The lower airway microbiota and inflammation in chronic obstructive pulmonary disease (COPD)

Thoughts on where to measure it, how to interpret it, and why it might matter

Solveig Tangedal

Thesis for the degree of Philosophiae Doctor (PhD)
University of Bergen, Norway
2024

UNIVERSITY OF BERGEN

The lower airway microbiota and inflammation in chronic obstructive pulmonary disease (COPD)

Thoughts on where to measure it, how to interpret it, and why it might matter

Solveig Tangedal

Thesis for the degree of Philosophiae Doctor (PhD)
at the University of Bergen

Date of defense: 01.02.2024

Year: 2024

Title: The lower airway microbiota and inflammation in chronic obstructive pulmonary disease (COPD)

Name: Solveig Tangedal

Print: Skipnes Kommunikasjon / University of Bergen

Table of contents

1. Scientific environment

The thesis presented here is based on data collected from three large chronic obstructive pulmonary disease (COPD) cohort studies. The studies were conducted at the Department of Thoracic Medicine at Haukeland University Hospital by the Bergen Respiratory Research Group. The research group is currently led by Professor Tomas Mikal Lind Eagan. The three studies include the *Bergen COPD Cohort Study (BCCS),* the *Bergen COPD Exacerbation Study (BCES),* and the *Bergen COPD microbiome study (MicroCOPD)*.

My research started in January 2013, when I was employed for a 50% clinical research position at the Department of Thoracic Medicine at Haukeland University Hospital. I have been affiliated with the Department of Clinical Science, Faculty of Medicine, University of Bergen since May 2017. The PhD work has since then been funded by the University of Bergen.

My main supervisor has been professor, MD Tomas Mikal Lind Eagan. Co-supervisors have been associate professor, MD Rune Nielsen and PhD, MD Marianne Aanerud. All supervisors have been affiliated with the Department of Thoracic Medicine and the Department of Clinical Science.

2. Acknowledgements

Where better to start, than at the beginning of things: Soon after signing my junior doctor employment contract, the ward director Kahtan Al-Azawy asked if I would be interested in conducting research as part of my employment. And he let it be clear: Saying yes would be a smart career move. He directed me to Professor Eagan´s office where I was welcomed into the research group, and here we are a decade later. So, thank you Al-Azawy!

From the very beginning I have wondered where my three supervisors hide their shared time machine. Anyone working with them will realise that they either have access to twice as many hours in a day as everyone else, or maybe they are just superhuman:

Marianne, as my co-supervisor you proved to be a true wizard both in Stata and with regards to your teaching skills. You offered invaluable help with statistics and introduced me to a new world of data handling. Not one time was I met with reluctancy when asking for advice. You are such an enthusiastic person making even the most tedious Stata workflow fun to untangle! And you worked days and during nights to help me while juggling your clinical work, union representation, raising kids, feeding your neighbours, colleagues, and fellow researchers, and succeeding at a repertoire of hobbies others can only admire!

Rune, you have been an amazingly valuable sparring partner. You have continuously been challenging established truths within this field of research, pushing for further work, new methods, critical analyses, and a deeper understanding of the matter. Without you I would not have known half of what I do today. The long talks in our office, during travels to conferences and courses has been enlightening and amusing, resulting in a will to keep on when challenging methods and results have put a dampener on my own enthusiasm. And amid everything you do for your PhD candidates, you have taken upon you commitments benefiting the entirety of Norwegian respiratory medicine doctors, and you somehow manage to prioritise time for, not to mention expand, your own wonderful family.

Tomas, I could not have gotten a better supervisor than you. From that meeting in your office back in 2012 you have offered me more of your time than many would ever have considered. And I have helped myself to your time shamelessly! As an example, my gmail account doesn't even quantify the e-mail correspondence between us by numbers, it just states *there are many*. Not only that, but you have kept several channels open for communication, so I am not going to try counting the number of messenger conversations, SMSs or WhatsApp messages sent and received. You have gracefully balanced the need to teach and instruct me, with the need to let me work independently. You challenge my ideas and thoughts, being just as eager with the red marker as my mother was correcting my homework back in the 1990s. You are an endless source of good advice and knowledge, putting things straight when confusion and despair is about to topple my days. Through the happiest and the darkest times, you have stepped in to avoid full stop in my research, without ever putting on any pressure. And during those rougher times you have been such a valuable support, sharing of your own experiences, giving advice reaching far passed the research you have been supervising. I wish to thank you from the bottom of my heart for the way you have helped me and three generations of my family when life was at its most chaotic. Without your support, I could never have completed this project.

My dear colleagues and friends, Marta, Christina, Bahareh and Anders: What should I have done without you? You bring me joy coming to work both at the hospital and doing my research. I love how we have progressed from being colleagues to being friends also outside the workplace. You keep me on the right path, help me remember all the scheduled stuff that never caught my attention, but that should have. I cherish having you to lean on, and for you to be able to lean on me. Life is full of joys and sorrows, and they are both best delt with in company of good friends like you.

Thanks to the Department of Thoracic Medicine and the entire Bergen Respiratory Research Group present and past, for providing with a great work environment, and the opportunity to become a microbiota nerd. A special thanks to Sverre Lehmann, Kristel Knudsen and Øystein Fløtten who have shown great flexibility with my clinical duties, allowing my participations at congresses and courses the past decade.

Thanks to the Department of Clinical Science for allowing me to combine research and clinical work. Thank you for offering the opportunity to be trained in my field outside the University of Bergen, but also offering interesting research courses broadening my perspectives and allowing me to meet other PhD candidates.

Thanks to co-authors Marianne Aanerud, Rune Nielsen, Tomas M. L. Eagan, Louise J. P. Persson, Karl A. Brokstad, Per S. Bakke, Christine Drengenes, Harald G. Wiker, Pieter S. Hiemstra, Gunnar R. Husebø, Sverre Lehmann and Kristel S. Knudsen both for insightful feedback and support.

Thanks to everyone that contributed to data collection and processing. In random order: Margrete Klemmetsby, Marie Waatevik, Dag Harald Skutlaberg, Lene Svendsen, Tina Endresen-Vinsjevik, Eli Nordeide, Inge Zwetzig, Hilda Andersen, Tuyen Hoang, Lise Ø Monsen, Hildegunn Fleten, Ingvild Haaland, Øistein Svanes, Marit Aardal, Randi Sandvik, Tharmini Kalanthanan. Without you my thesis would not have been realised. Some of you have also supported me with feedback on manuscripts, conference papers and presentations, you have taken part in group discussions and social events, or all mentioned above.

A special thanks to Heidi E. Haraldsen and Åse Fredriksen for patiently introducing me to and teaching me how to perform the laboratory work I needed to do for my research.

Thanks to all the wonderful study participants for sacrificing time and bodily fluids to make this research possible.

At last, but not the least – my family: A great thank you to my loved ones at home, Kristopher, Frida and Hedda for all your patience and support. Kristopher, thank you for dealing with everything at home while I travelled in parts to get educated, in parts to enjoy the opportunities offered. Thank you for listening to me ramble on about unsolvable challenges and statistical shortcomings. Thank you, my dear kids, for bearing over with the mother hidden behind the computer screen absentmindedly answering your questions. Though I´m sure you´ve gotten some ice creams out of it!

Thank you to my devoted parents Olav and Åshild for always supporting me through my education and studies, and through all the challenges life has thrown at me over the years. How can people manage without grandparents like you? The same goes without saying for my mother-in-law Unni. And thank you Målfrid for being the very best sister in this world! Who and where would I have been without you?

Many more could and should have been mentioned, but the professor has an unwavering demand for this thesis to be as short as possible so I´ll stop here.

3. Terms and abbreviations

16S rRNA	16S ribosomal ribonucleic acid
AMPs	Antimicrobial peptides
ANC	Absolute neutrophil counts
ANCOM-BC	Analysis of Compositions of Microbiomes with Bias Correction
ANOSIM	Analysis of similarities
ANOVA	Analysis of variance
ASVs	Amplicon sequence variants
(P)BAL	(Protected) bronchoalveolar lavage
BCCS	The Bergen COPD cohort study
BCES	The Bergen COPD exacerbation study
BLAST	Basic Local Alignment Search Tool
CAT	COPD Assessment Test
CF	Cystic fibrosis
COPD	Chronic obstructive pulmonary disease
CXCL8	C-X-C motif chemokine ligand 8
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
ELISA	Enzyme-linked immunosorbent assay
FEV_1	Forced expiratory volume at 1 seconds
FVC	Forced vital capacity
GOLD	The Global Initiative for Chronic Obstructive Lung Disease
HMP	The Human Microbiome Project
HOMD	The Human Oral Microbiome Database
ICS	Inhaled corticosteroids
IL-(#)	Interleukin-(#)
IP-10	Interferon gamma-induced protein 10
ISS	Induced sputum samples
KOLS	Kronisk obstruktiv lungesjukdom
LOA	Limits of agreement
LPK	Leucocyte particle count
MAFFT	Multiple Alignment using Fast Fourier Transform
MicroCOPD	The Bergen COPD microbiome study
MIG	Monokine induced by gamma interferon
NAST	Nearest Alignment Space Termination
NIH	National Institute of Health
NMDS	Non-metric multidimensional scaling
OTUs	Operational taxonomic units
OW	Oral wash
PBS	Phosphate-buffered saline
PCoA	Principal coordinate analysis
PCR	Polymerase chain reaction
PERMANOVA	Permutational multivariate analysis of variance

PSB	Protected specimen brushes
pyNAST	Python implementation of the NAST algorithm
QIIME (2)	Quantitative Insights Into Microbial Ecology
RCTs	Randomised controlled trials
RESCRIPt	Reference sequence annotation and curation pipeline
RNA	Ribosomal nucleic acid
SLPI	Secretory leucocyte protease inhibitor
SSS	Spontaneous sputum samples
SVL	Small volume lavage
TNF-α	Tumour necrosis factor alpha
UWUF	Unweighted UniFrac
WBC	White blood cell
WUF	Weighted UniFrac

4. List of Publications

Paper I: Tangedal S, Aanerud M, Persson LJ, Brokstad KA, Bakke PS, Eagan TM (2014): Comparison of inflammatory markers in induced and spontaneous sputum in a cohort of COPD patients. Respir Res. 15(1):138.

