas (found to be enriched in pre-PR samples), was also associated with decreased lung function and seems to have a role in the development, progression and exacerbation of COPD [75], and in our study it is shown that this genus decreases with pulmonary rehabilitation.

The genus *Granulicattella* has been reported to be increase in healthy people comparing to patients with COPD [86]. We found decrease of this genus after PR compared to other periods which corroborates that patient's microbiota does not approximate to the healthy state after PR. Additionally, the post-PR samples were enriched in family Pasteurellaceae and in *Staphylococcus* and *Selenomonas* genera compared to during-PR. Members of Pasteurellaceae family and Sthaphylococcus genus were found to be increased and associated with exacerbations in COPD [87], [88]. This seems to show that after PR patients' microbiota is enriched in in bacteria that are

associated with exacerbations, however, the sample we are using is unusual and this association may not be found in saliva.

3.2.1.2. Alpha diversity

To see if PR alters microbiota diversity, we calculated and compared the mean of alpha diversity for each period. No significant differences were observed (Friedman test with Dunn's multiple comparisons) for the four alpha diversity measures analysed (Fig. 7A and appendix III). Since differences between months might have been masked when considering means of alpha diversity per period, we also calculated differences in alpha diversity per month. So, we compared M3 (pre-PR) with M4, M5 and M6 (three months of PR). No significant differences were found (Friedman test Dunn's with multiple comparisons) (Fig. 7B and appendix III).

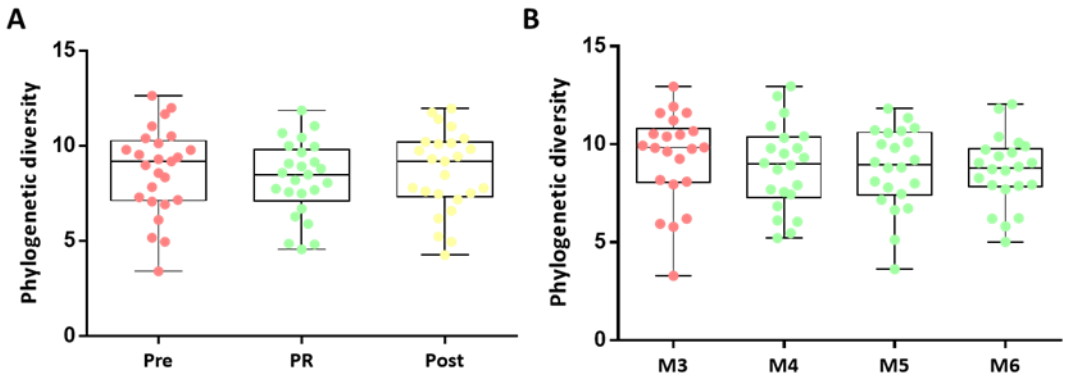

Figure 7. Mean of alpha diversity estimated with Faith's phylogenetic diversity index measured in A) all time points were pooled by period and B) months, M3 (before PR), M4, M5 and M6 (during PR). No significant differences were observed when comparing the three periods (Friedman test *P*=0.1409, with Dunn's multiple comparisons. No significant differences were found when comparing M3 (pre-PR) with M4, M5 and M6 (three months of PR) (Friedman test *P*=0.3916, with Dunn's with multiple comparisons).

In conclusion, PR does not seem to alter alpha diversity, but this does not mean that there have been no changes in microbiota composition.

Decreased alpha diversity is associated with disease worsening [74], [89], [90] and is often accompanied by an overrepresentation of some genera, frequently pathogenic genera [76]. In the gut, physical exercise increases alpha diversity [91]–[95], so we

hypothesized that alpha diversity could increase with PR, what was not observed. But it has also been shown that age decreases diversity [87], [96], [97] so PR may be attenuating this decrease and so we see no change.

3.2.2. Short term fluctuations of the microbiota

3.2.2.1. Microbiota Composition

Another way of querying the effect of PR is comparing short term fluctuations of the microbiota in steady state with those observed upon PR.

So, we calculate changes in relative abundance of phyla and most abundant genera of each phylum, between successive timepoints (Δ_1, one month interval: M2-M1, M3-M2, M4-M3) and between timepoints with two months difference (Δ_2, two month interval: M3-M1, M5-M3) (Fig. 8-10 and appendix IV). M2-M1, M3-M2 and M3-M1 correspond to the time intervals in steady state. M4-M3 and M5-M3 correspond to the transition between steady state and PR. Due to missing samples only 7 patients were included for analysis of Δ_1, and steady state of Δ_2.

Upon one month of PR, Firmicutes and Proteobacteria changed with higher magnitude than in steady state (Friedman test P=0.0515, Dunn's multiple comparisons, *P*=0.0486 and Friedman test P=0.0272, Dunn's multiple comparisons, *P*=0.0227, respectively) (Fig. 8A and 9A). In terms of genera, significant differences were found for *Streptococcus* relative frequency (Friedman test P=0.0084, Dunn's multiple comparisons, *P*=0.0066) (Fig. 10A). No other significant differences were observed (Fig. 8-10B and appendix IV).

These results suggest that the rate of change in the first month (M2-M1), pre-PR, is significantly different from the rate of change in the third month (M4-M3), transition period, particularly in phyla Firmicutes and Proteobacteria and genus *Streptococcus*. Moreover, in phylum Firmicutes and genus *Streptococcus*, we observed an increase of their relative abundances in the first month (M2-M1), in month 1 there was a lower relative abundance than in month 2 (positive difference) and a decrease in the third month (M4-M3), in month 3 there was a greater relative abundance than in month 4 (negative difference). In phylum Proteobacteria, the opposite trend was observed, we

found a decrease in the first month (M2-M1) and an increase in third month (M4-M3) (consistent with the results of the section 3.2.1.1).

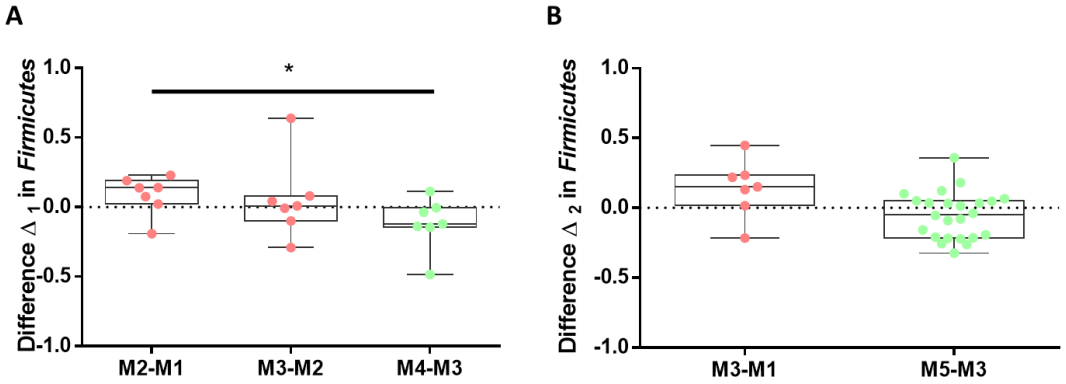

Figure 8. Differences in relative frequency of Firmicutes A) between successive timepoints, M2-M1 and M3-M2 (steady state), and M4-M3 (transition between steady state and PR) and B) between timepoints with two months difference M3-M1 (steady state) and M5-M3 (transition between steady state and PR). A significant difference was observed between M2-M1 and M4-M3 (Friedman test P=0.0515, Dunn's multiple comparison P=0.0486). No differences were observed when we considered two-month interval, Δ2, (Wilcoxon test). *P<0.05, **P<0.01, ***P<0.001

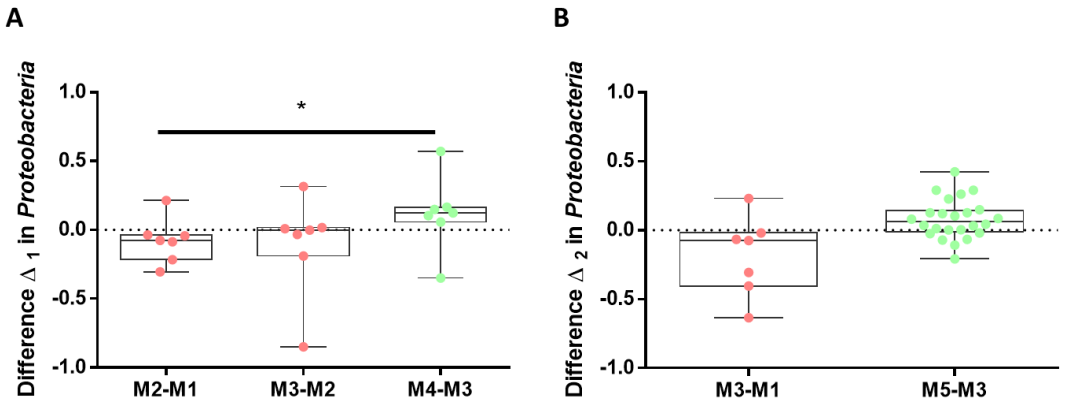

Figure 9. Differences in relative frequencies of Proteobacteria phylum A) between successive timepoints, M2-M1 and M3-M2 (steady state), and M4-M3 (transition between steady state and PR) and B) between timepoints with two months difference M3-M1 (steady state) and M5-M3 (transition between steady state and PR). A significant difference was observed between M2-M1 and M4-M3 (Friedman test *P*=0.0272, Dunn's multiple comparison *P*=0.0227). No differences were observed when we considered two-month interval, Δ_2, (Wilcoxon test). **P*<0.05, ***P*<0.01, ****P<0.001*

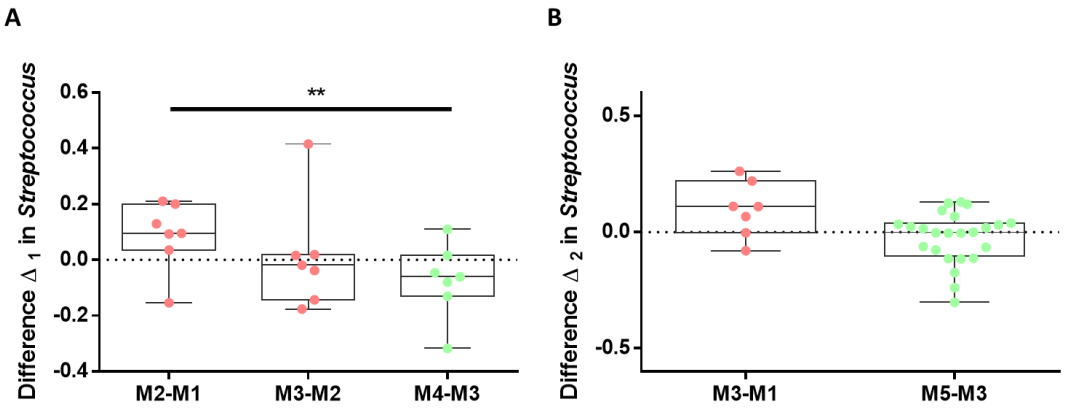

Figure 10. Differences in relative frequencies of *Streptococcus* genus A) between successive timepoints, M2-M1 and M3-M2 (steady state), and M4-M3 (transition between steady state and PR) and B) between timepoints with two months difference M3-M1 (steady state) and M5-M3 (transition between steady state and PR). A significant difference was observed between M2-M1 and M4-M3 (Friedman test P=0.0084, Dunn's multiple comparison P=0.0099). No differences were observed when we considered two-month interval, Δ_2, (Wilcoxon test). *P<0.05, **P<0.01, ***P<0.001*

Streptococcus is frequently identified as one of the most common genus implied in bacterial exacerbations of COPD [98]. Since, amongst others, PR reduces the frequency of exacerbations [2], we were expecting a decrease in the causative bacteria. Also, the increase of Firmicutes is observed in more severe states of the disease [99], [100] and, therefore, we also expected a decrease of Firmicutes relative abundance during PR in this phylum. These results should be interpreted carefully, due to M3-M2 (the other difference in steady state) is not significantly different from M4-M3, therefore, we cannot consider that the transition between steady state and PR is different from basal fluctuations and also because our small sample size and therefore, more research with larger samples are recommended.

3.2.2.2. Alpha diversity

Similarly to the analysis of phyla/genera relative abundances, we compared differences of alpha diversity of M2-M1, M3-M2 (steady state) and M4-M3 (transition between steady state and PR) in cases of Δ_1, and M3-M1 (steady state variation) with M5-M3 (transition between steady state and PR) in cases of Δ_2 (Fig. 11 and appendix V).

Due to missing samples only 7 patients were included for analysis of Δ_1, and steady state of Δ_2.

For phylogenetic diversity, no significant differences were found in Δ_1 comparisons of steady state with transition to PR. However, the two Δ_1 of steady state were significantly different between each other (Fig.11 A), which give us an idea of the steady state monthly variation in Phylogenetic diversity. No differences were observed when considering a time interval of two months (Fig. 11B). Similar results were observed with Shannon's diversity index, Evenness and Observed OTUs (appendix V). Interpretation of these results is challenging, since, there is no biological support for an increasing alpha diversity during steady state. We believe that the small sample size (seven patients) included is on the basis of these findings and increasing the sample size will be important to clarify these observations.

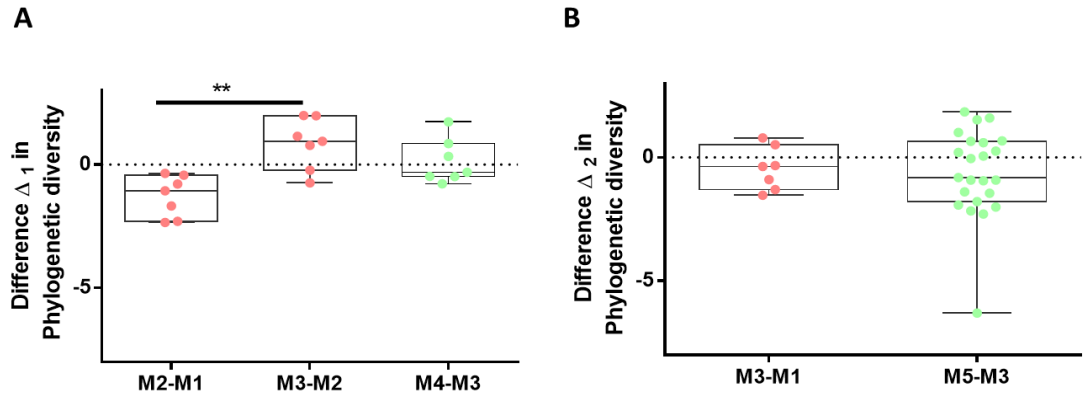

Figure 11. Differences in alfa diversity (Faith's phylogenetic diversity index) A) between successive timepoints, M2-M1 and M3-M2 (steady state), and M4-M3 (transition between steady state and PR) and B) between timepoints with two months difference M3-M1 (steady state) and M5-M3 (transition between steady state and PR). A significant difference was observed between M2-M1 and M3-M2 (Friedman test *P*=0.0036, Dunn's multiple comparison *P*=0.0040). No differences were observed when we considering two-month interval, Δ_2, (Wilcoxon test). **P*<0.05, ***P*<0.01, ****P<0.001*

3.2.2.3. Beta diversity

Then, we analysed if the microbiota dynamics, after PR, changes with higher magnitude than in steady state, i.e., if the microbiota heterogeneity varies with PR. As in previous sections, we compared the distances between pairs of samples (beta diversity)

of M2-M1, M3-M2 (steady state) and M4-M3 (transition between steady state and PR) in cases of Δ_1, and M3-M1 (steady state variation) with M5-M3 (transition between steady state and PR) in cases of Δ_2 (Fig. 12 and Appendix VI) .

No significant differences were obtained for all the comparisons, Δ_1 and Δ_2, made using weighted Unifrac distance metric (Fig. 12). Similar results were observed using different beta diversity metrics such unweighted Unifrac distance, Jaccard distance and Bray Curtis distance (appendix VI).

These results suggest that, at least until month 5, PR does not change beta diversity considerably, meaning that group heterogeneity remains the same than in steady state. However, once again, the sample is very small, and we must be careful in the conclusions we draw from these results.

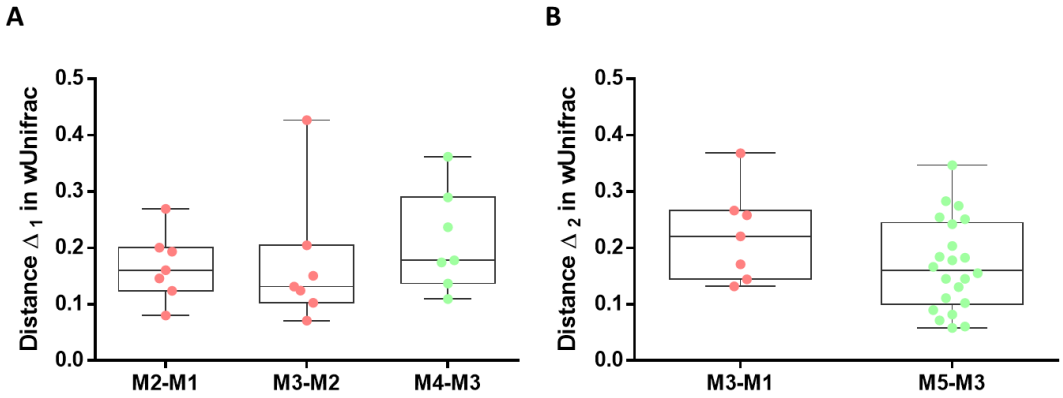

Figure 12. Pairwise distance comparisons in beta diversity (weighted Unifrac distance) A) between successive timepoints, M2-M1 and M3-M2 (steady state), and M4-M3 (transition between steady state and PR) and B) between timepoints with two months difference M3-M1 (steady state) and M5-M3 (transition between steady state and PR). No significant difference was observed (Friedman test with Dunn's multiple comparison and Wilcoxon test, respectively).

3.2.3. Temporal dynamics of microbiota composition

3.2.3.1. Alfa diversity

Next, we tested if alfa diversity changes over time within and between groups (intervention patients and control patients) and if this change was in response to PR. For this we tested changes from the 1st to 6th month in alfa diversity (Faith's Phylogenetic

Diversity, Shannon Observed OTUs and Evenness). We have only examined a 6 month period since control patients were only followed for 6 months.