Paper II: Tangedal S, Aanerud M, Gronseth R, Drengenes C, Wiker HG, Bakke PS, Eagan TM (2017): Comparing microbiota profiles in induced and spontaneous sputum samples in COPD patients. Respir Res. 18(1):164.

Paper III: Tangedal S, Nielsen R, Aanerud M, Persson LJ, Wiker HG, Bakke PS, Hiemstra PS, Eagan TM (2019) Sputum microbiota and inflammation at stable state and during exacerbations in a cohort of chronic obstructive pulmonary disease (COPD) patients. PLoS One.14(9):e0222449.

Paper IV: Tangedal S, Nielsen R, Aanerud M, Drengenes C, Husebø G, Lehmann S, Knudsen K, Hiemstra PS, Eagan TM (2023) The lower airway microbiota in COPD and healthy controls. Under peer-review.

5. Summaries

5.1 English

Background: Chronic obstructive pulmonary disease (COPD) is a complex inflammatory disease causing the death of millions annually. The lower airway bacterial community (microbiota) and immune responses could be important for the pathogenesis of COPD. The aims for this thesis were to study COPD cohorts considering if measures of inflammatory markers and microbiota 1) are affected by sputum sampling techniques, 2) differ with COPD state, and 3) if the microbiota differ in bronchoalveolar lavage (BAL) comparing patients with COPD with controls.

Methods: Sputum data originated from the *Bergen COPD Cohort* and exacerbations studies in which 433 patients with COPD were enrolled and 356 followed for exacerbations. BAL data originated from the *MicroCOPD* study in which 130 patients with COPD and 103 controls were enrolled. Inflammatory markers in sputum were measured by a bead based multiplex immunoassay and antimicrobial peptides by enzyme linked immunosorbent assay. DNA sequences were obtained by enzymatic and mechanical lysis extraction methods, PCR-amplification of the 16S rRNA gene and paired-end sequencing using the Illumina MiSeq System. Data were analysed in QIIME 1&2, Stata, and R.

Results: Inflammatory markers and microbiota differed significantly between induced and spontaneous sputum, and between stable state COPD and exacerbations. Differences related to disease state showed great heterogeneity looking at individual participants. The microbiota in BAL sampled in the COPD cohort had lower evenness and higher abundances of *Firmicutes* compared with controls. Sex, age, smoking, disease severity and use of inhaled corticosteroids were not clearly associated with the lower airway microbiota.

Conclusion: Sputum sampling methods influences on measurements of inflammation and microbiota. Exacerbations in COPD and the presence of disease are both associated with microbiota dysbiosis which indicate importance of the lower airway microbiota in the pathogenesis in COPD.

5.2 Norsk

Bakgrunn: Kronisk obstruktiv lungesjukdom (KOLS) er ein kompleks, inflammatorisk sjukdom som forårsakar millionar av dødsfall årleg. Bakteriane i dei nedre luftvegane (mikrobiotaet) og immunresponsar kan spele ei viktig rolle i KOLS. Målingar av dei begge kan tenkast å vera påverka av prøvetakingsmetode. Måla med denne oppgåva var å studere KOLS-kohortar med tanke på 1) høve for vekselbruk av indusert og spontant sputum for målingar av inflammasjonsmarkørar og mikrobiota, 2) endringar i inflammasjonsmarkørar og mikrobiota under sjukdomsforverringar samanlikna med stabil sjukdomsfase og 3) forskjellar i mikrobiota i bronkialskyllevæske (BAL) frå pasientar med KOLS samanlikna med friske kontrollar.

Metode: Sputum data kjem frå studiane *BergenKOLS* og tilhøyrande eksaserbasjonsstudie med 433 pasientar med KOLS inkludert, og 356 som vart følgt med tanke på forverringar. BAL data kjem frå studien *MikroKOLS* med 130 pasientar med KOLS og 103 friske kontrollar inkludert. Inflammasjonsmarkørar i sputum vart målt med bead based multiplex immunoassay og antimikrobielle peptid med enzyme linked immunosorbent assay. DNA sekvensar vart reinska ved hjelp av både enzym og mekanisk lysering. PCR-amplifisering av 16S rRNA og paired-end sequencing med Illumina MiSeq System vart utført. Data vart analysert i QIIME 1&2, Stata og R.

Resultat: Inflammasjonsmarkørar og mikrobiota var signifikant forskjellige i indusert og spontant sputum, og i stabil fase av KOLS samanlikna med under pågåande forverring. I høve til sjukdomsfase var ulikskapane heterogene når ein såg på kvart individ. Mikrobiota i BAL var meir ujamn og rikare på *Firmicutes* hos pasientar med KOLS samanlikna med friske. Kjønn, alder, røyking, sjukdomsgrad og bruk av inhalasjons-kortikosteroider var ikkje tydeleg assosiert til mikrobiotaet i dei nedre luftvegane.

Konklusjon: Sputumprøvetaking påverkar målingar av inflammasjonsmarkørar og mikrobiota. KOLS forverringar og KOLS i seg sjølv er begge assosiert med endringar i luftvegsmikrobiotaet. Det tyder på at mikrobiotaet spelar ei rolle i KOLS.

6. Introduction

6.1 Chronic obstructive pulmonary disease (COPD)

COPD is a multifactorial disease characterised by lower airway inflammation (1). Smoking is a trigger of inflammatory processes in the airways and considered the main cause of COPD. Examples of less prevalent causes include air pollution, exposure to other lung toxic substances, and rare genetic disease (2, 3). While COPD is mainly caused by smoking, far from all smokers develop COPD. This observation is yet to be explained (4, 5).

COPD is defined by chronic airflow obstruction assessed with spirometry (6). With a spirometer we measure how much air can be forcefully exhaled in the first second of expiration (FEV_1) and the total amount of air that can be forcefully exhaled (FVC). If the FEV_1/FVC ratio is less than 70% it indicates obstructed airways. In the case of COPD, this obstruction is caused by chronic remodelling of the airways and lung tissue (Illustration 1).

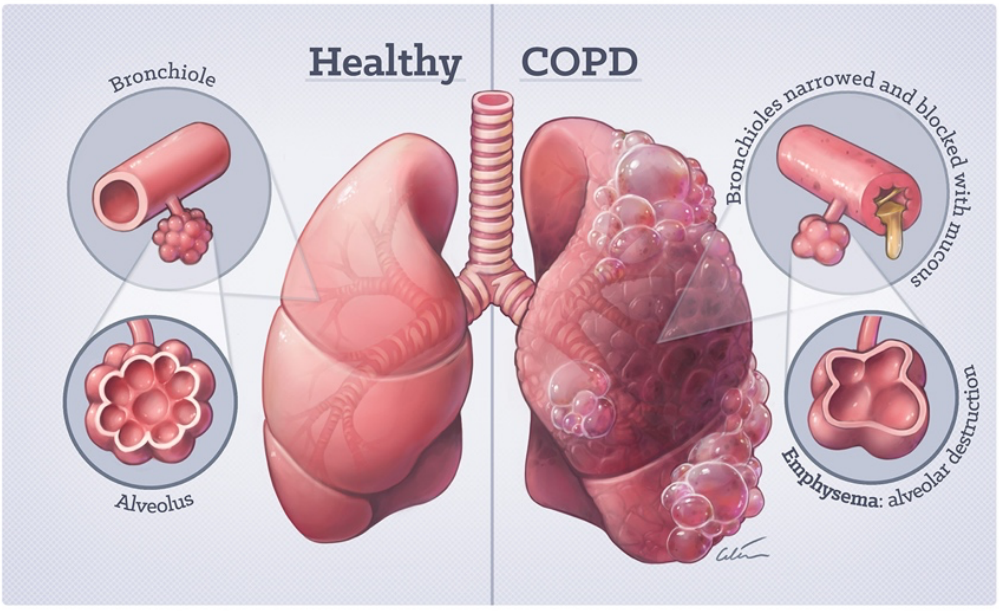

Illustration 1: Medical illustration of remodelling of the lower airways and lung parenchyma in COPD by Dr Ciléin Kearns (Artibiotics). Reprinted with permission.

Narrowing of bronchiole lumens are due to smooth muscle cell hypertrophy and alterations of mucus producing Goblet cells. Airflow obstruction is further caused by destruction of alveoli and a loss of structural integrity allowing for dynamic collapse of the small airways. The resulting emphysematous lung tissue also has a reduced surface for gas exchange compared with healthy lung tissue.