We also explored the first difference to baseline method, which calculates the difference of each time point to a static time point (baseline) and can help us to assess how a person differs from the start of the treatment. M3 was the baseline for the intervention group, since data from M1 and M2 was missing in most patients. Furthermore, for this analysis only the time interval of M3 to M6 of the intervention group was considered for comparisons with M1 to M4 period of the control group.

Regression plots show the relationship between time and metrics of alfa diversity. To test if changes were caused by PR, we used LME model. The results of LME model for Faith's Phylogenetic Diversity in the period of 1st to 6th month (Fig. 13A) showed that time had a significant impact (*P*=0.026) in phylogenetic diversity (Table 2), reducing Faith's Phylogenetic diversity. The same was observed for Shannon's diversity (*P*=0.021) (Fig. S 22A and Table S 1) and for OTUs (*P*=0.020) (Fig. S 23A and Table S 3). The results of LME model for Evenness in the period of 1st to 6th months (Fig. S 24A, Table S 5) showed that no factor had a significant impact in Evenness.

The results of LME model for Faith's Phylogenetic Diversity, Observed OTUs and Evenness, produced with difference to baseline method (Fig. 13B, S 23B and S 24B), showed that no factor had a significant impact in alfa-diversity (Table 3, S 4 and S 6). The results of LME model for Shannon's diversity, produced with difference to baseline method (Fig. S 22B), showed that the interaction between exposure to PR and time impacts Shannon index (Table S 2), indicate that daily diversity (Shannon' diversity) decreased by a factor of 0.01 in the control group.

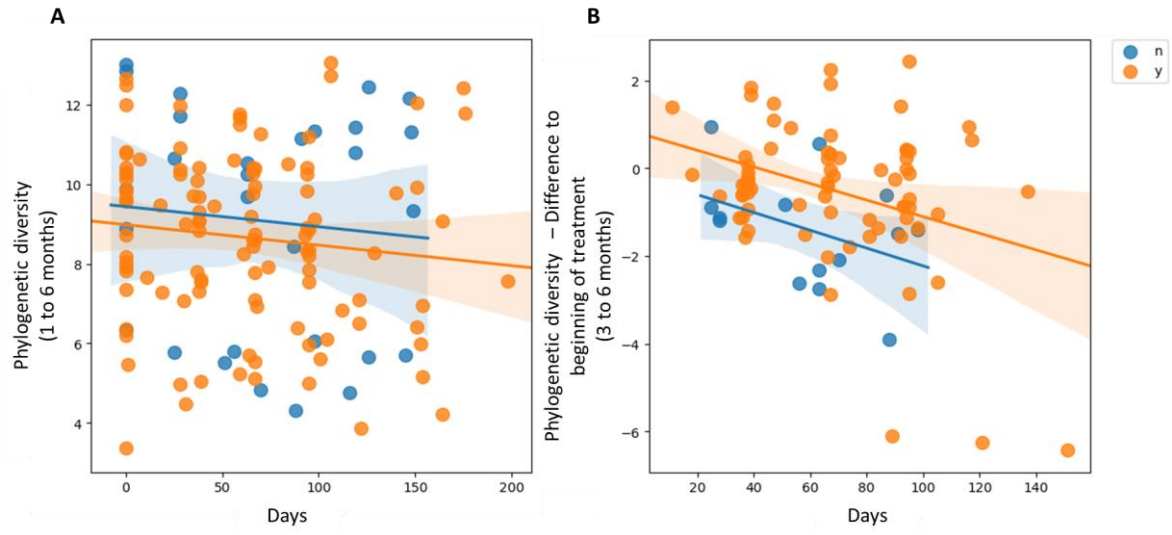

Figure 13. Alfa diversity regression plots **A)** over time in 1 to 6 months and **B)** difference of each time point to a baseline (3 to 6 month). **A)** Phylogenetic diversity decreases over time in both groups (intervention (y) and control group (n)). **B)** A trend towards a decrease of the difference of phylogenetic diversity to baseline (month 3) was observed in both groups over time.

Table 2. Linear mixed-effects model results for Faith's Phylogenetic Diversity between 1 and 6 months

Variable or parameter	**Coef.**	**Std.Err.**	**Z-score**	***P* value**
Intercept	9.542	0.976	9.776	0.000
Patientexposed[T.y]	-0.527	1.070	-0.493	0.622
Days	-0.006	0.004	-2.233	0.026
Days:Patientexposed[T.y]	0.002	0.004	0.418	0.676
Group Var	4.272	1.415		
Parameter estimate (coefficient), standard error, Z score, and *P* value for each model parameter.				

Table 3. Linear mixed-effects model results for Faith difference between samples collected between 4 and 6 months to baseline (month 3)

Variable or parameter	**Coef.**	**Std.Err.**	**Z-score**	***P* value**
Intercept	-0.597	0.827	-0.723	0.470

Patientexposed[T.y]	0.402	0.911	0.441	0.659
Days	-0.015	0.009	-1.762	0.078
Days:Patientexposed[T.y]	0.010	0.009	1.073	0.283
Group Var	1.836	0.885		
Parameter estimate (coefficient), standard error, Z score, and *P* value for each model parameter.				

In sum, alpha diversity shows a tendency to decrease over time. This decrease is due to the impact that time has on diversity (except Evenness), which is the same in both groups (intervention and control groups). Pulmonary microbiota diversity in COPD and in cystic fibrosis was found to be lower in older individuals [87], [96], [97]. So, the time period studied was sufficiently large to see a significant decline in diversity.

3.2.3.2. Beta diversity

Next, we tested if the dynamic of microbiota changed over time within and between groups, i.e., we tested if patients become more homogeneous over time and if this change was in response to PR. We applied first distances to examine how beta diversity (weighted Unifrac distance, unweighted Unifrac distance, Jaccard distance and Bray Curtis distance between successive samples collected from the same subject) changed over time. The first distance allows the analysis of individuals' rates of incremental change between time points [59], identifies the beta diversity distances between successive samples from the same subjects. We also examine beta-diversity in a period of 6 months, in the intervention group. We applied to the first distance to baseline, this method calculate distance from a static time point. Similar to what we did using alpha diversity, we considered baseline month 3 of the intervention patients and only analyse the time interval of month 3 to month 6 in intervention patients and month 1 to month 4 in control patients.

Regression plots show the relationship between time and metrics of beta diversity. To test if the change was in response to pulmonary rehabilitation, we use LME model. When we applied first distance in weighted Unifrac between successive samples

collected from the same individual (Fig. 14A), results of LME model indicated that no factor had a significant impact on weighted Unifrac distance (Table 4). The results of LME model for weighted Unifrac distance to baseline (Fig. 14B) also indicated that no factor had a significant impact (Table 5). Similar findings were obtained using unweighted Unifrac (Fig. S25 and Table S7 and S8).

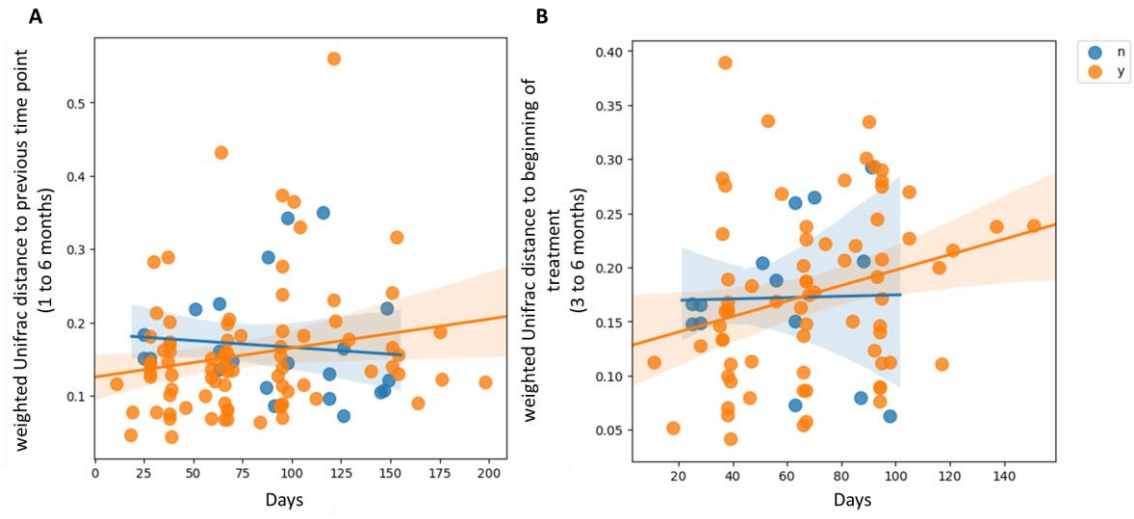

Figure 14. Beta diversity (weighted Unifrac distance) regression plots of A) distances to a previous time point (1 to 6 months) and B) distance of each time point to a baseline (3 to 6 month). No significant difference was observed.

Table 4. Linear mixed-effects model results for weighted Unifrac distance between successive samples collected between 1 and 6 months

Variable or parameter	**Coef.**	**Std.Err.**	**Z-score**	***P* value**
Intercept	0.191	0.043	4.480	0.000
Patientexposed[T.y]	-0.061	0.047	-1.290	0.179
Days	-0.000	0.000	-0.585	0.558
Days:Patientexposed[T.y]	0.001	0.000	1.160	0.246
Group Var	0.002	0.013		
Parameter estimate (coefficient), standard error, Z score, and *P* value for each model parameter.				

Table 5. Linear mixed-effects model results for weighted Unifrac distance between samples collected between 4 and 6 months to baseline (month 3)

Variable or parameter	Coef.	Std.Err.	Z-score	*P* value
Intercept	0.165	0.045	3.636	0.000
Patientexposed[T.y]	-0.028	0.050	-0.551	0.582
Days	0.000	0.001	0.234	0.815
Days:Patientexposed[T.y]	0.000	0.001	0.667	0.505
Group Var	0.003	0.025		
Parameter estimate (coefficient), standard error, Z score, and *P* value for each model parameter.				

In Jaccard distance between successive samples collected from the same person in period of 1 to 6 months, we observed that in the control group the beta diversity distances between successive timepoints decrease over time (Fig. S 26A). This difference was significantly impacted by time (*P*=0.002) and interaction between time and patient exposure (having or not PR) (*P*=0.022), indicating that daily Jaccard distance to previous timepoint decreased by a factor of 0.001 in the control group (Table S 9). When we applied Jaccard distance to baseline method (Fig. S 26B), a significant association was observed with patient exposure (*P*=0.020), indicating more distance to baseline in those composing the control group (by a factor of 0.153) (Table S 10).

When we used Bray Curtis distance between successive samples collected from the same person (Fig. S 27A), the LME models showed that no factors significantly impacted Bray Curtis distance (Table S 11). The results of LME model Bray Curtis distance to baseline (Fig. S 27B) showed an increase of distance to a baseline over time in the intervention group and a decrease in the control group. This difference was significantly impacted by patient exposure (P=0.005), indicating more Bray Curtis distance to baseline in the control group (by a factor of 0.229). Moreover, a significant interaction between time and patient exposure (*P*=0.023) was observed, indicate that daily Bray Curtis distance decreased by a factor of 0.002 in the control group (Table S 12).

So, in Unifrac distances, the time and exposure to PR no has impact in the metrics, suggesting that there are no significant differences between the two groups (intervention and control group), and therefore, exposure to PR seems not to be a predictor of changes in Unifrac distances. In Jaccard distance, time and exposure to PR had an impact in changes of the beta diversity distances between successive timepoints, so exposure to PR seems to be a predictor of changes in Jaccard distance to previous timepoint over time; patients who did not undertake PR were over time becoming more homogeneous than patients who participated in PR. When we compared to baseline, only PR had an impact, and it was the control group who had the greatest distance to the baseline, as it could be expected, since it was also this group that narrowed the gap over time. Using Bray Curtis distance, we observed that no factors significantly impacted Bray Curtis distance between timepoints successive. Bray Curtis distance to baseline are influenced by exposure to PR (by itself), indicating more Bray Curtis distance to baseline in control group, and by interaction between time and patient exposure, indicate that control group daily decreases Bray Curtis distance to baseline and therefore becoming more homogeneous compared to baseline than intervention group.

3.2.4. Does the microbiota converge to a similar composition among patients during pulmonary rehabilitation?

We wondered if during PR patients' microbiota converged to a similar state.

We generated a principle coordinates analysis (PCoA) plots with samples of Pre and PR periods, and tested if distances within groups were smaller than distances between groups, using PERMANOVA [58].

We chose PCoA plots of weighted Unifrac matrix because it is the measure that better explains the differences between groups (30.5%). Looking at the plot we were unable to visualize clustering by period and beta significance results indicated that there was no clustering (p=0.261, PERMANOVA), i.e., the distance between the samples from the PR period was similar to the distance from the samples from the Pre period (Fig. 15). Using the other metrics of beta diversity, the same was observed (appendix IX).

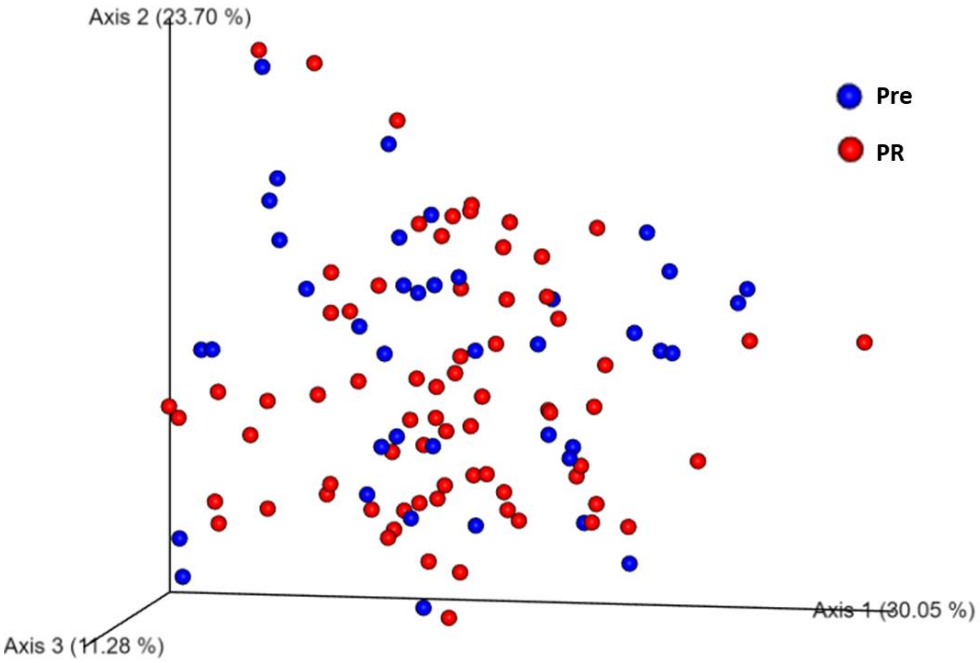

Figure 15. Microbiota of the different patients does not converge to a similar composition during pulmonary rehabilitation. Saliva's microbiota composition of samples of Pre-PR period is not significantly different from the saliva's microbiota composition of samples of during PR (PCoA analysis using Emperor of Weighted UniFrac distance matrix, PERMANOVA, *P*=0.261).

In conclusion, there is no clustering per period, so there was no convergence during rehabilitation. COPD is a complex and rather heterogeneous disease in terms of microbiota composition, that is, there are various forms of disease, whereas the healthy state is more homogeneous (unpublished data). Therefore, we expected to observe a convergence during PR, meaning that the treatment would make the patients' microbiota less heterogeneous and perhaps bring it closer to a healthy state. This has not happened, and this may mean that the impact of PR is not enough for the microbiota of patient samples to be more similar between samples from the same period than to samples from the same individual.

3.2.5. Does the microbiota converge to a similar composition among patients after pulmonary rehabilitation?

We wondered if convergence to a similar state of the patient's microbiota occurs after pulmonary rehabilitation. We generated a PCoA plots with samples of PR and post periods and we calculate beta diversity (PERMANOVA).

PCoA plots of weighted Unifrac matrix is the measure that better explains the differences between the groups (31,85%). Looking at the plot we also are unable to visualize clustering by period and beta significance results confirms that there is no clustering (p=0.899, PERMANOVA), that is, the distance between the samples from the PR period is similar to their distance from the samples from the post period (Fig. 16). Using the matrix of the other metrics of beta diversity, the same was observed (appendix X).

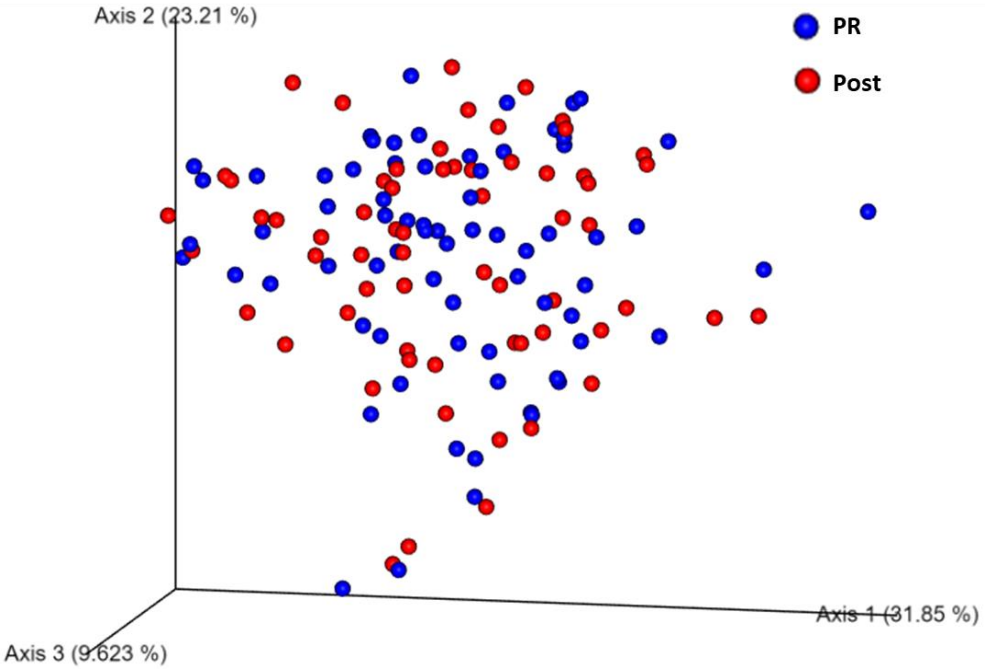

Figure 16. Microbiota of the different patients doesn't converge to a similar composition after Pulmonary Rehabilitation. Saliva's microbiota composition of samples of PR period isn't significantly different from the saliva's microbiota composition of samples of Post-PR period (PCoA analysis using Emperor of Weighted UniFrac distance matrix, PERMANOVA, *P*=0.261).