COPD development is progressive, heterogeneous, and unpredictable. While some patients may exhibit minimal symptoms, others can develop a debilitating condition marked by recurring episodes of increased difficulty in breathing, cough, and sputum production. These episodes are referred to as exacerbations, which have been associated with a decline in the overall quality of life and life expectancy, as well as an increase in disability and the need for hospitalisation (6). Exacerbations are often caused by infections, which again activate both the local and systemic immune system. Consequently, a combination of oral corticosteroids to decrease inflammation and antibiotics to target bacterial infections is frequently used in the treatment of exacerbations (6, 7).

Medicines used to treat COPD are otherwise aimed at dilating the bronchioles by relaxing the smooth muscles in the airways, while corticosteroids are recommended only for a subset of patients at stable state (6). These bronchodilators and corticosteroids are administered by various means of inhalation, but neither can reverse the airway remodelling nor loss of alveoli once established. A spirometry performed after inhalation of bronchodilators will therefor remain obstructive in COPD, and this is routinely done in clinical settings.

The use of antibiotics aimed at bacteria colonising the lower airways of patients with COPD has been investigated. Unfortunately, prophylactic antibiotics come with adverse side effects, a risk of inducing antibiotic resistance, and little if any benefit for the treated. It is therefore not routinely recommended (6, 8).

The key to prevent most cases of COPD is to keep people from smoking. Sadly, it remains a challenge to help smokers quit, and to effectively manage patients already diagnosed with COPD. There is a significant need to address questions such as which

healthy smokers are prone to developing smoking related disease, including COPD, and which patients with COPD are at a high risk of experiencing frequent exacerbations and disease advancement. Consequently, researchers studying COPD have analysed a diversity of airway samples in the hopes of identifying high risk individuals from the inflammation and the microbial community as measured in the airway samples.

6.2 Lower airway inflammation in COPD

It has been recognised that the lower airway immune system is altered in COPD compared with in health. Though the immune system in the airways is very complex, its main task is to eliminate pathogenic microorganisms and toxic substances before they can do harm to the respiratory organ. The airway immune system consists of barrier cells like the epithelial cells, specialised immune cells called leucocytes, and different molecules produced and released by these cells (Illustration 2).

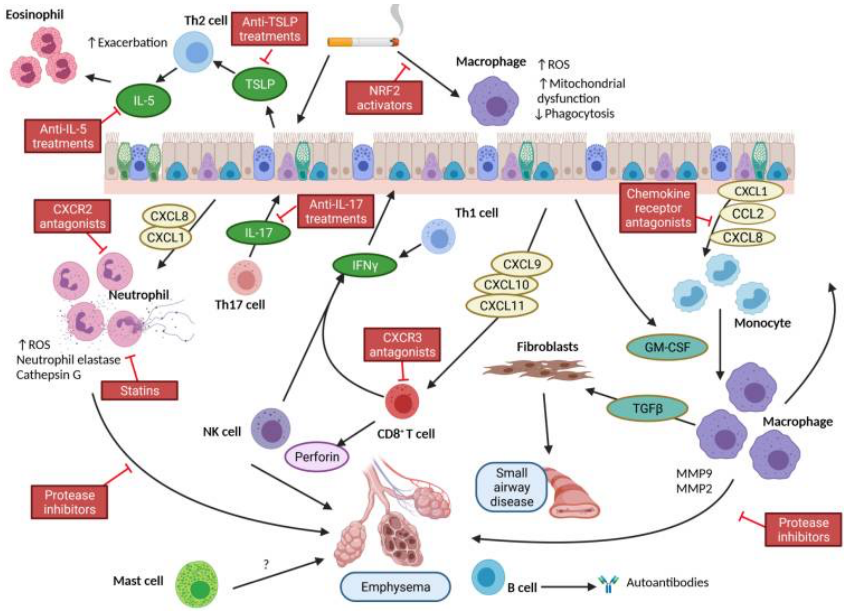

Illustration 2: Leukocytes and released molecules in chronic obstructive pulmonary disease (COPD). Shared under CC BY-NC 3.0 license: Baker et al. doi: 10.2147/COPD.S266394.

Typical for the lower airways of patients with COPD is a leucocyte pattern richer than normal in macrophages, neutrophils, cytotoxic T-cells, and helper T-cells type 1&17 (9, 10). Macrophages and neutrophils play a critical role in managing pathogenic microorganisms, but an imbalance in their activity in COPD is thought to contribute to the destruction of alveoli. This in part by increased release of protein degrading molecules called proteases (10). Furthermore, it has been found that the capacity of macrophages and neutrophils to eradicate microorganisms is reduced in the airways of patients with COPD (10). A subset of patients with COPD exhibits a higher prevalence of eosinophils, similar to what can be seen in many patients with asthma. For this group of patients inhaled corticosteroids (ICS) can be beneficial (11).

Important molecules in the immune response includes cytokines involved in cell signalling. These can be interleukins allowing for communication between leucocytes, and chemokines attracting immune cells to infected sites. The same molecule can have different functions, and therefore be recognised as both interleukins and chemokines. Interleukin-8 also known as chemokine ligand 8, is one example.

The immune system in the lower airways also contain cells with non-inflammatory functions that under the right circumstances engage with both pro- and anti-inflammatory activity. As an example, the airway epithelium is equipped with hair like cilia that with coordinated movements remove inhaled mucus and foreign particles and microorganisms. But the epithelium also possesses pattern recognition receptors that facilitate the identification of viruses and microorganisms, prompting the release of cytokines and antimicrobial peptides (AMPs) as a response. Moreover, the epithelium can release growth factors involved in leukocyte activity and the cell hypertrophy causing the bronchioles to narrow in COPD (9).

Smoking is one cause of COPD, but it is not the sole determinant as not all smokers develop COPD. Therefore, there must be unknown factors that play a role in triggering the remodelling of lung tissue and airways associated with COPD. Is it possible that the microbial community either residing in or visiting the lower airways could be such a factor?

6.3 The lower airway microbiota

Microbiota can be defined as the living microorganisms in a defined environment (12), as shown in Illustration 3.

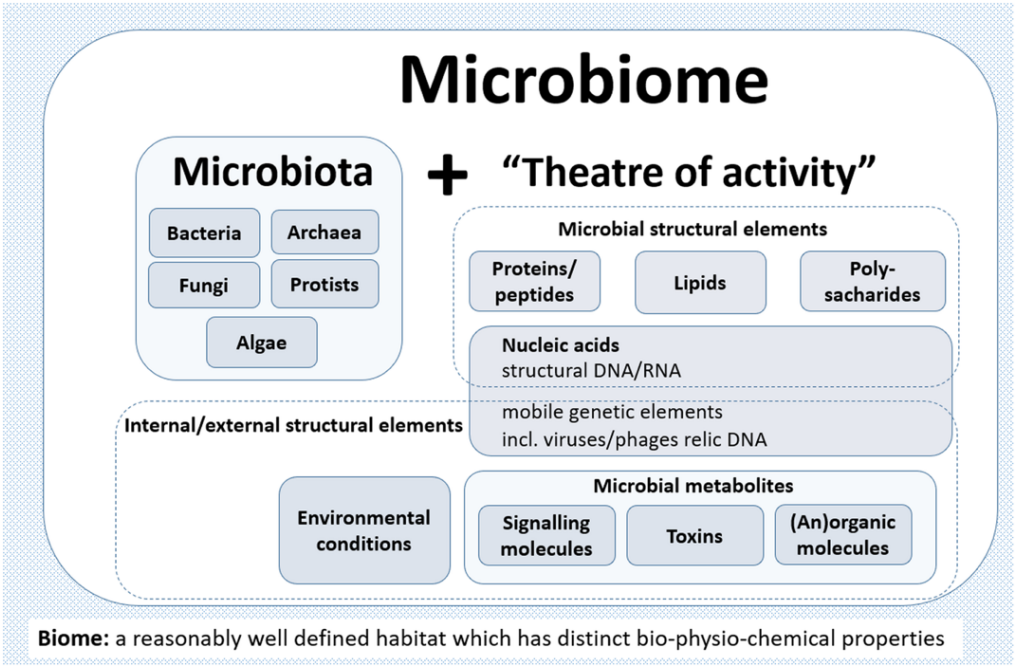

Illustration 3: Defining the microbiome. Shared under CC BY 4.0 license: Berg et al. doi.org/10.1186/s40168-020-00875-0.

From the work of Berg et al. we can see that the microbiota is not limited to bacteria, but also consists of fungi, archaea, protists, and algae. We also notice that the non-living viruses are not included in this definition. Today, bacteria are the most studied microorganism of them all.

Technological advancements have facilitated the classification of bacteria through sequencing of bacterial DNA, leading to valuable microbiota research findings. Currently, the most widely used sequencing technique is amplicon sequencing, (also called target gene sequencing), primarily because it is cost-effective. The 16S rRNA gene is particularly suited for amplicon sequencing due to its universal presence among all bacteria, enabling the differentiation of bacterial DNA from human DNA.

Additionally, the 16S rRNA gene contains both conserved and variable regions facilitating the classification of bacteria with bioinformatic analyses.

The Human Microbiome Project (HMP) financed by the United States National Institute of Health (NIH) was one of the first, large projects aiming to map the human microbiota (13). The HMP was initiated in 2008 and included 300 healthy participants sampled from many body sites with direct contact with the world around them. However, samples were not taken from the lower airways. The possibility of a "healthy" airway microbiota was thus not recognised until early 2010s when the works of Hilty et al., and Erb-Downward et al. presented a diverse collection of bacterial sequences from the lower airways of healthy controls (14). Studies on the airway microbiota (microbiome) in COPD available from the NIH online library (https://pubmed.ncbi.nlm.nih.gov) by the time my microbiota studies began in 2015 is presented in Table 1.

Table 1. Studies on the airway microbiota (microbiome) in COPD published by 2015.

First author: Title. Publisher	Aim	Participants	Sampling methods	Sequencing technique	Results
Hilty, Markus: Disordered microbial communities in asthmatic airways. PlosOne 2010	To describe and compare the airway microbiota in asthma, COPD, and healthy controls.	5 COPD 11 Asthma 8 Control	Oropharyngeal swabs Bronchial brushes	Pyro-sequencing	Alpha-diversity is rich in COPD, asthma, and in healthy controls. Asthma samples clustered with control samples from the upper respiratory tract, and with COPD samples from the lower respiratory tract. *Proteobacteria* was enriched in obstructive lung disease at the cost of *Bacteroidetes*.
Huang, Yvonne: A persistent and diverse airway microbiota present during chronic obstructive pulmonary disease exacerbations. OMICS 2010	To describe the airway microbiota in patients with exacerbated COPD treated with antibiotics.	8 COPD	Endotracheal aspirates	16S rRNA PhyloChip	Alpha-diversity in COPD is rich, and exacerbations could be a polymicrobial process, rather than caused by a single pathogenic bacterium. Taxonomy included genera known to inhabit the upper respiratory and gastric tract.
Erb-Downward, John: Analysis of the lung microbiome in the "healthy" smoker and in COPD. PlosOne 2011	To describe the airway microbiota in COPD and health, and to investigate for differences along the respiratory tract.	8 COPD 14 Control	Lung tissue Bronchoalveolar lavage (BAL)	Pyro-sequencing	The lower airway microbiota in health is diverse, and distinct from that of the upper respiratory tract. Compared with controls, the diversity of lower airway microbiota in COPD is lesser and depending on collection site along the respiratory tract.

Cabrera-Rubio, Raúl: Microbiome diversity in the bronchial tracts of patients with chronic obstructive pulmonary disease. J Clin Microbiol 2012	To describe the airway microbiota in stable COPD and compare upper and lower airways samples.	6 COPD	Induced sputum Bronchial aspirate BAL Bronchial mucosa	Pyro-sequencing	Alpha- and beta-diversity differed between upper and lower respiratory tract samples. The upper respiratory tract samples shared high relative abundances of common oropharyngeal phyla.
Pragman, Alexa: The lung microbiome in moderate and severe chronic obstructive pulmonary disease. PlosOne 2012	To describe differences between the airway microbiota in patients with COPD with that of healthy controls.	22 COPD 10 Control	BAL	Pyro-sequencing	Alpha-diversity associated with age, but not COPD. Associations between use of inhaled COPD medication and beta-diversity was suggested. The taxonomy in BAL indicated micro aspiration from the upper respiratory tract to be the source of the lower airway microbiota.
Sze, Marc: The lung tissue microbiome in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2012	To describe the microbiota in lung tissue in very severe COPD, non-smoking, and smoking controls, and in cystic fibrosis (CF).	8 COPD 8 CF 16 Control	Lung tissue	Pyro-sequencing	Alpha-diversity in lung tissue did not differ between COPD and controls, but associations between COPD and beta-diversity was found. The relative abundance of *Firmicutes* was significantly higher in COPD compared with controls.
Molyneaux, Philip: Outgrowth of the bacterial airway microbiome after rhinovirus exacerbation of chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2013	To describe the effect of rhinovirus infections on the airway microbiota in COPD.	14 COPD 17 Control	Induced sputum	Pyro-sequencing	Alpha- and beta-diversity were not associated with COPD. The most prevalent taxa were shared in COPD and controls, but *Proteobacteria* was significantly more abundant in controls at the cost of *Firmicutes*. Rhinovirus infections did affect the sputum microbiota in COPD, but not in controls.

Zakharkina, Tetyana: Analysis of the Airway Microbiota of Healthy Individuals and Patients with Chronic Obstructive Pulmonary Disease by T-RFLP and Clone Sequencing. PlosOne 2013	To describe the airway microbiota in COPD and health.	9 COPD 9 Control	BAL	ABI DNA sequencer	Alpha-diversity in COPD and healthy controls were equally rich and even assessed with Shannon diversity index. While several species were unique to one group, several were also shared by all participants.
Galiana, Antonio: Sputum microbiota in moderate versus severe patients with COPD Eur Respir J 2014	To study the airway microbiota across moderate and severe COPD.	19 COPD	Spontaneous sputum	Pyro-sequencing	The moderate COPD group had higher alpha-diversity of bacterial genera, and sputum samples clustered closer to each other compared with in the severe COPD group.
Garcia-Nuñez, Marian: Severity-related changes of bronchial microbiome in chronic obstructive pulmonary disease. J Clin Microbiol 2014	To study the airway microbiota across disease severity in COPD.	17 COPD	Spontaneous sputum	Pyro-sequencing	Alpha-diversity was inversely correlated with lung function measured as FEV_1. Associations between COPD severity and beta-diversity was not found. *Proteobacteria* was significantly more abundant with advanced disease at the cost of *Firmicutes*.
Huang, Yvonne: Airway Microbiome Dynamics in Exacerbations of Chronic Obstructive Pulmonary Disease. J Clin Microbiol 2014	To describe the airway microbiota dynamics associated with exacerbated COPD and the treatment provided.	12 COPD	Spontaneous sputum	16S rRNA PhyloChip	Alpha-diversity was not significantly different across sampling timepoints. Beta-diversity was associated with different treatments (corticosteroids and/or antibiotics) prescribed for exacerbations. Changes in taxonomic abundances was depending on treatment strategies.

Millares, Laura: Bronchial microbiome of severe COPD patients colonised by Pseudomonas aeruginosa. Eur J Clin Microbiol Infect Dis 2014	To study airway microbiota dynamics in stable state and exacerbated COPD related to colonisation by *P. aeruginosa*.	16 COPD	Spontaneous sputum	Pyro-sequencing	In stable state COPD, colonisation by *P. aeruginosa* was recognisable by increased relative abundance of the genus, while the sputum microbiota in exacerbated COPD was similar between colonised and non-colonised.
Tan, Lisi: 16S rDNA-based metagenomic analysis of dental plaque and lung bacteria in patients with severe acute exacerbations of chronic obstructive pulmonary disease. J Periodontal Res. 2014	To study the microbiota in dental plaques and tracheal aspirate to evaluate if oral microbiota causes COPD exacerbations.	53 COPD	Dental plaques Tracheal aspirates	ABI DNA sequencer	Several pathogenic genera, among them *P. aeruginosa*, *K. pneumoniae* and *S. pneumoniae* were found in both sample types collected from the same patients during exacerbations of COPD.
Aguirre, Estefania: Analysis of microbiota in stable patients with chronic obstructive pulmonary disease. APMIS 2015	To describe the airway microbiota in COPD and compare sequencing results with culture methods.	19 COPD	Spontaneous sputum	Pyro-sequencing	Pyrosequencing revealed a far more diverse bacterial microbiota in spontaneous sputum than did culturing methods. Still, detection of *Enterobacteriaceae* in cultures was only confirmed in 1 of 7 patients with pyrosequencing.
Huang, Yvonne: The Sputum Microbiome in Chronic Obstructive Pulmonary Disease Exacerbations. Ann Am Thorac Soc 2015	To study airway microbiota dynamics across stable state and exacerbated COPD	12 COPD	Spontaneous sputum	16S rRNA PhyloChip	Sputum microbiota changes were heterogenic. Bacteria diminished at exacerbation could be important for inflammatory activity, and the time to re-establish a stable microbiota was prolonged if exacerbations were treated with antibiotics alone.

Millares, Laura: Functional Metagenomics of the Bronchial Microbiome in COPD. PlosOne 2015	To identify the functional changes in the airway microbiota across stable state and exacerbated COPD.	8 COPD	Spontaneous sputum	Pyro-sequencing	No differences in the sputum microbiota were found between stability and exacerbation, but metabolic functional patterns changed during exacerbations.
Su, Jin: Sputum Bacterial and Fungal Dynamics during Exacerbations of Severe COPD. PlosOne 2015	To study day-to-day changes in the airway microbiota in hospitalised patients with exacerbated COPD.	6 COPD	Sputum, likely spontaneous	Illumina Hiseq	Variations in alpha-diversity of both bacteria and fungi showed no consistent patterns across patients. Taxonomic variations from day to day was common, but without recognisable patterns across patients.
Sze, Marc: Host Response to the Lung Microbiome in Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med 2015	To describe the airway microbiota and associated immune responses involved in the pathogenesis of very severe COPD.	5 COPD 4 Control	Lung tissue	Pyro-sequencing	Alpha-diversity richness and beta-diversity was associated with severe COPD. *Proteobacteria* was significantly more abundant in severe COPD than in controls, at the cost of *Firmicutes* and *Bacteroidetes*. Some associations between the airway microbiota and host immune responses were suggested.