So, PERMANOVA showed that there is no clustering per period, so there would be no convergence after pulmonary rehabilitation.

3.3. Limitations

This study has several limitations. Firstly, the quality and quantity of the saliva samples. Impurities and sputum contamination were detected in a few samples and, in some cases, led to QIAamp Mini spin column obstruction. To overcome this, we increased the time of vortexing and the number of washing steps, may have decreased the

concentration of the samples. In some cases, there was an insufficient amount of sample due to the difficulty that patients presented in producing saliva as a consequence of the effects of bronchodilators [98]. Another limitation was the absence of samples in the pre-PR period which limited the comparison analysis between different periods of time in the same group of patients. Increasing the sample size to strengthen our observations is recommended, especially of the he control group (which in this study was very small) matching mainly with age, which may be a fact with impact on the microbiota.

4. Conclusion

In sum, PR does not significantly alter the structure of microbiota (alpha diversity) but changes its composition. We observed an increase in the phylum Proteobacteria, particularly the genus *Neisseria*, as well as the genus *Alloscardovia* and a decrease in the *Pseudomonas* and *Shingomonas* genera upon PR compared to the previous period. However, changes in microbiota composition do not appear to be towards a healthy state. These changes do not persist after it ceases.

In general, alpha diversity seems to be influenced by the passage of time, decreasing over time, but not by PR.

The analyses of beta diversity showed that microbiota does not converge, among the intervention group during PR neither after PR. LME model showed that the interaction between PR and time passage are mainly contributing for microbiota dynamics, in non-phylogenetic metrics.

Future studies should also aim at understanding whether the observed changes in microbiota are correlated with clinical symptoms improvement and / or inflammatory markers changes. Integrating the several aspects of the disease (microbial, clinical and inflammatory) will contribute to enhance our understanding of the benefits of PR (the most cost-effective therapy for COPD) and find better ways of improving patients' health status.

5. References

[1] WHO, “Global Health Estimates 2016: Deaths by Cause, Age, Sex, by Country and by

Region, 2000-2016," 2018.
[2] GOLD, "Global Strategy for the Diagnosis, Management and Prevention of COPD," *Glob. Initiat. Chronic Obstr. Lung Dis.*, 2019.
[3] A. C. Santos, "Panorama das doenças respiratórias em Portugal," 2018.
[4] J. L. López-Campos, W. Tan, and J. B. Soriano, "Global burden of COPD," *Respirology*, vol. 21, no. 1, pp. 14–23, 2016.
[5] P. J. Barnes *et al.*, "Chronic obstructive pulmonary disease," *Nat. Rev. Dis. Prim.*, vol. 1, pp. 1–22, 2015.
[6] D. Price *et al.*, "Spirometry in primary care case-identification , diagnosis and management of COPD," *Prim. Care Respir. J.*, vol. 18, no. 3, pp. 216–223, 2009.
[7] K. F. Rabe and H. Watz, "Chronic obstructive pulmonary disease," *Lancet*, vol. 389, pp. 1931–1940, 2017.
[8] S. Dhandha and W. A. Hogge, "Alpha 1-antitrypsin deficiency," *Lancet*, vol. 365, no. 2, pp. 2225–36, 2005.
[9] D. N. O'Dwyer, R. P. Dickson, and B. B. Moore, "The Lung Microbiome, Immunity and the Pathogenesis of Chronic Lung Disease," *J Immunol.*, vol. 196, no. 12, pp. 4839–4847, 2016.
[10] M. A. Spruit *et al.*, "An Official American Thoracic Society / European Respiratory Society Statement : Key Concepts and Advances in Pulmonary Rehabilitation," *Am. J. Respir. Crit. Care Med.*, vol. 188, no. 8, pp. e13–e64, 2013.
[11] S. D. Shukla, K. F. Budden, R. Neal, and P. M. Hansbro, "Microbiome effects on immunity, health and disease in the lung," *Clin. Transl. Immunol.*, vol. 6, no. 3, pp. e133-12, 2017.
[12] K. Miki, R. Maekura, S. Kitada, and K. Yoshimura, "Pulmonary rehabilitation for COPD improves exercise time rather than exercise tolerance : effects and mechanisms," *Int. J. COPD*, vol. 12, pp. 1061–1070, 2017.
[13] D. E. O'Donnell, M. McGuire, L. Samis, and K. A. Webb, "General Exercise Training Improves Ventilatory and Peripheral Muscle Strength and Endurance in Chronic Airflow Limitation," *Am. J. Respir. Crit. Care Med.*, vol. 157, no. 12, pp. 1489–1497, 1998.
[14] P. Laveneziana and P. Palange, "Physical activity , nutritional status and systemic inflammation in COPD," *Eur. J. Epidemiol.*, vol. 40, no. 3, pp. 522–529, 2012.
[15] O. O. Sullivan *et al.*, "Exercise and the microbiota," *Gut Microbes*, vol. 6, no. 2, pp. 131–136, 2015.
[16] A. Schwiertz and V. Rusch, "A Short Definition of Terms," in *Advances in Experimental Medicine and Biology*, 2016, pp. 1–3.
[17] I. Cho and M. J. Blaser, "The Human Microbiome: at the interface of health and disease," *Nat. Rev. Genet.*, vol. 13, no. 4, pp. 260–270, 2012.
[18] H. A. Filyk and L. C. Osborne, "The Multibiome: The Intestinal Ecosystem's Influence on Immune Homeostasis, Health, and Disease," *EBioMedicine*, vol. 13, pp. 46–54, 2016.
[19] R. Sender, S. Fuchs, and R. Milo, "Revised Estimates for the Number of Human and Bacteria Cells in the Body," *PLoS Biol.*, vol. 14, no. 8, pp. 1–14, 2016.
[20] P. J. Turnbaugh, R. E. Ley, M. Hamady, C. Fraser-liggett, R. Knight, and J. I. Gordon, "The human microbiome project: exploring the microbial part of ourselves in a changing world," *Nature*, vol. 449, no. 7164, pp. 804–810, 2007.
[21] R. P. Dickson, J. R. Erb-Downward, F. J. Martinez, and G. B. Huffnagle, "The Microbiome and the Respiratory Tract," *Annu. Rev. Physiol.*, vol. 78, no. 2, pp. 481–504, 2016.
[22] W. A. A. de Steenhuijsen Piters, E. A. M. Sanders, and D. Bogaert, "The role of the local microbial ecosystem in respiratory health and disease," *Philos. Trans. R. Soc. B*, vol. 370, no. 1675, 2015.
[23] T. P. Wypych, L. C. Wickramasinghe, and B. J. Marsland, "The influence of the microbiome on respiratory health," *Nat. Immunol.*, vol. 20, no. 10, pp. 1279–1290, 2019.
[24] W. H. Man, W. A. A. D. S. Piters, and D. Bogaert, "The microbiota of the respiratory tract:

gatekeeper to respiratory health," *Nat. Rev. Microbiol.*, vol. 15, no. 5, pp. 259–270, 2017.
[25] C. M. Bassis *et al.*, "Analysis of the Upper Respiratory Tract Microbiotas as the Source of the Lung and Gastric Microbiotas in Healthy Individuals," *MBio*, vol. 6, no. 2, pp. e00037-15, 2015.
[26] M. F. Moffatt and W. O. Cookson, "The Lung Microbiome in Health and Respiratory Diseases," *Clin. Med. (Northfield. Il).*, vol. 17, no. 6, pp. 525–9, 2017.
[27] D. Mayhew *et al.*, "Longitudinal profiling of the lung microbiome in the AERIS study demonstrates repeatability of bacterial and eosinophilic COPD exacerbations," *Thorax*, vol. 73, no. 5, pp. 422–430, 2018.
[28] Z. Wang *et al.*, "Sputum microbiome temporal variability and dysbiosis in chronic obstructive pulmonary disease exacerbations : an analysis of the COPDMAP study," *Thorax*, vol. 73, no. 4, pp. 1–8, 2018.
[29] L. Wang, K. Hao, T. Yang, and C. Wang, "Role of the lung microbiome in the pathogenesis of chronic obstructive pulmonary disease," *Chin. Med. J. (Engl).*, vol. 130, no. 17, pp. 2107–2111, 2017.
[30] M. L. K. Han *et al.*, "Significance of the microbiome in obstructive lung disease," *Thorax*, vol. 67, no. 5, pp. 456–463, 2012.
[31] S. Sethi and T. F. Murphy, "Infection in the Pathogenesis and Course of Chronic Obstructive Pulmonary Disease," *N. Engl. J. Med.*, vol. 359, no. 22, pp. 2355–2365, 2008.
[32] M. A. J. Mammen and S. A. Sethi, "COPD and the microbiome," *Respirology*, vol. 21, pp. 590–599, 2016.
[33] N. R. Anthonisen, "The British hypothesis revisited," *Eur. Respir. J.*, vol. 23, pp. 657–658, 2004.
[34] T. M. A. Wilkinson, I. S. Patel, M. Wilks, G. C. Donaldson, and J. A. Wedzicha, "Airway Bacterial Load and FEV 1 Decline in Patients with Chronic Obstructive Pulmonary Disease," *Am. J. Respir. Crit. Care Med.*, vol. 167, pp. 1090–1095, 2003.
[35] R. E. Kanner, N. R. Anthonisen, J. E. Connett, and L. Health, "Lower Respiratory Illnesses Promote FEV 1 Decline in Current Smokers But Not Ex-Smokers with Mild Chronic Obstructive Pulmonary Disease," *Am. J. Respir. Crit. Care Med.*, vol. 164, pp. 358–364, 2001.
[36] N. A. Clayton, G. D. Carnaby, M. J. Peters, and A. J. Ing, "Impaired laryngopharyngeal sensitivity in patients with COPD: The association with swallow function," *Int. J. Speech. Lang. Pathol.*, vol. 16, no. 6, pp. 615–623, 2014.
[37] R. D. Gross, C. W. Atwood, S. B. Ross, J. W. Olszewski, and K. A. Eichhorn, "The coordination of breathing and swallowing in chronic obstructive pulmonary disease," *Am. J. Respir. Crit. Care Med.*, vol. 179, no. 7, pp. 559–565, 2009.
[38] A. A. Pragman *et al.*, "The lung tissue microbiota of mild and moderate chronic obstructive pulmonary disease," *Microbiome*, vol. 6, no. 7, pp. 1–19, 2018.
[39] R. P. Dickson and G. B. Huffnagle, "The Lung Microbiome: New Principles for Respiratory Bacteriology in Health and Disease," *PLoS Pathog.*, vol. 11, no. 7, p. e1004923, 2015.
[40] Y. J. Huang, J. R. Erb-Downward, R. Dickson, J. L. Curtis, G. B. Huffnagle, and M. K. Han, "Understanding the role of the microbiome in COPD: Principles, Challenges and Future Directions," *Transl Res.*, vol. 179, pp. 71–83, 2017.
[41] E. S. Charlson *et al.*, "Topographical Continuity of Bacterial Populations in the Healthy Human Respiratory Tract," *Am. J. Respir. Crit. Care Med.*, vol. 184, pp. 957–963, 2011.
[42] A. A. Pragman, H. B. Kim, C. S. Reilly, C. Wendt, and R. E. Isaacson, "The Lung Microbiome in Moderate and Severe Chronic Obstructive Pulmonary Disease," *PLoS One*, vol. 7, no. 10, p. e47305, 2012.
[43] A. Morris *et al.*, "Comparison of the Respiratory Microbiome in Healthy Nonsmokers and Smokers," *Am. J. Respir. Crit. Care Med.*, vol. 187, no. 10, pp. 1067–1075, 2013.

[44] G. Chard, “International Classification of Functioning, Disability and Health,” *Br. J. Occup. Ther.*, vol. 67, no. 1, p. 1, 2004.
[45] M. R. Miller *et al.*, “Standardisation of spirometry,” *Eur. Respir. J.*, vol. 26, no. 2, pp. 319–338, 2005.
[46] F. H. M. George, “Diagnóstico e Tratamento da Doença Pulmonar Obstrutiva Crónica,” *Direção-Geral de Saúde*, 2013.
[47] T. Hajiro, K. Nishimura, M. Tsukino, A. Ikeda, H. Koyama, and T. Izumi, “Analysis of Clinical Methods Used to Evaluate Dyspnea in Patients with Chronic Obstructive Pulmonary Disease,” *Am. J. Respir. Crit. Care Med.*, vol. 158, no. 4, pp. 1185–1189, 1998.
[48] R. Geuther, “The Molecular Basis of Cell Structure and Function,” *Thermo Fish. Sci.*, vol. 17, no. 1, pp. 86–87, 2007.
[49] D. Wilmington, “Interpretation of Nucleic Acid 260/280 Ratios,” *Thermo Sci. Nanodrop Prod.*, pp. 302–479, 2012.
[50] G. B. Rogers, “The lung microbiome,” *Emerg. Top. Life Sci.*, vol. 1, pp. 313–324, 2017.
[51] N. A. Bokulich *et al.*, “Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing,” *Nat. Methods*, vol. 10, no. 1, pp. 57–59, Jan. 2013.
[52] A. Amir *et al.*, “Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns,” *mSystems*, vol. 2, no. 2, pp. e00191-16, /msys/2/2/e00191-16.atom, Apr. 2017.
[53] N. A. Bokulich *et al.*, “Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2’s q2-feature-classifier plugin,” *Microbiome*, vol. 6, no. 1, p. 90, Dec. 2018.
[54] D. McDonald *et al.*, “An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea,” *ISME J.*, vol. 6, no. 3, pp. 610–618, Mar. 2012.
[55] K. Katoh, K. Misawa, K. Kuma, and T. Miyata, “MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform,” *Nucleic Acids Res.*, vol. 30, no. 14, pp. 3059–3066, Jul. 2002.
[56] “FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments.” .
[57] B. D. Wagner *et al.*, “On the Use of Diversity Measures in Longitudinal Sequencing Studies of Microbial Communities,” *Front. Microbiol.*, vol. 9, no. 1037, 2018.
[58] M. J. Anderson, “A new method for non-parametric multivariate analysis of variance,” *Austral Ecol.*, vol. 26, no. 1, pp. 32–46, 2001.
[59] N. A. Bokulich *et al.*, “q2-longitudinal : Longitudinal and Paired-Sample Analyses of Microbiome Data,” *mSystems*, vol. 3, no. 6, pp. e00219-18, 2018.
[60] S. Mandal, W. Van Treuren, R. A. White, M. Eggesbø, R. Knight, and S. D. Peddada, “Analysis of composition of microbiomes: a novel method for studying microbial composition,” *Microb. Ecol. Health Dis.*, vol. 26, p. 27663, 2015.
[61] N. Segata *et al.*, “Metagenomic biomarker discovery and explanation,” *Genome Biol.*, vol. 12, no. 6, p. R60, Jun. 2011.
[62] G. B. Gloor, J. M. Macklaim, V. Pawlowsky-Glahn, and J. J. Egozcue, “Microbiome datasets are compositional: And this is not optional,” *Front. Microbiol.*, vol. 8, no. NOV, pp. 1–6, 2017.
[63] Mandal, Siddhartha, “Research - Dr. Siddhartha Mandal: ANCOM 2.0 - updated code.” .
[64] C. Huttenhower, “Galaxy / Hutlab - Harvard.” .
[65] J. A. Navas-Molina *et al.*, “Advancing our understanding of the human microbiome using QIIME,” *Methods Enzym.*, vol. 531, pp. 371–444, 2013.
[66] T. Z. DeSantis, P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, and K. Keller, “Greengenes , a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB,” *Appl. Environ. Microbiol.*, vol. 72, no. 7, pp. 5069–5072, 2006.
[67] C. Shannon and W. Weaver, “A Mathematical Theory of Communication,” *Univ. Illinois*

Press, vol. 11, pp. 1–117, 1964.
[68] E. C. Pielou, "The Measurement of Diversity in Different Types of Biological Colledions," *J. Theor. Biol.*, vol. 13, pp. 131–144, 1966.
[69] D. P. Faith, "Conservation evaluation and phylogenetic diversity," *Biol. Conserv.*, vol. 61, pp. 1–10, 1992.
[70] C. A. Lozupone, M. Hamady, S. T. Kelley, and R. Knight, "Quantitative and Qualitative β Diversity Measures Lead to Different Insights into Factors That Structure Microbial Communities," *Appl. Environ. Microbiol.*, vol. 73, no. 5, pp. 1576–1585, 2007.
[71] C. Lozupone and R. Knight, "UniFrac : a New Phylogenetic Method for Comparing Microbial Communities," *Appl. Environ. Microbiol.*, vol. 71, no. 12, pp. 8228–8235, 2005.
[72] T. Sorensen, "A method of establishing groups of equal amplitude in plant sociology based on similarity of species content," *K. Danske Vidensk. Selsk.*, vol. 5, no. 4, pp. 1–34, 1948.
[73] P. Jaccard, "Nouvellesrecherches sur la distribution florale," *Bull. Soc. V and. Sci. Nat*, vol. 44, pp. 223–270, 1908.
[74] G. G. Einarsson *et al.*, "Community dynamics and the lower airway microbiota in stable chronic obstructive pulmonary disease , smokers and healthy non-smokers," *Thorax*, vol. 71, pp. 795–803, 2016.
[75] J. R. Erb-downward *et al.*, "Analysis of the Lung Microbiome in the '" Healthy "' Smoker and in COPD," *PLoS One*, vol. 6, no. 2, p. e16384, 2011.
[76] L. Millares, X. Pomares, R. Ferrari, M. Gallego, and M. Espasa, "Severity-Related Changes of Bronchial Microbiome in Chronic Obstructive Pulmonary Disease," *J. Clin. Microbiol.*, vol. 52, no. 12, pp. 4217–4223, 2014.
[77] L. J. Caverly, Y. J. Huang, and M. A. Sze, "Past, Present, and Future Research on the Lung Microbiome in Inflammatory Airway Disease," *Chest J.*, vol. 156, no. 2, pp. 376–382, 2019.
[78] I. Mukhopadhya, R. Hansen, E. M. El-Omar, and G. L. Hold, "IBD-what role do Proteobacteria play?," *Nat. Rev. Gastroenterol. Hepatol.*, vol. 9, no. 4, pp. 219–230, 2012.
[79] G. Rizzatti, L. R. Lopetuso, G. Gibiino, C. Binda, and A. Gasbarrini, "Proteobacteria: A common factor in human diseases," *Biomed Res. Int.*, vol. 2017, pp. 1–7, 2017.
[80] K. F. Budden *et al.*, "Functional effects of the microbiota in chronic respiratory disease," *Lancet Respir.*, vol. 2600, no. 18, pp. 1–14, 2019.
[81] Y. Litvak, M. X. Byndloss, R. M. Tsolis, and A. J. Bäumler, "Dysbiotic Proteobacteria expansion: a microbial signature of epithelial dysfunction," *Curr. Opin. Microbiol.*, vol. 39, pp. 1–6, 2017.
[82] P. M. Clarkson and M. J. Hubal, "Exercise-Induced Muscle Damage," *Am. J. Phys. Med. Rehabil.*, vol. 81, no. 11, pp. 52–69, 2002.
[83] J. Canavan, R. Garrod, J. Marshall, D. Jackson, P. Ansley, and A. Jewell, "Measurement of the acute inflammatory response to walking exercise in COPD: Effects of pulmonary rehabilitation," *Int. J. COPD*, vol. 2, no. 3, pp. 347–353, 2007.
[84] H. A. C. van Helvoort, M. H. J. van de Pol, Y. F. Heijdra, and P. N. R. Dekhuijzen, "Systemic inflammatory response to exhaustive exercise in patients with chronic obstructive pulmonary disease," *Respir. Med.*, vol. 99, no. 12, pp. 1555–1567, 2005.
[85] A. Sciriha *et al.*, "Systemic inflammation in COPD is not influenced by pulmonary rehabilitation," *Eur. J. Physiother.*, vol. 19, no. 4, pp. 194–200, 2017.
[86] Y. Du, Q. Ke, G. B. Huffnagle, and M. R. Gyetko, "Symptom-related sputum microbiota in stable chronic obstructive pulmonary disease," *Int. J. COPD*, vol. 13, pp. 2289–2299, 2018.
[87] L. Millares *et al.*, "Relationship between the respiratory microbiome and the severity of airflow limitation, history of exacerbations and circulating eosinophils in COPD patients," *BMC Pulm. Med.*, vol. 19, no. 112, pp. 1–9, 2019.
[88] E. Jubinville *et al.*, "Exacerbation induces a microbiota shift in sputa of COPD patients," *PLoS One*, vol. 13, no. 3, pp. 1–16, 2018.

[89] M. J. Mammen and S. Sethi, “Microbiome in Chronic Lung Diseases,” *BRN Rev.*, vol. 3, pp. 102–120, 2017.
[90] T. Zakharkina *et al.*, “Analysis of the Airway Microbiota of Healthy Individuals and Patients with Chronic Obstructive Pulmonary Disease by T-RFLP and Clone Sequencing,” *PLoS One*, vol. 8, no. 7, p. e68302, 2013.
[91] S. Tudies, M. U. Sohail, H. M. Yassine, A. Sohail, and A. A. Al Thani, “Impact of Physical Exercise on Gut Microbiome , Inflammation , and the Pathobiology of Metabolic Disorders,” *Rev. Diabet. Stud.*, vol. 15, pp. 35–48, 2019.
[92] C. Gubert, G. Kong, T. Renoir, and J. Anthony, “Exercise, diet and stress as modulators of gut microbiota: Implications for neurodegenerative,” *Neurobiol. Dis.*, vol. 104621, 2019.
[93] L. Schmitz *et al.*, “Impact of endurance exercise and probiotic supplementation on the intestinal microbiota : a cross-over pilot study,” *Pilot Feasibility Stud.*, vol. 5, no. 76, pp. 1–8, 2019.
[94] F. Gallè, F. Valeriani, M. S. Cattaruzza, F. Ubaldi, and V. R. Spica, “Exploring the association between physical activity and gut microbiota composition : a review of current evidence,” *Ann Ig*, vol. 31, no. 6, pp. 582–589, 2019.
[95] T. Kern, M. Bæk, B. Tue, H. Hansen, M. Rosenkilde, and J. Salling, “Structured exercise alters the gut microbiota in humans with overweight and obesity — A randomized controlled trial,” *Int. J. Obes.*, no. 7, pp. 1476–5497, 2019.
[96] B. Coburn *et al.*, “Lung microbiota across age and disease stage in cystic fibrosis,” *Nature*, vol. 5, no. 10241, pp. 1–12, 2015.
[97] E. T. Zemanick *et al.*, “Airway microbiota across age and disease spectrum in cystic fibrosis,” *Eur. J. Epidemiol.*, vol. 50, no. 1700832, pp. 1–13, 2017.
[98] T. Hansel and P. Barnes, “Tiotropium bromide: a novel once-daily anticholinergic bronchodilator for the treatment of COPD,” *Drugs Today*, vol. 38, no. 9, pp. 585-600., 2002.
[99] M. A. Sze, J. C. Hogg, and D. D. Sin, “Bacterial microbiome of lungs in COPD,” *Int. J. COPD*, vol. 9, pp. 229–238, 2014.
[100] M. A. Sze *et al.*, “The Lung Tissue Microbiome in Chronic Obstructive Pulmonary Disease,” *Am. J. Respir. Crit. Care Med.*, vol. 185, no. 19, pp. 15–18, 2012.

6. Annex

Annex I - Ethics' approval

CENTRO HOSPITALAR DO BAIXO VOUGA, E.P.E. / AVEIRO

Avenida Artur Ravara – 3814-501 AVEIRO
Tel. 234 378 300 – Fax 234 378 395
sec-geral@chbv.min-saude.pt
Matrícula na Conservatória do Registo Comercial de Aveiro
Capital Social 40.284.651 €
Pessoa Colectiva nº 510 123 210

Exma. Senhora

Dra. Alda Sofia Pires de Dias Marques
Escola Superior de Saúde
Universidade de Aveiro
Agras do Crasto - Campus Universitário de Santiago
3810-193 Aveiro

S/ Ref.ª *S/ Comunicação de* *N/ Ref.ª* 777638 *Aveiro,* 22.03.2017

ASSUNTO: Resposta ao V/ Pedido de confirmação para a realização de estudo no CHBV, E.P.E.

Em resposta à V/ solicitação subordinada ao tema "GENIAL – Marcadores genéticos e clínicos na trajetória da DPOC", vimos, pelo presente, informar que por deliberação do Conselho de Administração, nesta data, se encontra autorizado o pedido formulado.

Nesse sentido, solicitamos a V. Exa se digne enviar um relatório final ao Serviço de Investigação e Formação do CHBV, E.P.E.

Com os melhores cumprimentos,

A Diretora do Serviço de Investigação e Formação

Joana Guimarães.

(Dra. Joana Guimarães)

Na resposta indicar o número e as referências deste documento. Em cada ofício tratar só de um assunto.

CHBV – 347

Autorização n.º 8828/ 2016

Universidade de Aveiro , NIPC 501461108, notificou à Comissão Nacional de Protecção de Dados (CNPD) um tratamento de dados pessoais com a finalidade de realizar um Estudo Clínico com Intervenção, denominado GENIAL - Marcadores Genéticos e Clínicos na Trajetória da DPOC .

Existe justificação específica, validada pela Comissão de Ética Competente (CEC), para o tratamento do dado pessoal raça/etnia.

O participante é identificado por um código especificamente criado para este estudo, constituído de modo a não permitir a imediata identificação do titular dos dados; designadamente, não são utilizados códigos que coincidam com os números de identificação, iniciais do nome, data de nascimento, número de telefone, ou resultem de uma composição simples desse tipo de dados. A chave da codificação só é conhecida do(s) investigador(es).

É recolhido o consentimento expresso do participante ou do seu representante legal.

A informação é recolhida diretamente do titular e indiretamente do processo clínico.

As eventuais transmissões de informação são efetuadas por referência ao código do participante, sendo, nessa medida, anónimas para o destinatário.

A CNPD já se pronunciou na Deliberação n.º 1704/2015 sobre o enquadramento legal, os fundamentos de legitimidade, os princípios aplicáveis para o correto cumprimento da Lei n.º 67/98, de 26 de outubro, alterada pela Lei n.º 103/2015, de 24 de agosto, doravante LPD, bem como sobre as condições e limites aplicáveis ao tratamento de dados efetuados para a finalidade de investigação clínica.

No caso em apreço, o tratamento objeto da notificação enquadra-se no âmbito daquela deliberação e o responsável declara expressamente que cumpre os limites e condições aplicáveis por força da LPD e da Lei n.º 21/2014, de 16 de abril, alterada pela Lei n.º 73/2015, de 27 de junho – Lei da Investigação Clínica –, explicitados na Deliberação n.º 1704/2015.

O fundamento de legitimidade é o consentimento do titular.

Lisboa, 23-08-2016

A Presidente

Filipa Calvão

7. Appendixes

Appendix I – DNA Extraction and Quantification Protocol

First, 400µL of saliva was centrifugated at 10000g, at 4°C for 10 minutes. The supernatant was stored in aliquots of 120µL at -80°C. The pellet follows to DNA extraction. Then, 350µL of PBS and 40µL of Qiagen protease was added. Next, 400µL of buffer AL was added to lysis cells. To ensure an efficient lysis, the microtube with the samples and buffer AL were mixed by vortexing for 15 seconds and incubated at 56°C for 10 minutes. After incubation, samples were briefly spinned to remove drops produced by the condensation. In order to precipitate DNA, 400µL of 100% ethanol were added to microtube followed by a 15 seconds vortexing and spinning to remove droplets. Afterwards, the mixture of microtube was carefully applied in a QIAamp Mini spin column. The column was washed with both washing buffers: AW1 and AW2. Finally, the DNA was eluted with 100µL of milli Q water, incubated at room temperature (15-25 °C) for 1 minute to assure an efficient elution and the final extract was obtained by centrifugation at 6000 g for 1 minute.

Protocol:

Each patient has 2 microtubes of saliva.
The microtube containing the most sample will be used in this protocol.

Saliva samples are stored at -80 ° C and should be thawed at room temperature.

From the microtube chosen to continue the protocol, 400 ul will be used to extract DNA (Point 1). The remaining saliva should be centrifuged for 10 min at 10 000 g at 4°C, the supernatant collected and stored in 120 µl aliquots at -80 ° C. The pellet should be stored at -80ºC.

1. **Centrifuge the 400 µl of saliva for 10 min at 10,000 g at 4 ° C**
2. **Collect about 350 µl of supernatant and store in 120 µl aliquots at -80 ° C.**
3. **add the pellet with about 350 µl of PBS.**
4. **Pipette 40 µl Quiagen protease into the tube with remaining saliva / pellet**
5. **Add 400 µl Buffer AL to the pellet and vortex 15 seconds**

To ensure efficient lysis, it is essential that the sample and Buffer AL are thoroughly mixed to produce a homogeneous solution.

Note: Do not add QIAGEN protease or proteinase K directly to Buffer AL.

6. **Incubate at 56 ° C for 10 minutes**

DNA yield reaches a maximum after lysis for 10 min at 56°C. Longer incubation times have no effect on yield or quality of the purified DNA.

7. **Make a spin (just to remove the drops from the lid)**
8. **Add 400ul of ethanol (compatible with sample volume) and mix again by pulse-vortexing for 15 s. Then spin to remove e to remove drops from the inside of the lid.**

If the sample volume is greater than 200 µl, increase the amount of ethanol proportionally; for example, a 400 µl sample will require 400 µl of ethanol.

9. **Carefully apply the mixture from step 6 to the QIAamp Mini spin column (in a 2 ml collection tube) without wetting the rim. Close the cap, and centrifuge at 6000 g for 1 min. Place the QIAamp Mini spin column in a clean 2 ml collection tube (provided), and discard the tube containing the filtrate.**

If the preparation is larger than the column, repeat this step as many times as necessary.

Close each spin column to avoid aerosol formation during centrifugation.

Centrifugation is performed at 6000 g (8000 rpm) to reduce noise. Centrifugation at full speed will not affect the yield or purity of the DNA. If the lysate has not completely passed through the column after centrifugation, centrifuge again at higher speed until the QIAamp Mini spin column is empty.

Note: When preparing DNA from buffy coat or lymphocytes, centrifugation at full speed is recommended to avoid clogging.

10. **Add 500 µl of Buffer AW1 and centrifuge at 6000 g (8000 rpm) for 1 minute.**

It is not necessary to increase the volume of Buffer AW1 if the original sample volume is larger than 200 µl.

11. **Add 500 µl of Buffer AW2 and centrifuge at 20000 g (14000 rpm) for 3 minutes**

12. **Place the QIAamp Mini spin column in a new 2 ml collection tube (not provided) and discard the old collection tube with the filtrate. Centrifuge at full speed for 1 min.**

This step helps to eliminate the chance of possible Buffer AW2 carryover

13. **Place the column in a clean 1.5 ml tube and add 50 µl H20 MilliQ. Incubate at room temperature (15–25°C) for 1 min, and then centrifuge at 6000 x g (8000 rpm) for 1 min.**

Repeat this step 2 more times.

14. **Quantify the DNA by reading the optical density at 260nm and 280nm.**

Appendix II – Relative frequency plots, by month, of top 5 phyla and most representative genera of each phyla

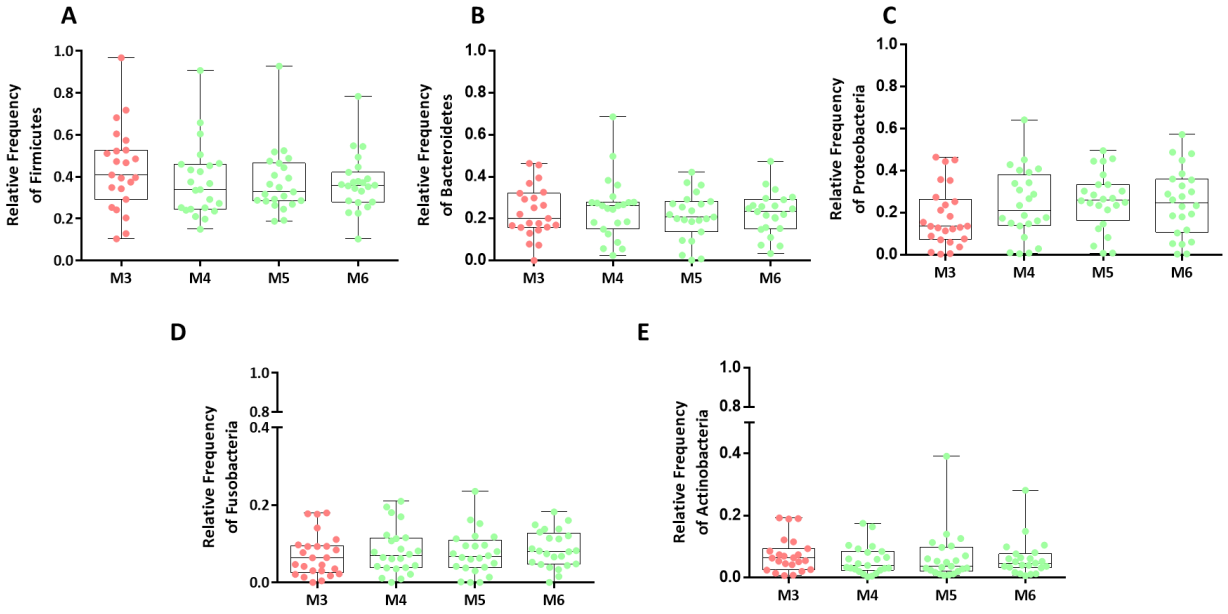

Figure S 1. Relative frequency of A) Firmicutes, B) Bacteroidetes, C) Proteobacteria, D) Fusobacteria and E) Actinobacteria phyla in months M3 (before PR), M4, M5 and M6 (during PR). No significant differences were found when we compared M3 (pre-PR) with M4, M5 and M6 (three months of PR) (Friedman test with Dunn's with multiple comparisons: Firmicutes *P*=0.3387, Bacteroidetes *P*=0.3916, Proteobacteria *P*=0.0730, Fusobacteria *P*=0.8055, Actinobacteria *P*=0.1415).

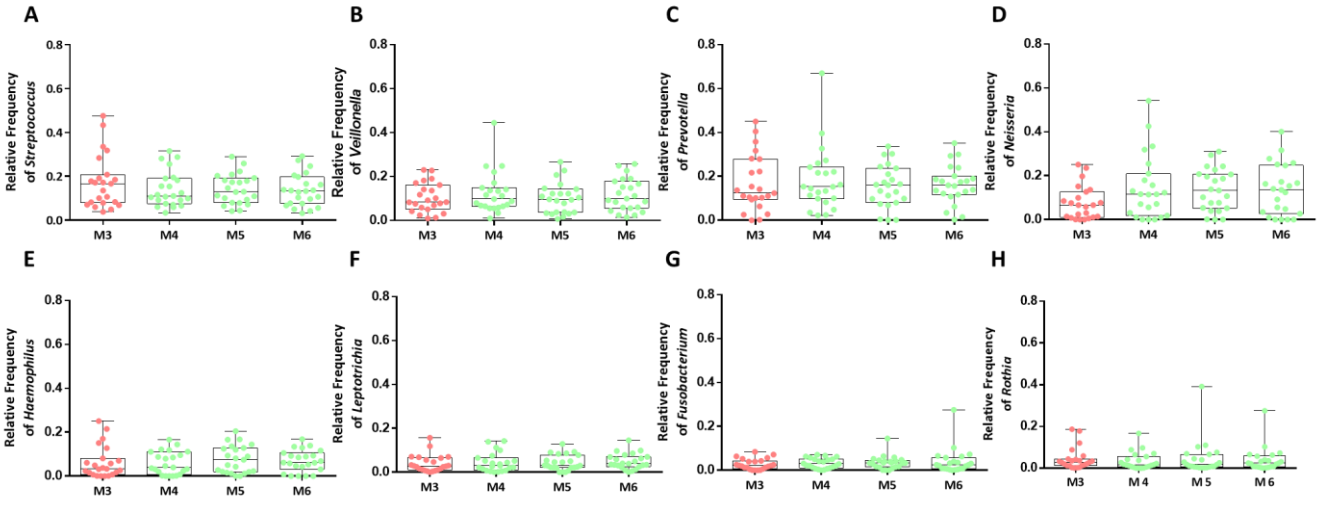

Figure S 2. Relative frequency of **A)** *Streptococcus*, **B)** *Veillonella*, **C)** *Prevotella*, **D)** *Neisseria*, **E)** *Haemophilus*, **F)** *Leptotrichia*, **G)** *Fusobacterium* and **H)** *Rothia* genera in months M3 (before PR), M4, M5 and M6 (during PR). No significant differences were found when we compared M3 (pre-PR) with M4, M5

and M6 (three months of PR) (Friedman test with Dunn's with multiple comparisons: *Streptococcus* *P*=0.4698, *Veillonella* *P*=0.5925, *Prevotella* *P*=0.6379, *Neisseria* *P*=0.0898, *Haemophilus* *P*=0.7361, *Leptotrichia* *P*=0.7141, *Fusobacterium* *P*=0.2610 and *Rothia* *P*=0.1993).