From Table 1 we can observe that the earliest lower airway microbiota studies were mainly descriptive, had limited numbers of participants, and presented airway microbiota from various sample types. Certain aspects from Table 1 can be highlighted:

First and foremost, all studies including healthy controls confirmed the existence of a healthy lower airway microbiota as first suggested by Hilty et al. Hence, the belief that the healthy lower airways were sterile had already been disregarded by 2015.

Secondly, Huang et al. challenged the notion that exacerbations were caused by a single pathogenic bacterium (15). She proposed that a more complex bacterial dysbiosis, involving multiple bacteria, could be responsible for exacerbations.

Thirdly, measures of alpha- and beta-diversity in both health and COPD were inconsistent between studies, as well as in relation to the severity of COPD. Similarly, various studies have provided conflicting results on the proportions of prevalent bacterial phyla in relation to both health and COPD, as well as in relation to the severity of COPD.

Furthermore, the measurements of the microbiota were affected by the choice of sampling sites and methods. In the research conducted by Hilty et al. differences in sample sites were only found among patients with obstructive lung disease, not in healthy controls. Erb-Downward et al., and Cabrera-Rubio et al. discovered that sampling methods and sites influenced the composition of the airway microbiota.

Notably, there was a lack of studies investigating the relationship between the airway microbiota and airway inflammation, as evident in Table 1.

Studies of the airway microbiota in COPD have grown rapidly in size and complexity, with researchers attempting to link the microbiota to host characteristics and host inflammation. This shift is exemplified in Table 2, which provides an overview of larger studies conducted after 2015.

Table 2. Selected publications on the airway microbiota in COPD focusing on larger studies.

First author: Title. Publisher	Aim	Participants	Sampling methods	Sequencing technique	Results
Gronseth, Rune: Protected sampling is preferable in bronchoscopic studies of the airway microbiome. ERJ Open Res 2017	To investigate the susceptibility of oropharyngeal contamination of various bronchoscopic samples from a COPD cohort.	67 COPD 58 Control	Oral wash (OW), small-volume lavage (SVL), protected broncho-alveolar lavage (BAL), protected specimen brushes (PSB)	Illumina MiSeq	Alpha-diversity and *Firmicutes* decreased and *Proteobacteria* increased in the order OW, SVL, BAL, PSB. Beta-diversity varied by sampling method, and visualisation of principal coordinates analyses indicated that differences in diversity were smaller between OW and SVL and OW and protected BAL samples than for OW and the protected PSB samples.
Wang, Zhang: Sputum microbiome temporal variability and dysbiosis in chronic obstructive pulmonary disease exacerbations: an analysis of the COPDMAP study. Thorax 2018	To investigate sputum microbiota in COPD across stable state and exacerbations with a longitudinal study design.	281 COPD	Induced and spontaneous sputum	Illumina MiSeq	Alpha-diversity reduced during exacerbations, and with increasing *Moraxella.* Exacerbations with eosinophilic sputum: Higher alpha-diversity and decreased *Proteobacteria* compared with bacterial exacerbations. A definition of dysbiosis was met in 49 of 119 exacerbations. A combination of serum eosinophilia and dysbiosis associated with greater lung function decline and CAT scores.

Mayhew, David: Longitudinal profiling of the lung microbiome in the AERIS study demonstrates repeatability of bacterial and eosinophilic COPD exacerbations. Thorax 2018	To investigate changes in the lung microbiome across stable state and exacerbations for one year, and how the microbiota associated with COPD outcomes.	101 COPD	Induced and spontaneous sputum	Illumina MiSeq	Alpha-diversity reduced with increasing COPD severity. *Haemophilus* enriched in very severe COPD, at the cost of both *Prevotella* & *Veillonella*. *Moraxella* relatively more abundant during exacerbations and opposed to *Lactobacillus*, *Moraxella* was positively associated with exacerbation frequencies. Tendency to repeat exacerbation phenotype. Lower difference in beta-diversity within individuals than between individuals, and larger differences when exacerbations occurred.
Ramsheh, Mohammadali: Lung microbiome composition and bronchial epithelial gene expression in patients with COPD versus healthy individuals: a bacterial 16S rRNA gene sequencing and host transcriptomic analysis. Lancet Microbe 2018	To investigate associations between lower airway microbiota across COPD and health, across use of inhaled corticosteroids (ICS), and with lower airway inflammation.	339 COPD 207 Control	Bronchial brushes	Illumina MiSeq	Alpha-diversity reduced in COPD, no association with use of ICS. *Prevotella, Streptococcus, and Moraxella* separated control and COPD samples best. *Prevotella*: Reduced in ICS users, positively associated with lung function, negatively with COPD symptoms, linked with inflammation-associated host-genes.

Williams, Nicholas: Impact of radiologically stratified exacerbations: insights into pneumonia aetiology in COPD. Respir Res 2018	To investigate associations between sputum microbiota, inflammation, and radiological findings in COPD exacerbations.	127 COPD	Induced and spontaneous sputum	Illumina MiSeq	Chest x-ray infiltrates not associated with diversity nor taxonomy based on 16S rRNA sequencing, but it associated with increased *H. Influenzae* from cultures. Sputum neutrophils and serum inflammation were positively associated with infiltrates.
Diver, Sarah: Sputum microbiomic clustering in asthma and chronic obstructive pulmonary disease reveals a Haemophilus-predominant subgroup. Allergy 2020	To investigate for effects of *Proteobacteria* and *Firmicutes* balances on COPD and asthma.	78 COPD 63 Asthma	Induced and spontaneous sputum	Pyro-sequencing	*Haemophilus* dominance differentiated participants in a *Haemophilus*-high and a *Haemophilus*-low group. *Proteobacteria*:*Firmicutes* ratio could distinguish between groups. The *Haemophilus*-high group was COPD-high relative to asthma and associated with low alpha-diversity and increased neutrophils, IL1β and TNF-α.
Wang, Zhang: A Refined View of Airway Microbiome in Chronic Obstructive Pulmonary Disease at Species and Strain-Levels. Front Microbiol 2020	To investigate species and strain level microbiota in sputum collected from patients with COPD.	98 COPD 27 Control	Induced sputum	Pacific Biosciences	Beta-diversity differences with weighted UniFrac between COPD and controls, but not across smoking or ICS use. Species data allowed for better discrimination between COPD and controls than genera data in a random forest classification model. Heterogenous changes were observed for species within the same genus, indicating that changes can be overlooked when considering genera. Species level taxa associated both with clinical variables and inflammatory markers.

Wang, Zhang: Inflammatory Endotype–associated Airway Microbiome in Chronic Obstructive Pulmonary Disease Clinical Stability and Exacerbations: A Multicohort Longitudinal Analysis. Am J Respir Crit Care Med 2021	To investigate associations between sputum microbiota and leukocyte patterns across stable state and exacerbations in COPD. Longitudinal study design.	510 COPD	Induced and spontaneous sputum	Pyro-sequencing Illumina MiSeq	Alpha-diversity in neutrophilic sputum lower than in sputum classified by other leukocytes. Neutrophilic sub-classes based on *Haemophilus*, *Moraxella* or *Streptococcus* dominance was less stable if balanced compared with eosinophilic and un-balanced neutrophilic sputum. Neutrophilic sub-classes associated with different cytokines. ICS did not affect stability of sputum classes.
Dicker, Alison: The sputum microbiome, airway inflammation, and mortality in chronic obstructive pulmonary disease. J Allergy Clin Immunol 2021	To investigate associations between sputum microbiota and blood leukocytes, disease characteristics and mortality in COPD.	252 COPD	Induced sputum	Illumina MiSeq	Alpha-diversity associated with risk of death. *Proteobacteria* associated with low blood-eosinophiles, neutrophil release products, chronic bronchitis (CB), frequent exacerbations, reduced lung-function, and increased risk of death. *Streptococcus* associated with high blood-eosinophils, frequent exacerbations, and absence of CB.
Opron, Kristopher: Lung microbiota associations with clinical features of COPD in the SPIROMICS cohort. NPJ Biofilms Microbiomes 2021	Explore associations between differences in lung bacterial composition and COPD and/or clinical features that reflect COPD pathophysiology.	78 COPD 103 Control	BAL	Illumina MiSeq	Beta-diversity not associated with COPD, but with pulmonary function measures and CAT scores. *Streptococcus* and *Lactobacillales* OTUs were enriched in COPD compared with controls. ICS use did not affect the microbiota. Smoking associated with OTUs assigned to *Streptococcus*, *Lactobacillales* and *Veillonella*. *Streptococcus* associated with increased BAL neutrophils.

6.4 Airway sampling procedures

It is apparent from both Table 1 and Table 2 that different sample types have been used to characterise the airway microbiota. To sampling sputum is convenient as the procedures are of a less invasive nature, and require less resources compared with bronchoscopic sampling procedures. However, one drawback of sputum is that it must pass through the oral cavity and the much richer microbiota present there. This will inevitably lead to sample contamination.

Two types of sputum can be collected: Participants can deliver spontaneous sputum simply by coughing and expectorating, or they can be induced. With this procedure a machine (nebuliser) turns saline into a mist that can be inhaled. This increases sputum production making sampling from non-expectorating participants possible. The standardisation of induction techniques has been proposed by a European task force over 20 years ago (16). Limited publications exist regarding the potential impact of sputum sampling methods on measurements of inflammatory markers and microbiota in patients with COPD. When commencing this thesis work in 2013, it had been reported that induced sputum from patients with asthma and COPD had higher cell viability compared to spontaneous sputum, IL-8 levels were similar across different types of sputum in stable COPD (17, 18), and no studies comparing microbiota in induced and spontaneous sputum in patients with COPD had been published.

The use of standardised sampling methods, such as induced sputum, would be a logical choice for research purposes, also due to the ability to include a larger number of participants. However, it is important to exercise caution when sampling patients with exacerbated COPD and asthma, as induction has been shown to obstruct the airways (16). The European task force recommended using isotonic saline instead of concentrated saline for high-risk patients, gradually increasing the concentration if sampling failed and isotonic saline was well-tolerated. By 2013, the safety of induction with concentrated saline had been evaluated in stable asthma, stable COPD, and exacerbated mild to moderate COPD (19-22). In 2013 Gao et al. published their assessment of the safety of induction during exacerbations in severe and very severe

COPD using isotonic saline. They concluded that this modified procedure was safe for most exacerbated patients with severe and very severe COPD (23).

Although sputum is commonly used, and safely sampled during most exacerbations of COPD, it is obtained from the proximal parts of the lower airways (24). To obtain biological samples from the distal airways, bronchoscopy should be performed. Bronchoscopic samples, such as bronchoalveolar lavage (BAL), bronchial brushes, and transbronchial biopsies, carry substantially less risk of oral contamination compared to sputum. This is particularly true when employing procedural steps aimed at reducing contamination. One example of measures that should be taken is to refrain from suction through the bronchoscope until it has reached the lower airways (25-27). Equipment specifically designed to reduce the risk of contamination, such as wax-sealed sterile specimen brushes and wax-sealed sterile inner catheters for BAL, can further decrease the risk of oral contamination (26, 28).

Currently, there is no consensus on which type of sample is preferable for assessing the airway microbiota in COPD. It is reasonable to assume that samples obtained near the sites of airway remodelling would provide more accurate insight into the environment responsible for such changes. However, both BAL and bronchial brush samples cover a more limited portion of the lower airways compared to sputum, which is likely to be supplied from both lungs. Another important limitation of research bronchoscopy is that the procedure has only been conducted in stable state COPD, and not in patients with ongoing exacerbations.

6.5 How to plan a study?

The overall objective of research is to acquire knowledge about a specific subject. The selection of a study design is crucial in achieving this objective. In the field of medical research, randomised, controlled trials (RCTs) are usually considered the superior study design (29), but they do have their limitations, as discussed by Bosdriesz et al. (30). For instance, envision a study looking into the correlation between smoking and dementia, where one group of healthy individuals is randomly assigned to smoke 20 cigarettes per day for 20 years, while another group remains smoke-free. Due to the well-known health risks associated with smoking, such an RCT would (hopefully) never receive approval from an ethics committee. However, an observational cohort study could be conducted to examine whether there is a higher incidence of dementia among smokers compared to non-smokers, without imposing smoking on the participants. Additionally, cohort studies have the advantage of being more cost-effective, allowing for larger-scale studies compared to RCTs (29). A potential drawback with observational cohort studies is the time needed to observe outcomes. Case-control studies in which the outcome identifies cases and controls, and inclusion depends on available exposure data can effectively overcome this issue. Important limitations to such case-control studies include confounding, recall biases and inappropriate inclusion of controls (31). Confounding refers to the influence of factors other than those under study. To mitigate this, stratification can be used (31). For example, dementia is more prevalent in the elderly, who are also more likely to have smoked for a longer duration than younger participants. Therefore, studies on smoking and dementia should analyse data from young and elderly individuals separately to reduce the confounding effect of age. Another approach is to use cases as their own controls in a crossover design (31). Traditionally, case reports have been regarded as the least reputable study design, with the focus on in-depth investigations of individual participants (29). Study designs can either be longitudinal, comparing measurements over time, or cross-sectional, where measurements are taken at a single point in time. The latter design does not allow for causal inference.

It is essential to obtain and process samples according to validated methods (26, 32). The internal validity must be acceptable, meaning that the results should accurately reflect the sampled environment. Additionally, external validity is important, as we need to ensure that what we observe in the study population is representative of the general population. Furthermore, we want the results to be repeatable and reliable. It is necessary to determine if observations made are random or not, and if the data obtained will suffice to detect significant patterns in our data. This requires us to assess if the statistical power is high enough to allow us to draw conclusions.

Considering the factors mentioned above, it becomes evident that it is important to clearly formulate the questions one intends to address and select an appropriate study design and sampling methods in alignment with those questions.

6.6 Microbiota data and statistical analyses

Some challenges arise when working with microbiota data and statistical analyses. One of these challenges is the high zero-counts in sequencing data. Many bacterial sequences exist in low quantities and can only be found in a few samples. Several established statistical methods struggle to handle datasets with high zero-counts, but there are ongoing efforts to develop mathematical approaches to address this issue (33). Additionally, the compositionality of sequencing data complicates the interpretation of changes and differences (34). This was effectively demonstrated in Leiten's PhD thesis *The airway microbiota of stable COPD* (Illustration 4).

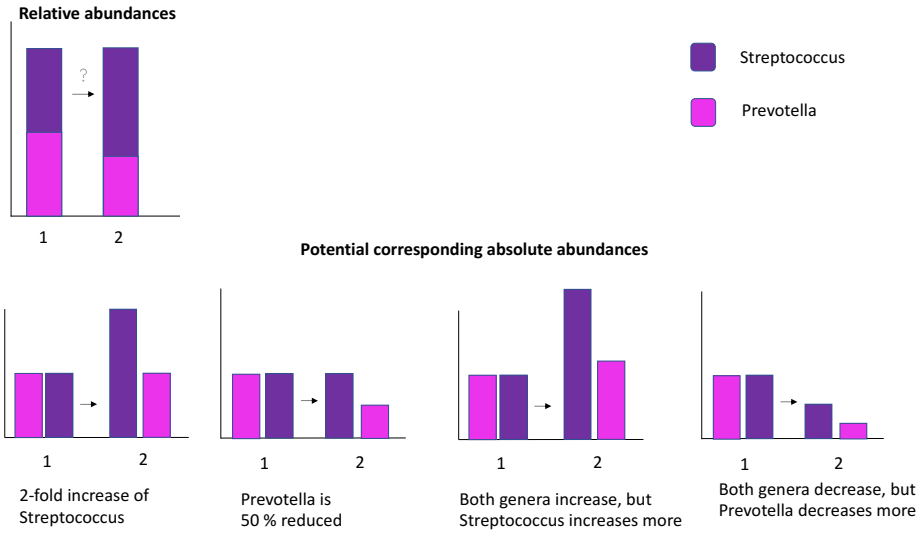

Illustration 4: Four scenarios of absolute abundances that correspond to the relative abundance of *Streptococcus* and *Prevotella* in samples 1 and 2. Illustration reprinted with permission from PhD, MD Elise Leiten.

6.6.1 Bioinformatic pipelines

Microbiota data comprises an extensive number of sequences (reads), rendering manual curation impractical. Consequently, these data need to be processed with bioinformatical pipelines, of which Quantitative Insights Into Microbial Ecology (QIIME) 1&2 and Mothur are commonly employed (35, 36). Furthermore, data curation can be conducted using diverse tools created within the open statistics environment known as R (37). The bioinformatic tools aid in transforming the millions of sequences, depicted as lengthy strings of letters, with each letter symbolising one base in the amplicon sequence, into comprehensible data.

6.6.2 Microbiota characteristics

Descriptions of microbiota typically include *diversity* and *taxonomy*. In 1972, biologist Whittaker introduced the concepts of alpha-, beta-, and gamma-diversity,

which help to explain the diversity both within and between samples (38). Since then, a wide range of mathematical diversity models have been developed improving our understanding of the complexity of microbiota data. It is important to recognise that there are differences between diversity analyses to fully grasp the intricacies of the microbial diversity.

The taxonomy of microbiota is described at various hierarchical levels, beginning with *kingdom* followed by *phyla*, *classes*, *orders*, *families*, *genera*, *species*, and even *sub-species*. However, caution must be taken when interpreting bacterial taxonomy below the genus rank, due to limitations in 16s rRNA target gene sequencing (39). Additionally, including more ranks in analyses increases the number of comparisons made and the risk of false positive observations.

The complexity of microbiota data has led to the development of several statistical methods for comparing taxonomy between groups. These differential abundance analyses have been constantly evolving, and the methods used for this thesis are summarised in the methods section (Table 3).

6.7 Potential implications of microbiota research for patient care

The most notable illustration of utilising microbiota to treat disease is faecal transplants for antibiotic resistant *Clostridium difficile* colitis (40). Unfortunately, diseases like COPD are not likely to be caused or worsened by a single treatable microorganism, as in the case of *C. difficile* infections. Therefore, targeting individual pathogens would not be effective in preventing COPD.

Hopefully, methodical, and descriptive studies have laid the foundation for future research endeavours focused on unravelling the intricate interplay between host inflammation and the airway microbiota and its disease propagating characteristics. Aided by continually advancing statistical and bioinformatic tools this research field is far from exhausted and can be important for improved patient care. Although we may not achieve the same level of success for patients with COPD, as for the *C. difficile* sufferers, there is reason to believe that the lower airway microbiota plays an important role for the human health as summarised by Man et al. (41). Perhaps, in the future, it may be feasible to administer personalised treatments to distinct sub-groups of smokers and patients with COPD by considering their microbiota-inflammatory profiles.