Appendix III – Alfa diversity plots, by period and month, estimated with Shannon diversity index, observed OTUs and evenness

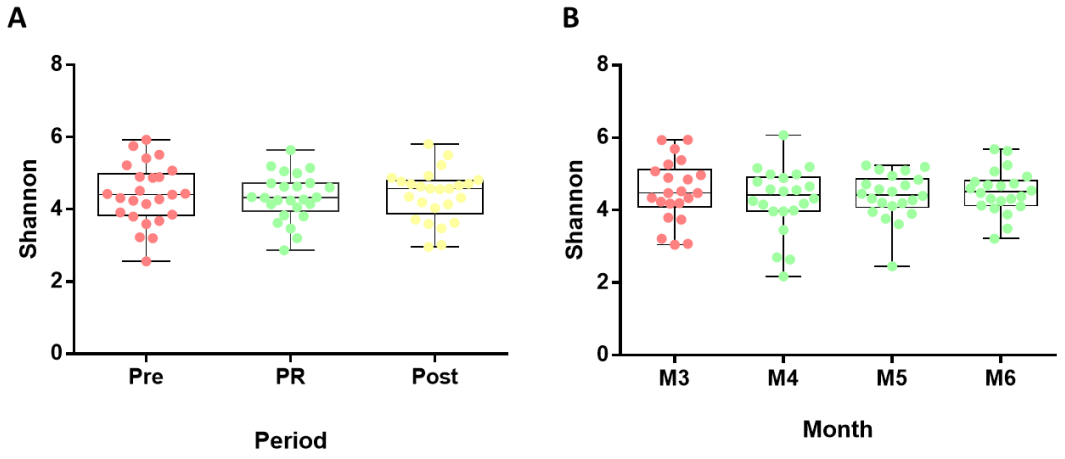

Figure S 3. Mean of alpha diversity estimated with Shannon's diversity index measured **A** all the time points of each tree periods and **B)** in months, M3 (before PR), M4, M5 and M6 (during PR). No significant differences were observed (Friedman test with Dunn's multiple comparisons).

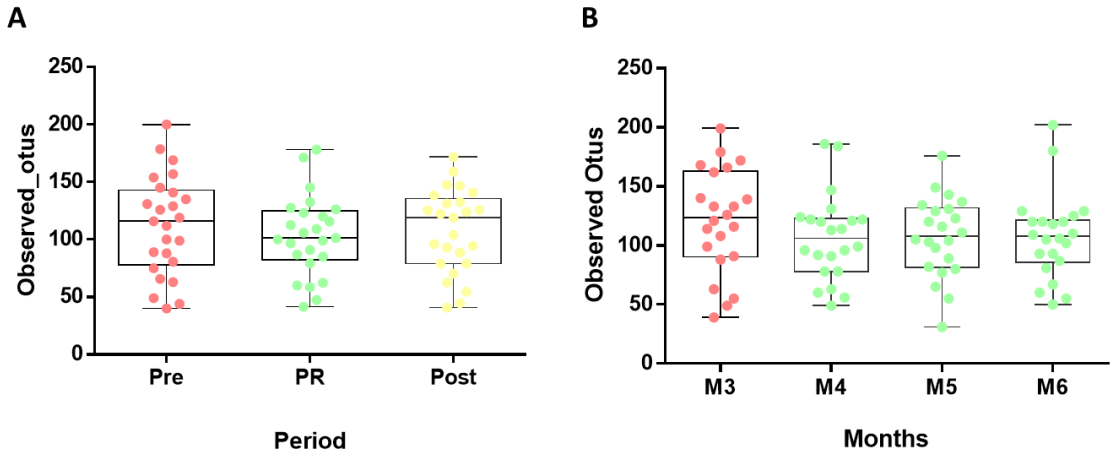

Figure S 4. Mean of alpha diversity estimated with observed OTUs measured **A** all the time points of each tree periods and **B)** in months, M3 (before PR), M4, M5 and M6 (during PR). No significant differences were observed (Friedman test with Dunn's multiple comparisons.)

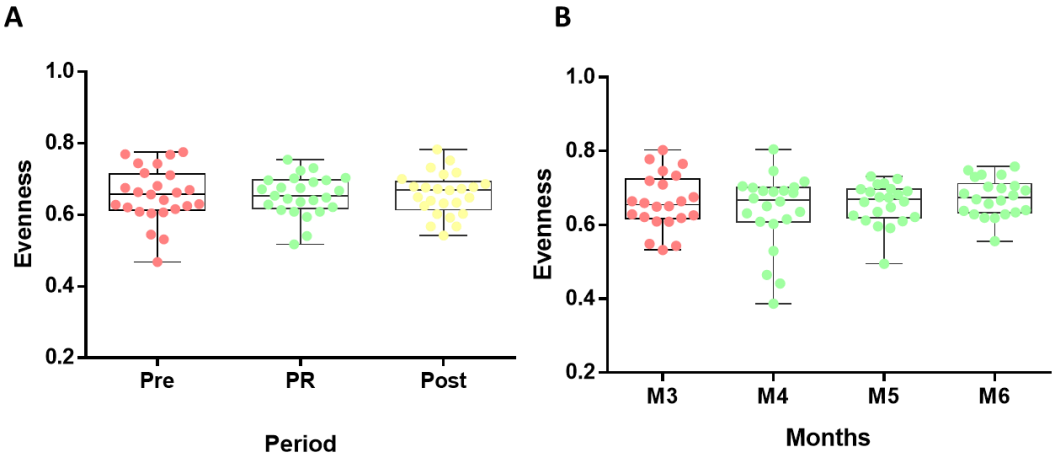

Figure S 5. Mean of alpha diversity estimated with Evenness measured **A** all the time points of each tree periods and **B)** in months, M3 (before PR), M4, M5 and M6 (during PR). No significant differences were observed (Friedman test with Dunn's multiple comparisons).

Appendix IV – Differences in relative frequencies of phyla and most representative genera of each phyla, between successive timepoints and between timepoints with two months of difference

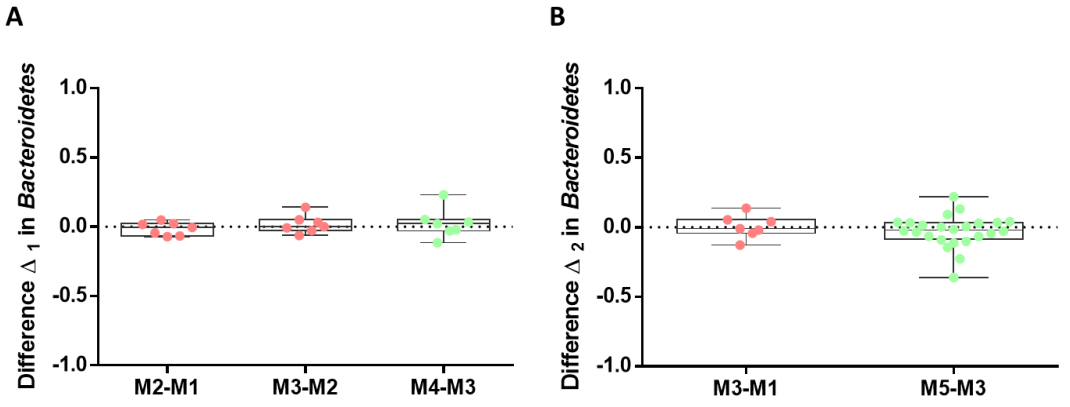

Figure S 6. Differences in relative frequencies of Bacteroidetes phylum **A)** between successive timepoints, M2-M1 and M3-M2 (steady state), and M4-M3 (transition between steady state and PR) and **B)** between timepoints with two months difference M3-M1 (steady state) and M5-M3 (transition between steady state and PR). No significant differences were observed (Friedman test with Dunn's multiple comparisons and Wilcoxon test, respectively).

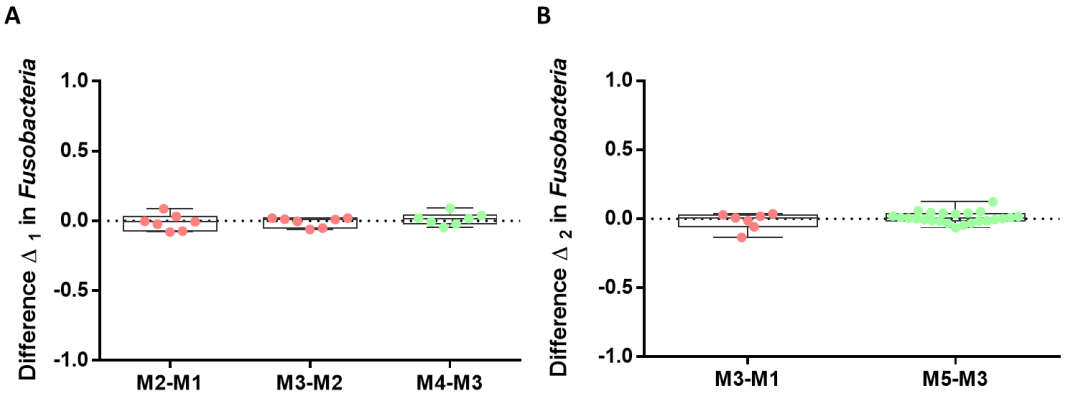

Figure S 7. Differences in relative frequencies of Fusobacteria phylum **A)** between successive timepoints, M2-M1 and M3-M2 (steady state), and M4-M3 (transition between steady state and PR) and **B)** between timepoints with two months difference M3-M1 (steady state) and M5-M3 (transition between steady state and PR). No significant differences were observed (Friedman test with Dunn's multiple comparisons and Wilcoxon test, respectively).

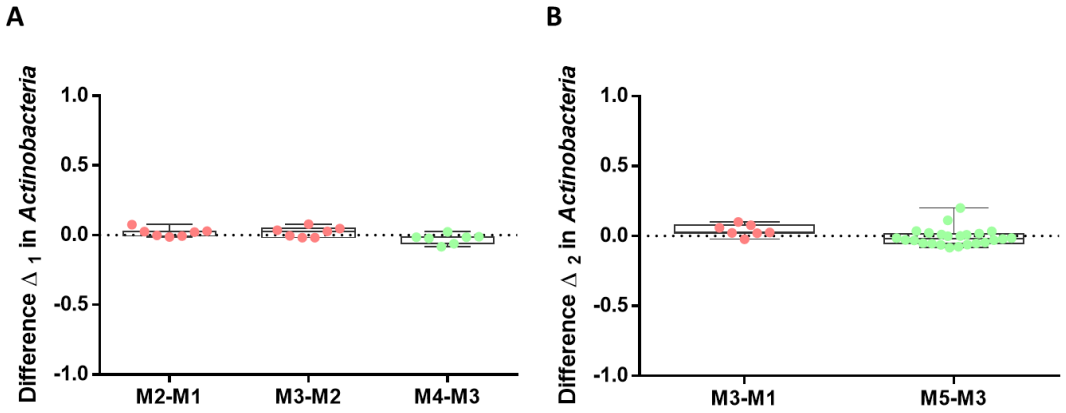

Figure S 8. Differences in relative frequencies of Actinobacteria phylum **A)** between successive timepoints, M2-M1 and M3-M2 (steady state), and M4-M3 (transition between steady state and PR) and **B)** between timepoints with two months difference M3-M1 (steady state) and M5-M3 (transition between steady state and PR). No significant differences were observed (Friedman test with Dunn's multiple comparisons and Wilcoxon test, respectively).

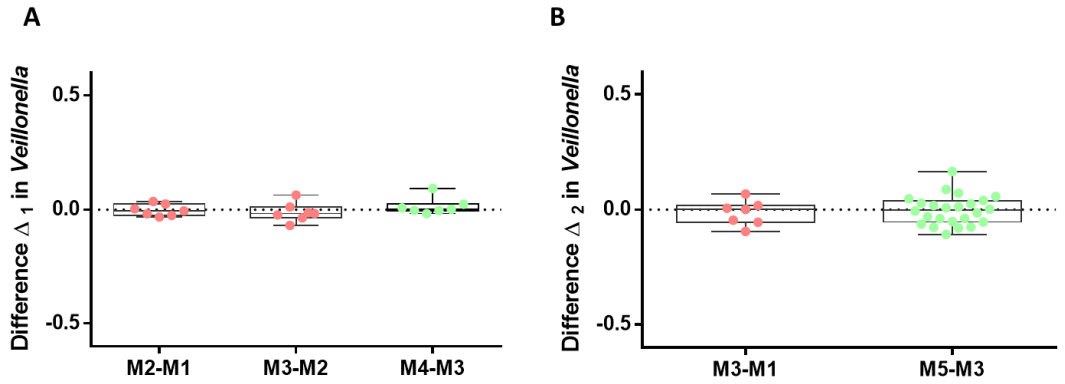

Figure S 9. Differences in relative frequencies of *Veillonella* genus **A)** between successive timepoints, M2-M1 and M3-M2 (steady state), and M4-M3 (transition between steady state and PR) and **B)** between timepoints with two months difference M3-M1 (steady state) and M5-M3 (transition between steady state and PR). No significant differences were observed (Friedman test with Dunn's multiple comparisons and Wilcoxon test, respectively).

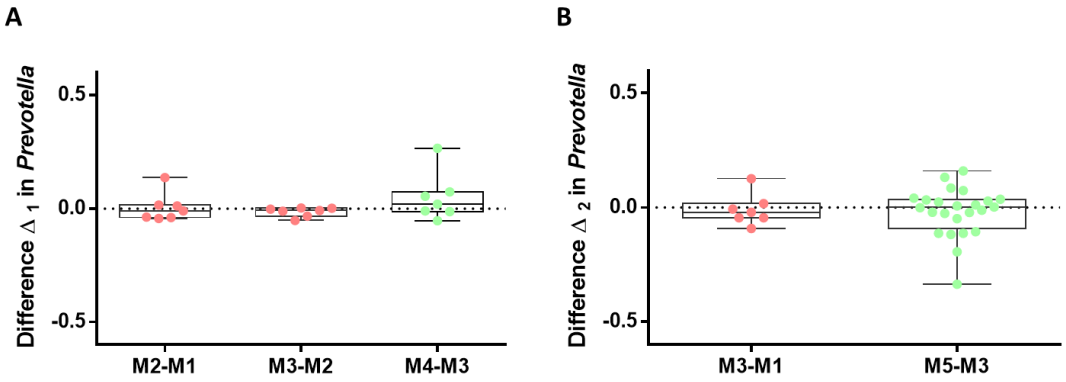

Figure S 10. Differences in relative frequencies of *Prevotella* genus **A)** between successive timepoints, M2-M1 and M3-M2 (steady state), and M4-M3 (transition between steady state and PR) and **B)** between timepoints with two months difference M3-M1 (steady state) and M5-M3 (transition between steady state and PR). No significant differences were observed (Friedman test with Dunn's multiple comparisons and Wilcoxon test, respectively).

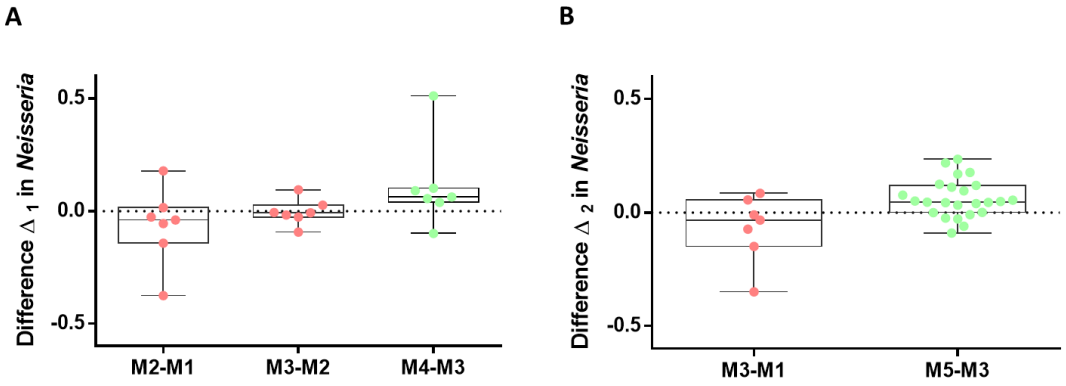

Figure S 11. Differences in relative frequencies of *Neisseria* genus **A)** between successive timepoints, M2-M1 and M3-M2 (steady state), and M4-M3 (transition between steady state and PR) and **B)** between timepoints with two months difference M3-M1 (steady state) and M5-M3 (transition between steady state and PR). No significant differences were observed (Friedman test with Dunn's multiple comparisons and Wilcoxon test, respectively).