7. Objectives of this thesis

I conducted the two first studies with the intention to understand the influence of different methods of sputum sampling on the measurement of lower airway inflammation and microbiota and to use the results to direct the sample selection for my third study.

Next, my goal was to identify specific patterns of change in inflammation and microbiota in induced sputum from patients with COPD enrolled if sampled both during stable and exacerbated states.

Lastly, I aimed to find potentially important differences in the lower airway microbiota between a large group of patients with COPD and a similarly sized group of healthy participants, as measured in BAL.

8. Material and methods

8.1 Research studies

The data in this thesis was acquired from three extensive cohort studies concerning patients with COPD and healthy controls residing in western Norway:

The *Bergen COPD Cohort Study* (*BCCS*) followed 433 patients with COPD over a span of three years, from 2006 to 2010 (42-45). To be eligible for inclusion as a COPD patient a study physician had to clinically confirm the COPD diagnosis based on both respiratory symptoms and spirometric criteria, with post-bronchodilator measures of forced expiratory volume in 1 sec/forced vital capacity (FEV_1/FVC) <0.7 and FEV_1 <80% of predicted. In addition, patients with COPD were required to have a smoking history of ≥10 pack years. Candidates with other lung diseases, connective tissue disorders, inflammatory bowel disease, or cancer during the last five years were excluded. The data collected within the *BCCS* and used in this thesis consists of medical history including age, sex, smoking history, information on medications, exacerbation history, body composition measurements, lung function testing with COPD severity classification (6), induced sputum samples and spontaneous sputum samples.

The *Bergen COPD Exacerbation study* (*BCES*) followed 356 patients from the *BCCS* who resided within the Bergen hospital district (46, 47). These patients were given detailed descriptions of COPD exacerbation symptoms and were encouraged to contact the hospital if they experienced such symptoms. A telephone service staffed by a study nurse was open 12 hours per day, seven days a week, for the duration of the study. The study nurse determined if immediate hospitalisation was needed, or if a next working day outpatient clinic visit with a study physician was sufficient. The data collected within the *BCES* and used for this thesis is identical to the data from the *BCCS*, but with the addition of serum white blood cell counts (WBC) and absolute neutrophil counts (ANC) (9).

The *Bergen COPD microbiome study* (*MicroCOPD*) included 249 participants, with data collected between 2012 and 2016 (26, 48, 49). Among these participants, 130 were diagnosed with COPD and 103 served as healthy controls. The confirmation process for a COPD diagnosis followed the same criteria as in the *BCCS*, except that COPD with the Global initiative for chronic Obstructive Lung Disease (GOLD) stage I were also allowed. Participants were excluded if they were deemed unsuitable for bronchoscopy. Factors that would prevent bronchoscopy included oxygen saturation <90% with oxygen supplementation, pCO_2 in arterial blood >6.65 kPa, increased risk of bleeding and a known allergy towards the premedication. Cardiac risks such as acute coronary syndrome last six weeks, severe pulmonary hypertension, or valve prosthesis ruled out bronchoscopy. The use of antibiotics and/or systemic corticosteroids two weeks prior to the scheduled bronchoscopy, as well as ongoing airway infections, would only delay inclusion. The data collected within the *MicroCOPD* study and used for this thesis include medical history (including age, sex, smoking history, information on medications, exacerbation history), lung function testing with COPD severity classification (6) and BAL.

Flowcharts depicts how participants from the three studies were selected for my studies, based on availability and stringent quality screening of all biological samples (Figure 1 and Figure 2).

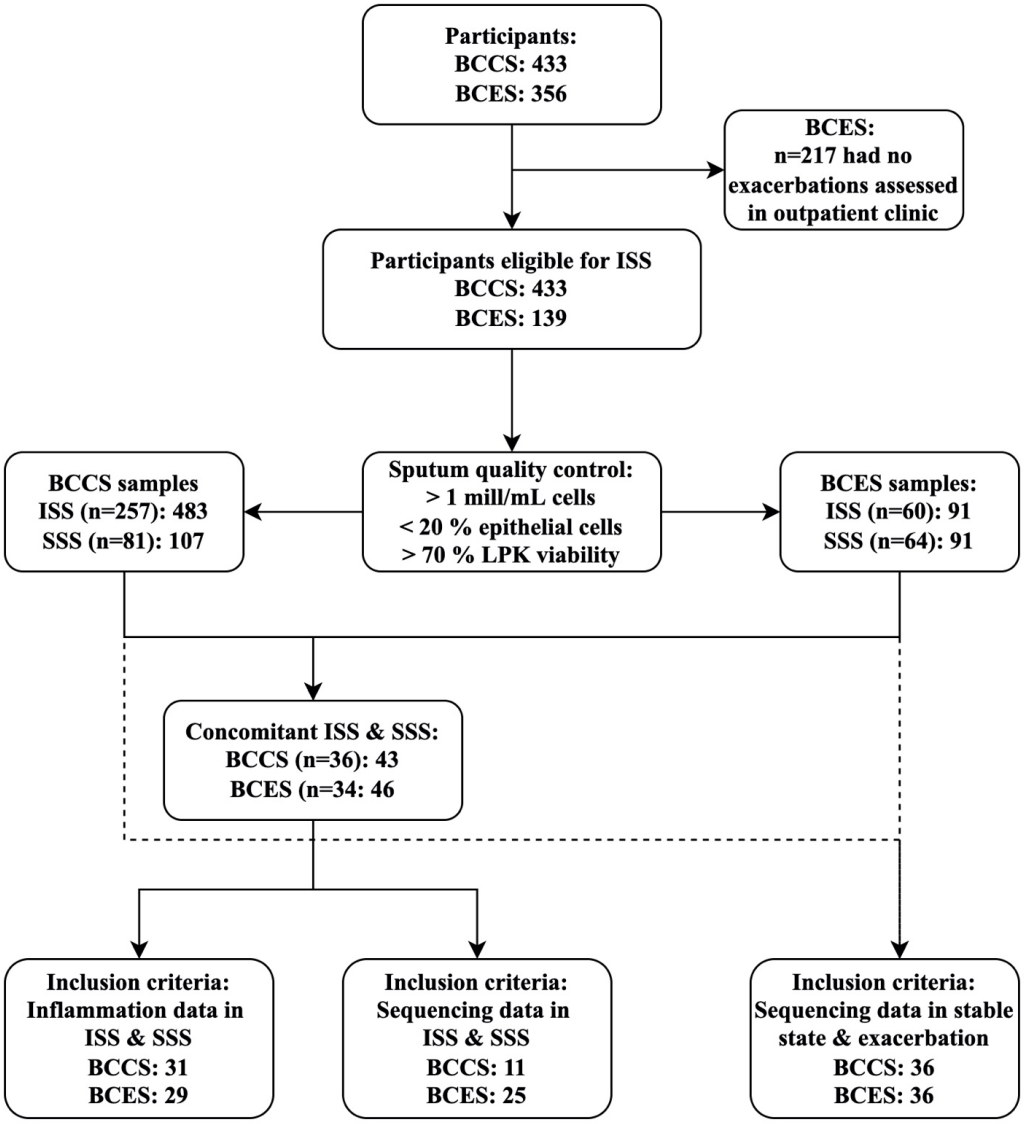

Figure 1: Sample selection for sputum studies: Flowchart depicting the sample selection process from the *Bergen COPD cohort study* (*BCCS*) and the *Bergen Exacerbation Cohort study* (*BCES*). Induced sputum samples (ISS) and spontaneous sputum samples (SSS) for evaluation of sampling effect and induced sputum for evaluation of exacerbation effect on inflammation and microbiota.