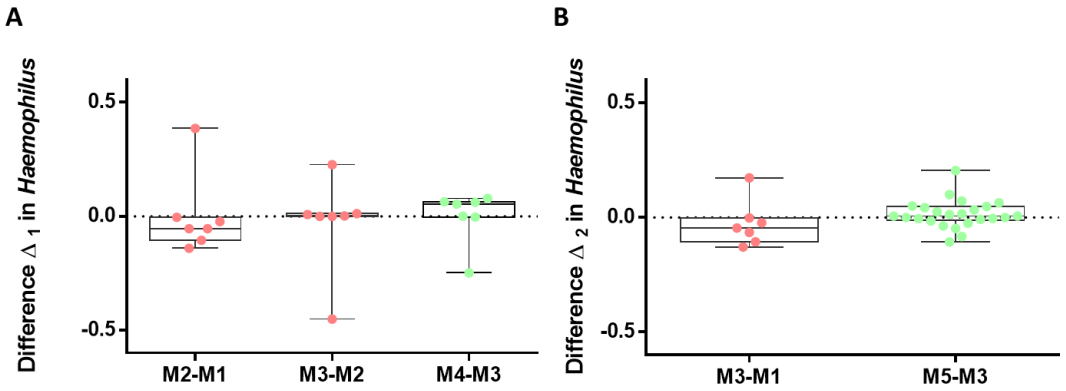

Figure S 12. Differences in relative frequencies of *Haemophilus* genus **A)** between successive timepoints, M2-M1 and M3-M2 (steady state), and M4-M3 (transition between steady state and PR) and **B)** between timepoints with two months difference M3-M1 (steady state) and M5-M3 (transition between steady state and PR). No significant differences were observed (Friedman test with Dunn's multiple comparisons and Wilcoxon test, respectively).

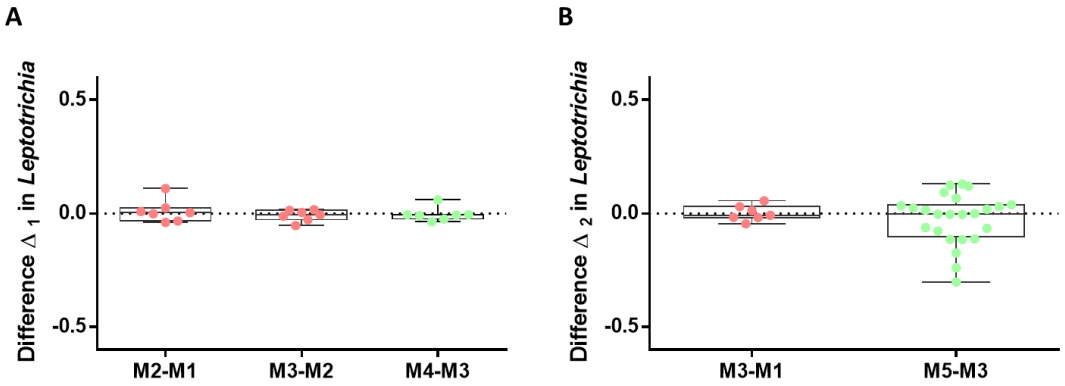

Figure S 13. Differences in relative frequencies of *Leptotrichia* genus **A)** between successive timepoints, M2-M1 and M3-M2 (steady state), and M4-M3 (transition between steady state and PR) and **B)** between timepoints with two months difference M3-M1 (steady state) and M5-M3 (transition between steady state and PR). No significant differences were observed (Friedman test with Dunn's multiple comparisons and Wilcoxon test, respectively).

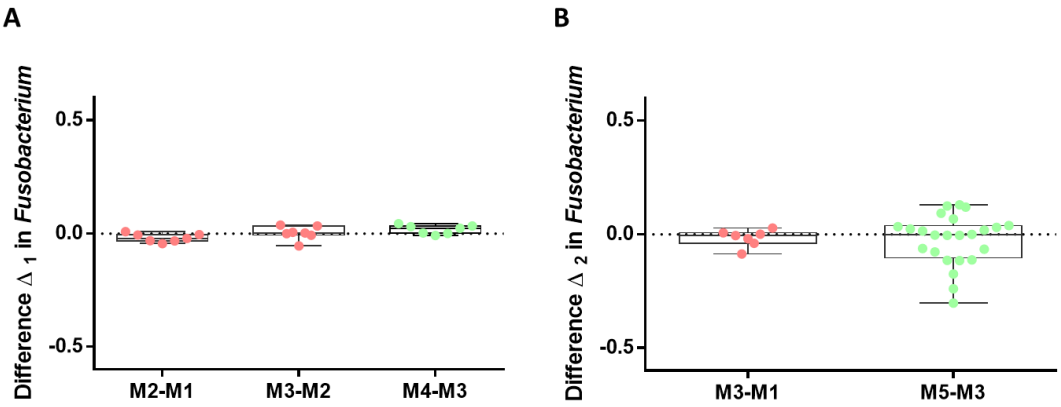

Figure S 14. Differences in relative frequencies of *Fusobacterium* genus **A)** between successive timepoints, M2-M1 and M3-M2 (steady state), and M4-M3 (transition between steady state and PR) and **B)** between timepoints with two months difference M3-M1 (steady state) and M5-M3 (transition between steady state and PR). No significant differences were observed (Friedman test with Dunn's multiple comparisons and Wilcoxon test, respectively).

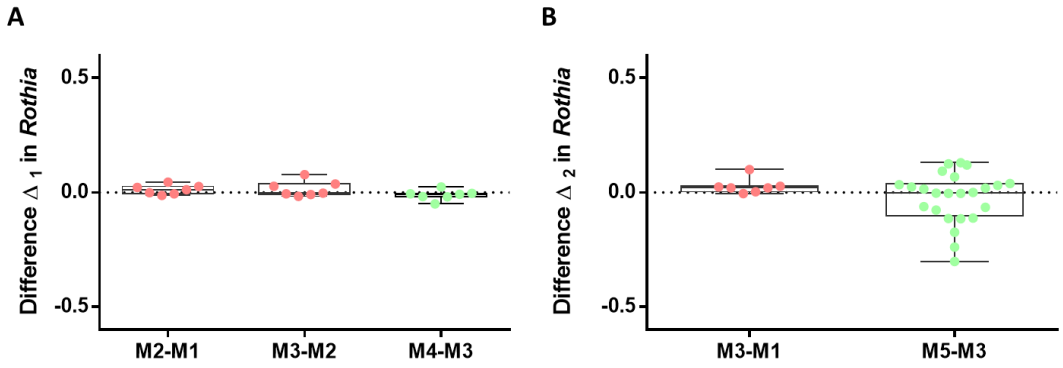

Figure S 15. Differences in relative frequencies of *Rothia* genus **A)** between successive timepoints, M2-M1 and M3-M2 (steady state), and M4-M3 (transition between steady state and PR) and **B)** between timepoints with two months difference M3-M1 (steady state) and M5-M3 (transition between steady state and PR). No significant differences were observed (Friedman test with Dunn's multiple comparisons and Wilcoxon test, respectively).

Appendix V – Alfa diversity differences plots, by period and month, estimated with Shannon diversity index, observed OTUs and evenness, between successive timepoints and between timepoints with two months of difference

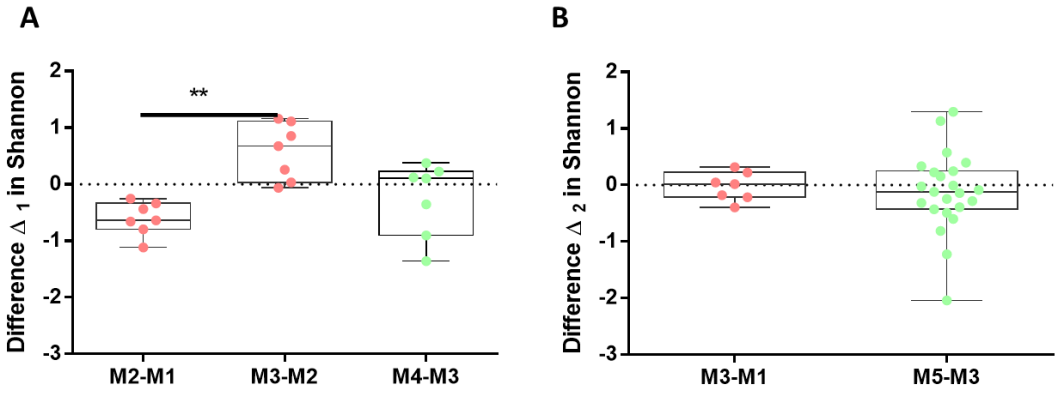

Figure S 16. Differences in alfa diversity (Shannon's diversity index) A) between successive timepoints, M2-M1 and M3-M2 (steady state), and M4-M3 (transition between steady state and PR) and B) between timepoints with two months difference M3-M1 (steady state) and M5-M3 (transition between steady state and PR). A significant difference was observed between M2-M1 and M3-M2 (Friedman test P=0.0084, Dunn's multiple comparison P=0.0099). No differences were observed when we considering two-month interval, Δ_2, (Wilcoxon test). *P<0.05, **P<0.01, ***P<0.001*

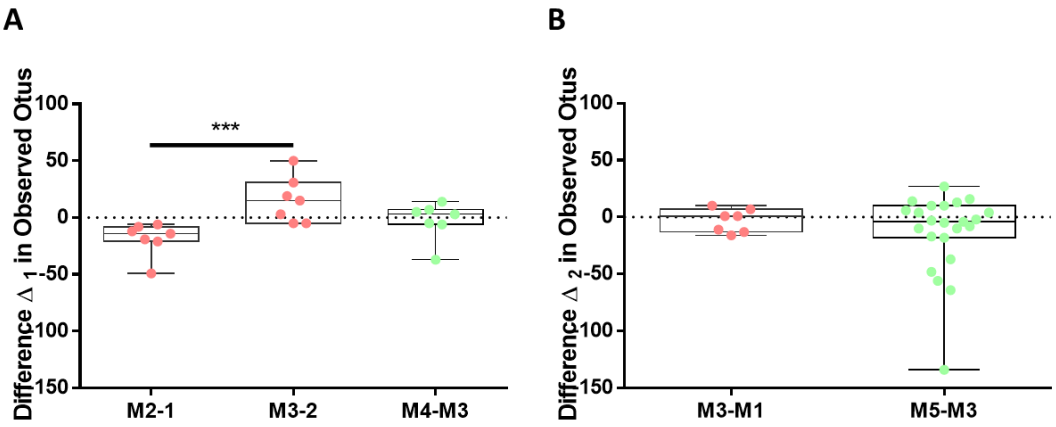

Figure S 17. Differences in alfa diversity (observed OTUs) **A)** between successive timepoints, M2-M1 and M3-M2 (steady state), and M4-M3 (transition between steady state and PR) and **B)** between timepoints with two months difference M3-M1 (steady state) and M5-M3 (transition between steady state and PR). A significant difference was observed between M2-M1 and M3-M2 (Friedman test *P*=0.0003, Dunn's multiple comparison *P*=0.0015). No differences were observed when we considering two-month interval, Δ_2, (Wilcoxon test). **P*<0.05, ***P*<0.01, ****P<0.001*

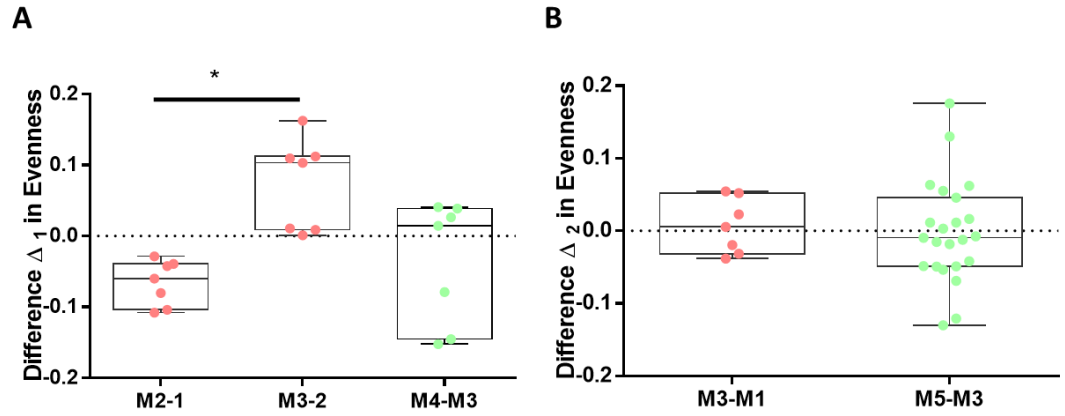

Figure S 18. Differences in alfa diversity (Evenness) **A)** between successive timepoints, M2-M1 and M3-M2 (steady state), and M4-M3 (transition between steady state and PR) and **B)** between timepoints with two months difference M3-M1 (steady state) and M5-M3 (transition between steady state and PR). A significant difference was observed between M2-M1 and M3-M2 (Friedman test *P*=0.0272, Dunn's multiple comparison *P*=0.0227). No differences were observed when we considering two-month interval, Δ_2, (Wilcoxon test). **P*<0.05, ***P*<0.01, ****P<0.001*

Appendix VI – Beta diversity distance plots, by period and month, estimated with unweighted Unifrac distance, Bray Curtis distance and Jaccard distance,

between successive timepoints and between timepoints with two months of difference

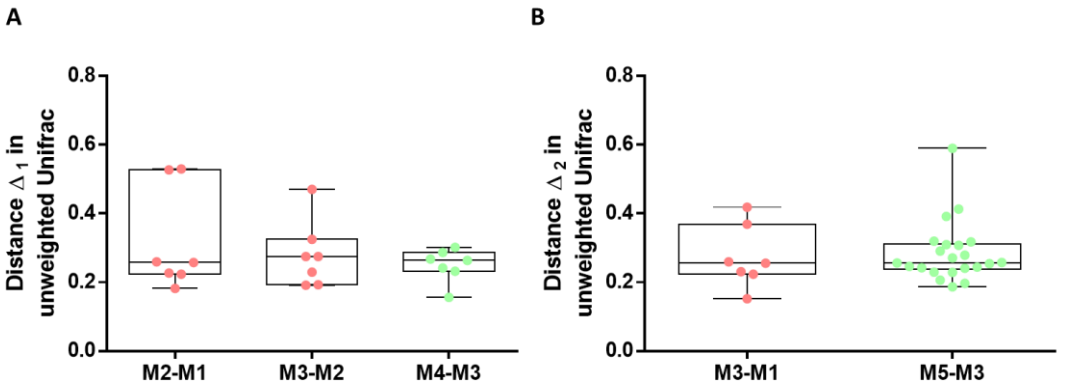

Figure S 19. Pairwise distance comparisons in beta diversity (unweighted Unifrac distance) **A)** between successive timepoints, M2-M1 and M3-M2 (steady state), and M4-M3 (transition between steady state and PR) and **B)** between timepoints with two months difference M3-M1 (steady state) and M5-M3 (transition between steady state and PR). No significant differences were observed (Friedman test with Dunn's multiple comparisons and Wilcoxon test, respectively).

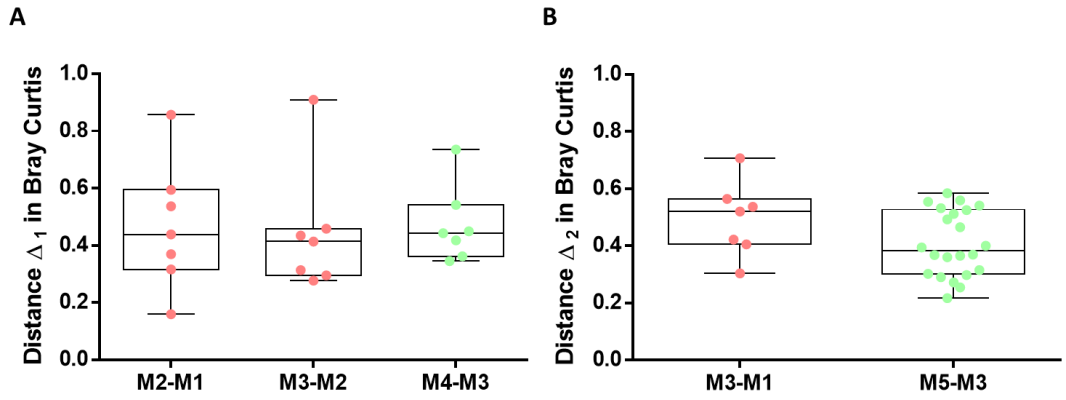

Figure S 20. Pairwise distance comparisons in beta diversity (Bray Curtis distance) **A)** between successive timepoints, M2-M1 and M3-M2 (steady state), and M4-M3 (transition between steady state and PR) and **B)** between timepoints with two months difference M3-M1 (steady state) and M5-M3 (transition between steady state and PR). No significant differences were observed (Friedman test with Dunn's multiple comparisons and Wilcoxon test, respectively).

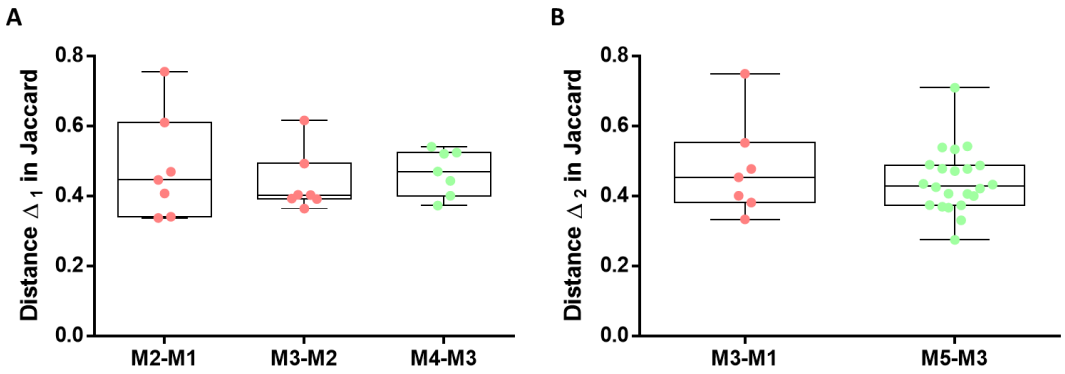

Figure S 21. Pairwise distance comparisons in beta diversity (Jaccard distance) **A)** between successive timepoints, M2-M1 and M3-M2 (steady state), and M4-M3 (transition between steady state and PR) and **B)** between timepoints with two months difference M3-M1 (steady state) and M5-M3 (transition between steady state and PR). No significant differences were observed (Friedman test with Dunn's multiple comparisons and Wilcoxon test, respectively).

Appendix VII – Linear mixed effects models using alfa diversity metrics (Shannon's diversity index, Observed OTUs and evenness)

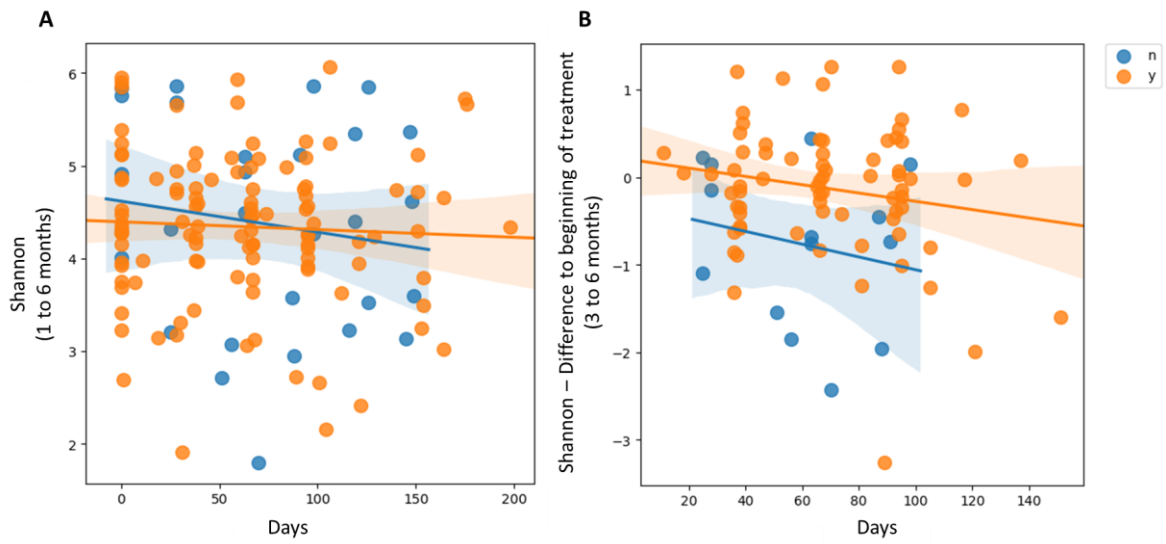

Figure S 22. Alfa diversity regression plots **A)** over time in 1 to 6 months and **B)** difference of each time point to a baseline (3 to 6 month). **A)** Diversity (Shannon index) decreases over time in time in both groups (intervention (y) and control group (n)). However, was observed a tended of the diversity of the intervention group seems to decrease much less over time than the diversity of the control group, but it is not significant difference. **B)** Difference of diversity to baseline (month 3) decreases in both groups over time, but more in control group.

Table S 1. Linear mixed-effects model results for Shannon's Diversity index between 1 and 6 months

Variable or parameter	Coef.	Std.Err.	Z-score	*P* value
Intercept	4.665	0.355	13.158	0.000
Patientexposed[T.y]	-0278	0.389	-0.715	0.475
Days	-0.004	0.002	-2.304	0.021
Days:Patientexposed[T.y]	0.004	0.002	1.774	0.076
Group Var	0.488	0.318		

Parameter estimate (coefficient), standard error, Z score, and *P* value for each model parameter. Brackets indicate reference groups for interpreting fixed-effect estimates.

Table S 2. Linear mixed-effects model results for Shannon difference between samples collected between 4 and 6 months to baseline (month 3)

Variable or parameter	Coef.	Std.Err.	Z-score	*P* value
Intercept	-0.303	0.427	-0.709	0.478
Patientexposed[T.y]	0.077	0.471	0.163	0.870
Days	-0.009	0.005	-1.868	0.062
Days:Patientexposed[T.y]	0.010	0.005	1.978	0.048
Group Var	0.454	0.412		

Parameter estimate (coefficient), standard error, Z score, and *P* value for each model parameter.

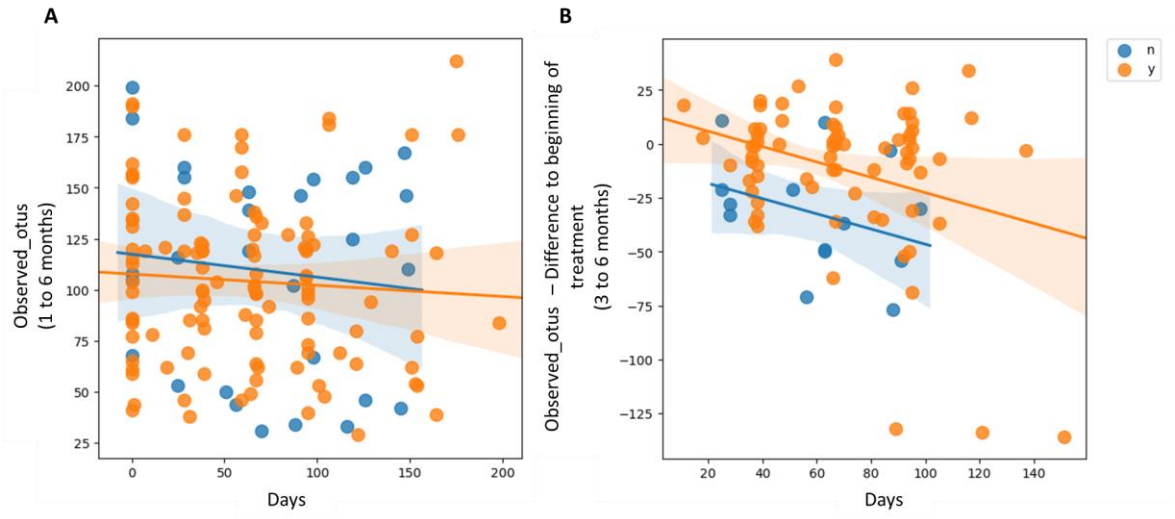

Figure S 23. Alfa diversity regression plots **A)** over time in 1 to 6 months and **B)** difference of each time point to a baseline (3 to 6 month). **A)** Observed OTUs over time in both groups (intervention (y) and control group (n)). **B)** No significant difference was observed.

Table S 3. Linear mixed-effects model results for observed OTUs between 1 and 6 months

Variable or parameter	**Coef.**	**Std.Err.**	**Z-score**	***P* value**
Intercept	119.047	17.569	6.776	0.000
Patientexposed[T.y]	-10.422	19.258	-0.541	0.588
Days	-0.159	0.068	-2.329	0.020
Days:Patientexposed[T.y]	0.078	0.077	1.006	0.314
Group Var	1360.322	23.447		
Parameter estimate (coefficient), standard error, Z score, and *P* value for each model parameter.				

Table S 4. Linear mixed-effects model results for observed OTUs difference between samples collected between 4 and 6 months to baseline (month 3)

Variable or parameter	**Coef.**	**Std.Err.**	**Z-score**	***P* value**
Intercept	-21.822	16.462	-1.326	0.185
Patientexposed[T.y]	14.338	18.126	0.791	0.429

Days	-0.216	0.159	-1.387	0.165
Days:Patientexposed[T.y]	0.159	0.169	0.939	0.348
Group Var	842.111	21.765		
Parameter estimate (coefficient), standard error, Z score, and *P* value for each model parameter.				

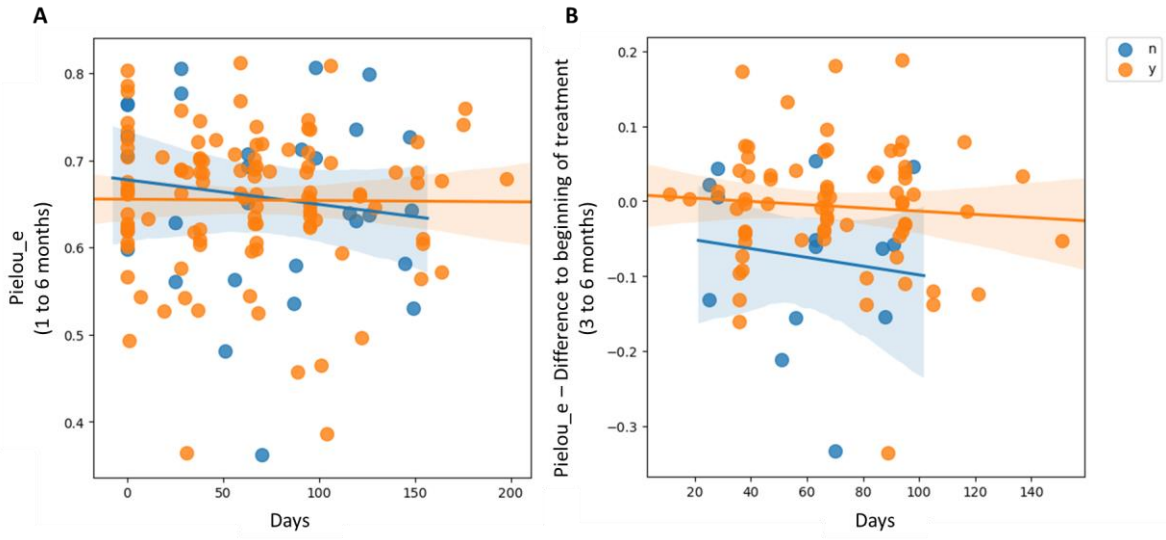

Figure S 24. Alfa diversity regression plots **A)** over time in 1 to 6 months and **B)** difference of each time point to a baseline (3 to 6 month). No significant difference was observed.

Table S 5. Linear mixed-effects model results for evenness, samples collected between 1 and 6 months

Variable or parameter	**Coef.**	**Std.Err.**	**Z-score**	***P* value**
Intercept	0.683	0.033	20.785	0.000
Patientexposed[T.y]	-0.029	0.036	-0.809	0.419
Days	-0.000	0.000	-1.583	0.113
Days:Patientexposed[T.y]	0.000	0.000	1.549	0.121
Group Var	0.003	0.019		
Parameter estimate (coefficient), standard error, Z score, and *P* value for each model parameter.				

Table S 6. Linear mixed-effects model results for evenness difference between samples collected between 4 and 6 months to baseline (month 3)

Variable or parameter	Coef.	Std.Err.	Z-score	*P* value
Intercept	-0.018	0.051	-0.348	0.728
Patientexposed[T.y]	-0.007	0.057	-0.128	0.898
Days	-0.001	0.001	-1.608	0.108
Days:Patientexposed[T.y]	0.001	0.001	1.853	0.064
Group Var	0.005	0.035		
Parameter estimate (coefficient), standard error, Z score, and *P* value for each model parameter.				

Appendix VIII – Linear mixed effects models using beta diversity distances (unweighted Unifrac distance, Bray Curtis Distance and Jaccard distance)

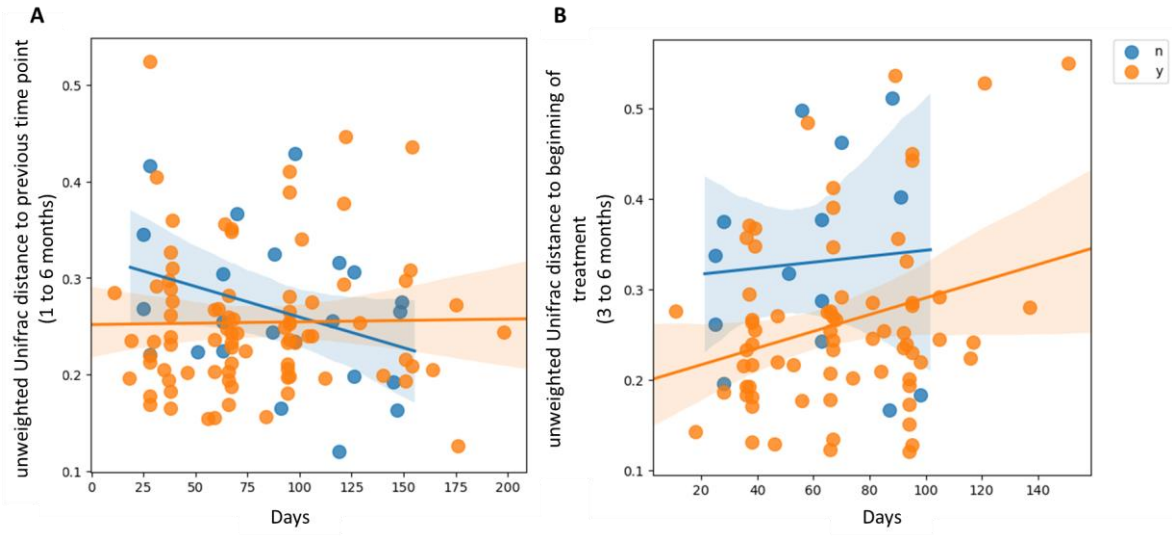

Figure S 25. Beta diversity (unweighted Unifrac distance) regression plots of **A)** distances to a previous time point (1 to 6 months) and **B)** distance of each time point to a baseline (3 to 6 month). No significant difference was observed

Table S 7. Linear mixed-effects model results for unweighted Unifrac distance between successive samples collected between 1 and 6 months

Variable or parameter	Coef.	Std.Err.	Z-score	*P* value
Intercept	0.315	0.037	8.595	0.000
Patientexposed[T.y]	-0.061	0.040	-1.508	0.131
Days	-0.001	0.000	-1.793	0.073
Days:Patientexposed[T.y]	0.001	0.000	1.591	0.112
Group Var	0.002	0.016		
Parameter estimate (coefficient), standard error, Z score, and *P* value for each model parameter.				

Table S 8. Linear mixed-effects model results for unweighted Unifrac distance between samples collected between 4 and 6 months to baseline (month 3)

Variable or parameter	Coef.	Std.Err.	Z-score	*P* value
Intercept	0.332	0.053	6.215	0.000
Patientexposed[T.y]	-0.093	0.059	-1.575	0.115
Days	0.000	0.001	0.227	0.821
Days:Patientexposed[T.y]	0.000	0.001	0.355	0.723
Group Var	0.08	0.058		
Parameter estimate (coefficient), standard error, Z score, and *P* value for each model parameter.				

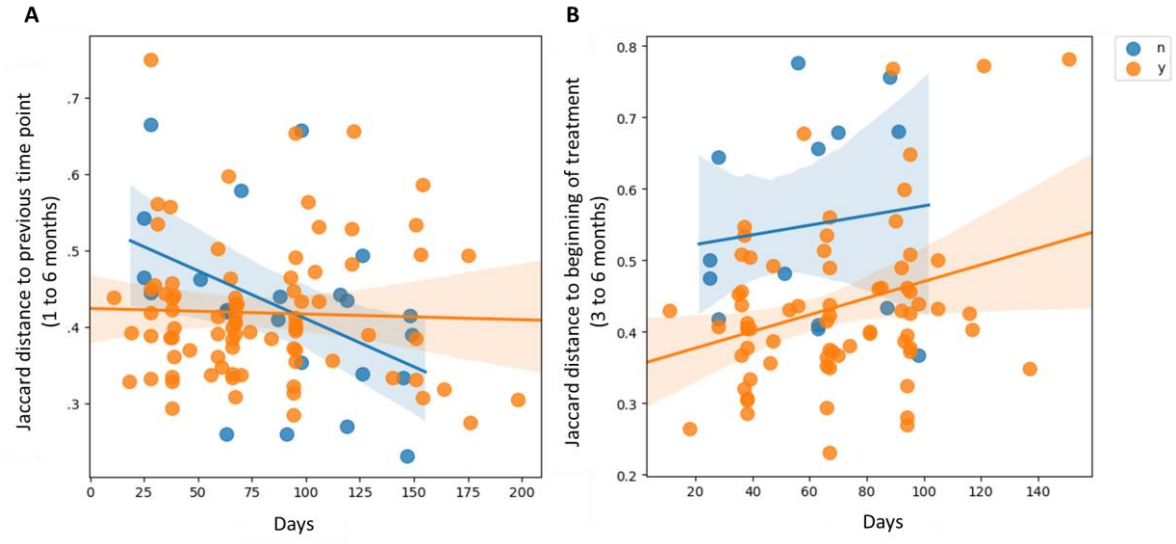

Figure S 26. Beta diversity (Jaccard distance) regression plots of **A)** distances to a previous time point (1 to 6 months) and **B)** distance of each time point to a baseline (3 to 6 month). **A)** Jaccard distance to a previous time point decrease in control group (n) and seems to keep constant in intervention group (y) over time. **B)** Jaccard distance to baseline (month 3) is higher in control group than in intervention group.

Table S 9. Linear mixed-effects model results for Jaccard distance between successive samples collected between 1 and 6 months

Variable or parameter	Coef.	Std.Err.	Z-score	*P* value
Intercept	0.522	0.046	11.448	0.000
Patientexposed[T.y]	-0.093	0.050	-1.861	0.063
Days	-0.001	0.000	-3.085	0.002
Days:Patientexposed[T.y]	0.001	0.000	2.283	0.022
Group Var	0.004	0.025		
Parameter estimate (coefficient), standard error, Z score, and *P* value for each model parameter.				

Table S 10. Linear mixed-effects model results for Jaccard distance between samples collected between 4 and 6 months to baseline (month 3)

Variable or parameter	Coef.	Std.Err.	Z-score	*P* value

Intercept	0.548	0.060	9.186	0.000
Patientexposed[T.y]	-0.153	0.066	-2.327	0.020
Days	0.000	0.001	0.409	0.683
Days:Patientexposed[T.y]	0.000	0.001	0.584	0.559
Group Var	0.011	0.078		

Parameter estimate (coefficient), standard error, Z score, and *P* value for each model parameter.

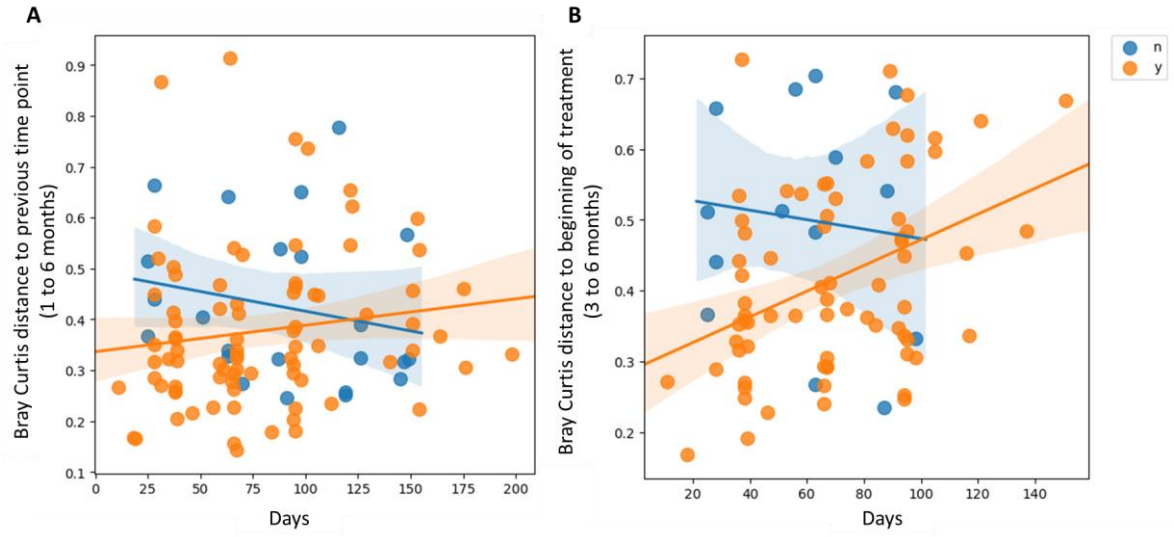

Figure S 27. Beta diversity regression plots of **A)** distances to a previous time point (1 to 6 months) and **B)** distance of each time point to a baseline (3 to 6 month). **A)** No significative differences was observed. **B)** Bray Curtis distance to baseline (month 3) increases in intervention group and decrease in control group over time.

Table S 11. Linear mixed-effects model results for Bray Curtis distance between successive samples collected between 1 and 6 months

Variable or parameter	**Coef.**	**Std.Err.**	**Z-score**	***P* value**
Intercept	0.520	0.076	6.865	0.000
Patientexposed[T.y]	-0.162	0.083	-1.952	0.051
Days	-0.001	0.001	-1.574	0.115
Days:Patientexposed[T.y]	0.001	0.001	1.596	0.111

Group Var	0.009	0.033		
Parameter estimate (coefficient), standard error, Z score, and *P* value for each model parameter.				

Table S 12. Linear mixed-effects model results for Bray Curtis distance between samples collected between 4 and 6 months to baseline (month 3)

Variable or parameter	**Coef.**	**Std.Err.**	**Z-score**	***P* value**
Intercept	0.562	0.074	7.643	0.000
Patientexposed[T.y]	-0.229	0.081	-2.817	0.005
Days	-0.001	0.001	-1.068	0.285
Days:Patientexposed[T.y]	0.002	0.001	2.276	0.023
Group Var	0.011	0.057		
Parameter estimate (coefficient), standard error, Z score, and *P* value for each model parameter.				

Appendix IX – PCoA plots between Pre-PR and PR period using unweighted Unifrac distance matrix, Bray Curtis distance matrix and Jaccard distance matrix

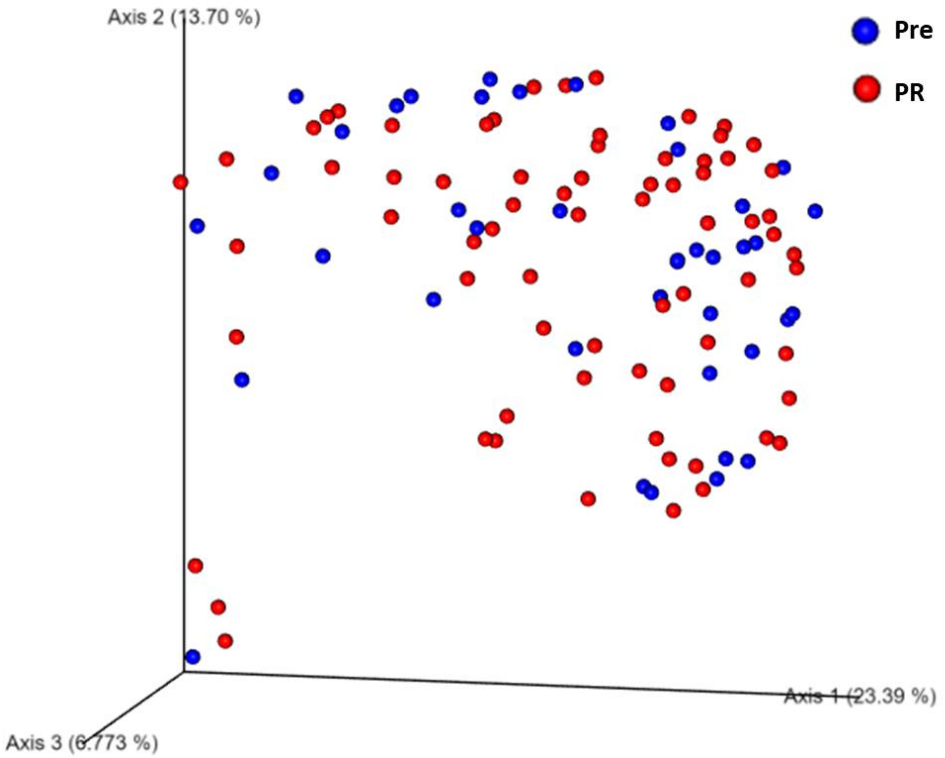

Figure S 28. Microbiota of the different patients doesn't converge to a similar composition during Pulmonary Rehabilitation. Saliva's microbiota composition of samples of Pre-PR period isn't significantly different from the saliva's microbiota composition of samples of PR period (PCoA analysis using Emperor of unweighted UniFrac distance matrix, PERMANOVA, *P*=0.593).

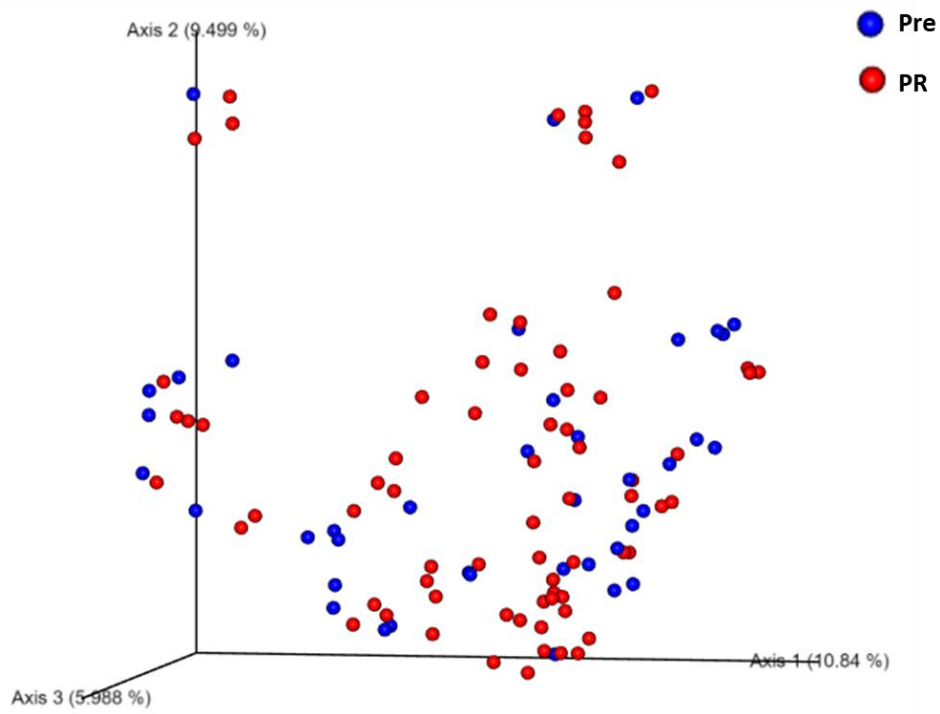

Figure S 29. Microbiota of the different patients doesn't converge to a similar composition during Pulmonary Rehabilitation. Saliva's microbiota composition of samples of Pre-PR period isn't significantly different from the saliva's microbiota composition of samples of PR period (PCoA analysis using Emperor of Jaccard distance matrix, PERMANOVA, *P*=0.919).

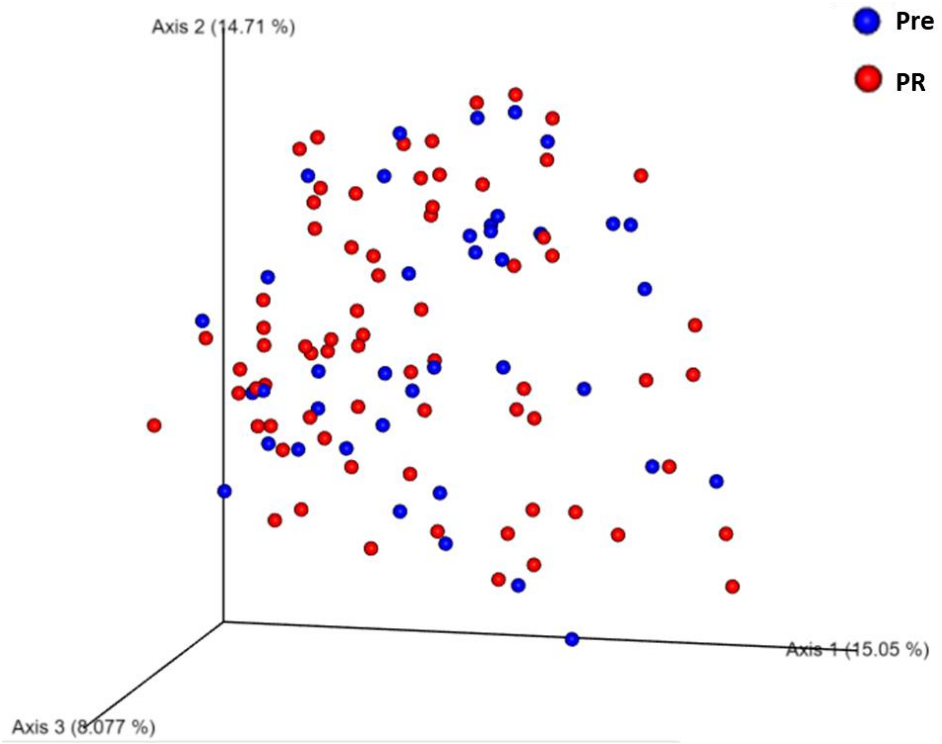

Figure S 30. Microbiota of the different patients doesn't converge to a similar composition during Pulmonary Rehabilitation. Saliva's microbiota composition of samples of Pre-PR period isn't significantly different from the saliva's microbiota composition of samples of PR period (PCoA analysis using Emperor of Bray Curtis distance matrix, PERMANOVA, *P*=0.312).

Appendix X – PCoA plots between PR and Post-PR period using unweighted Unifrac distance matrix, Bray Curtis distance matrix and Jaccard distance matrix

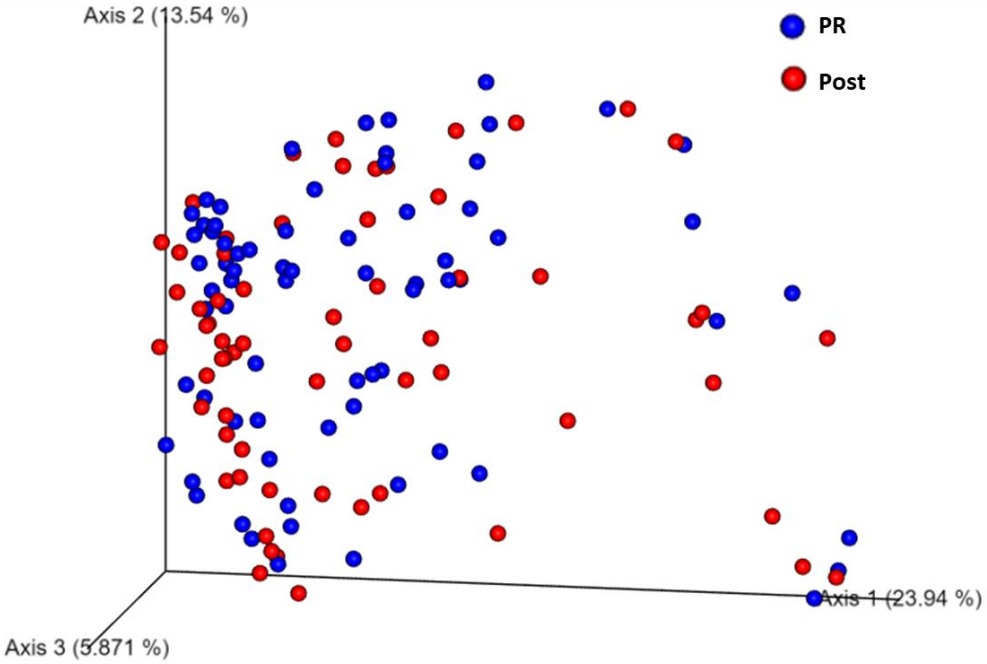

Figure S 31. Microbiota of the different patients doesn't converge to a similar composition after Pulmonary Rehabilitation. Saliva's microbiota composition of samples of PR period isn't significantly different from the saliva's microbiota composition of samples of Post-PR period (PCoA analysis using Emperor of unweighted UniFrac distance matrix, PERMANOVA, *P*=0.743).

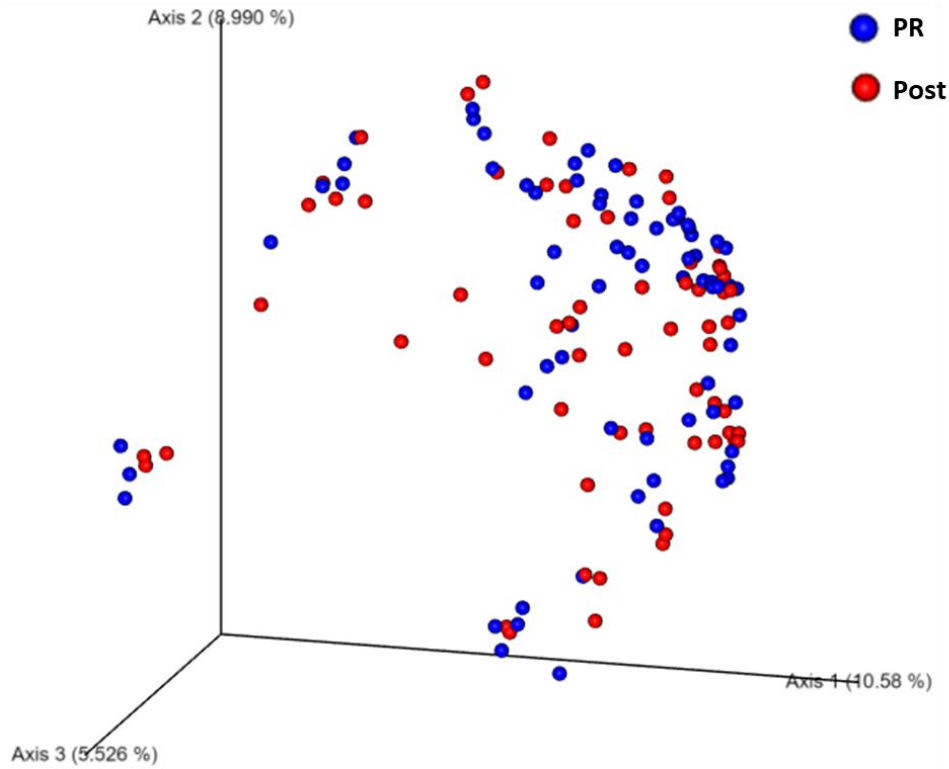

Figure S 32. Microbiota of the different patients doesn't converge to a similar composition after Pulmonary Rehabilitation. Saliva's microbiota composition of samples of PR period isn't significantly different from the saliva's microbiota composition of samples of Post-PR period (PCoA analysis using Emperor of Jaccard distance matrix, PERMANOVA, *P*=0.957).

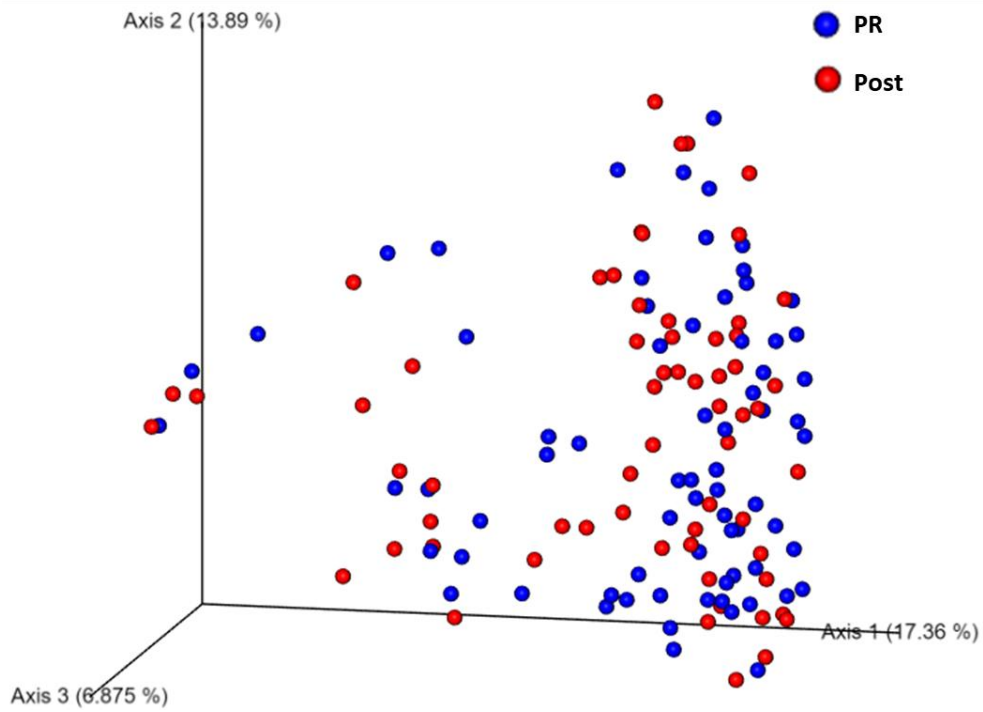

Figure S 33. Microbiota of the different patients doesn't converge to a similar composition after Pulmonary Rehabilitation. Saliva's microbiota composition of samples of PR period isn't significantly different from the saliva's microbiota composition of samples of Post-PR period (PCoA analysis using Emperor of Bray Curtis distance matrix, PERMANOVA, *P*=0.587).