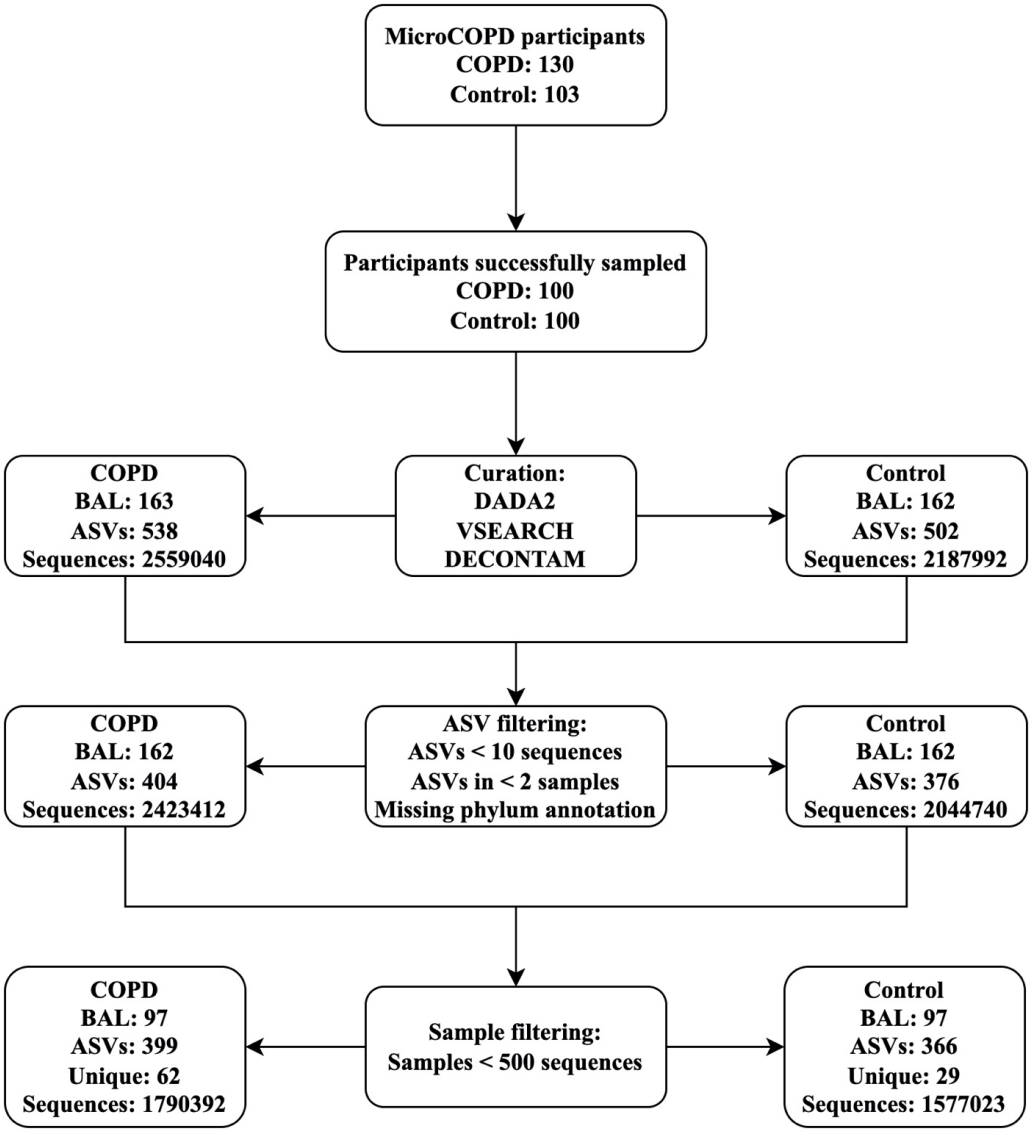

Figure 2: Sample selection for the bronchoalveolar lavage (BAL) study: The *Bergen microbiome study* (*MicroCOPD*) providing BAL for evaluation of differences in microbiota in a COPD cohort and healthy controls. COPD: Chronic obstructive pulmonary disease. ASVs: Amplicon sequence variants. Unique: Found only in COPD or only in control.

8.2 Sampling and sample processing.

8.2.1 Sputum sampling

The collection of sputum began by attempting to collect spontaneous sputum. Participants were instructed to rinse their mouths with water and blow their noses before expectorating into two separate cups. After spontaneous sputum collection was attempted, sputum induction was performed. To prevent bronchoconstriction during and after induction and to obtain a post-bronchodilator spirometry, participants were given salbutamol inhalations as a pre-treatment (19, 50). Induction was only carried out if participants had an oxygen saturation level of at least 90% after the salbutamol treatment. Inhalations of saline were administered using an ultrasonic wave nebuliser, with each inhalation lasting seven minutes and being repeated three times consecutively. Sterile hypertonic saline (3%) was the preferred choice, but in 10 cases, the procedure was modified to use 0.9% saline due to participants experiencing clinical obstruction or refusing to inhale the hypertonic saline. Spirometry tests were conducted before and after each induction, using the Vitalograph Ltd. S-model in stable states and the EasyOne model 2001 Ndd Medizintechnik AG during exacerbations. The induction procedure was stopped if the FEV_1 declined by more than 20%, if symptoms worsened, or if the participants chose not to continue.

8.2.2 Sputum sample processing

Both types of sputum samples were immediately placed on ice until processed for quality control and storage, typically within 30 minutes. Dithiothreitol 0.1% (DTT) was added to break disulphide bonds in mucin at a ratio of 4 ml/gram sputum (51). The samples were then homogenised using an Eppendorf homogeniser at 600 rounds per minute for 15 minutes at a temperature of 4° Celsius. Phosphate-buffered saline (PBS) was added, and the samples were filtered to enhance homogenisation. After centrifugation at 450g for 15 minutes at 4° Celsius, the supernatants were removed, divided into 0.5ml tubes, and stored at −80° Celsius. These aliquots were used to measure inflammatory markers and AMPs.

The cell pellets were re-suspended in PBS for quality evaluation and leukocyte differential counts. Trained personnel assessed viability after staining with tryptan blue. The samples were considered acceptable if they contained >1 million cells/mL, <20% epithelial cells, and the leucocyte viability were >70%.

The cell suspension remaining after preparation of Cytospin slides for leukocyte differential counts was centrifuged at 450g (5 minutes, 4° Celsius). The resulting cell pellet was resuspended in cold TRIzol (Invitrogen, Product No. 15596-026), divided into 0.5 ml tubes, and stored at −80° Celsius. These aliquots were used for microbiota analyses. Negative controls were not stored.

Sputum samples collected after December 2006 were cultured at the Department of Microbiology, Haukeland University Hospital. Albumin measurements were conducted by enzyme immunoassay for paper I to enable concentration correction of measured inflammatory markers. $Albumin_{induced}/Albumin_{spontaneous}$ was multiplied with each concentration of the inflammatory markers.

Cytokine analyses were performed using the Luminex® xMAP® technology (Luminex Corporation, Austin, Texas) (52). Standards from BioRad (Bio-Plex Pro Human Cytokine Standards Group I 27-Plex #171-D50001, Lot No 5022130. Bio-Plex Pro Human Cytokine Standards Group II 23-Plex #171-D10502 Lot No 5015357) along with bead-based multiplex assays allowed for the analysis of Interleukin-6, interleukin-8, interleukin-18, interferon gamma-inducible protein-10, monokine induced by gamma interferon, and tumour necrosis factor-alpha (IL-6, IL-8, IL-18, IP-10, MIG and TNF-α) in simplex analyses. A Luminex 100 instrument was used according to the manufacturer´s instructions, and the results were collected and stored using STarStation software version 2.0 (STarStation Software Version 2.0, Applied Cytometry, Sheffield, UK).

Two AMPs, LL-37, and secretory leucocyte protease inhibitor (SLPI) were examined. Both AMPs were obtained from a previous study (47). LL-37 was measured using an enzyme-linked immunosorbent assay (ELISA) according to the manufacturer´s

instructions (Hycult Biotech, Uden, the Netherlands). SLPI was measured using an ELISA protocol developed at the Laboratory for Respiratory Cell Biology and Immunology, Leiden University Medical Centre, the Netherlands (53).

Albumin was used to adjust for the dilution of sputum following induction. Albumin measurements were conducted for 58 pairs of induced/spontaneous sputum samples.

8.2.3 BAL sampling

Preceding bronchoscopy, participants were fasting for at least four hours and pre-treated with salbutamol for a post-bronchodilator spirometry. Sterile PBS was used for all sampling. Each day a new bottle of 500 mL sterile PBS was opened. All participants first delivered an oral wash sample consisting of 10 mL PBS gargled for one minute. Then PBS was used for BAL, storage of sterile brushes, and small-volume lavage samples. It was also used for negative control samples. BAL was not sampled if post-bronchodilator FEV_1 <30% predicted and <1.0 litre. Participants were monitored during bronchoscopy with non-invasive blood pressure, three-lead electrocardiogram, and pulse oximetry. Oxygen was supplied nasally (3 litres/min). Premedication was administered intravenously (Alfentanil 0.25-1.0 mg according to participant preference), and topically (Lidocaine oral spray 10 mg/dose). Bronchoscopy was performed via oral access with patients lying in a supine position. Topical anaesthesia was administered to the bronchi through the bronchoscope. To reduce the risk of upper airway contamination suction was not used until the bronchoscope was below the vocal cords. Furthermore, a sterile wax-plug inner catheter (Plastimed Combicath, France) was used to instil and collect the BAL fluid in two 50 ml fractions from the right middle lobe. The total BAL yield (x mL/100mL) was recorded for each sample. After the procedure, the participants were monitored for two hours and provided with contact information for the ward in case of any adverse effects after discharge.

8.2.4 BAL processing

BAL fluid was immediately processed and partitioned using sterile equipment and processed for 16S rRNA sequencing. For the sequencing analyses, 2 mL aliquots of BAL fluid were stored in ultra-freezers at -80° C until processed in our laboratory.

8.2.5 DNA extraction and 16S rRNA sequencing

The detailed procedure for DNA processing and sequencing of the *MicroCOPD* samples can be found on the open access protocols.io website (54). The sputum samples from *BCCS* and *BCES* were processed using the same procedure.

To ensure bacterial DNA access, sputasol treatment was performed along with an enzymatic pre-lysis step involving Lysostaphin (4000 U/mL), Lysozyme (10 mg/mL), and Mutanolysin (25,000 U/mL) (Sigma-Aldrich). Mechanical and chemical lysis were carried out using the FastPrep-24 instrument and reagents from the FastDNA Spin Kit (MP Biomedicals, LLC, Solon, OH, USA) as the previous steps did not guarantee access to bacterial DNA within whole cells. Centrifugation was used to separate the samples into supernatants and pellets, and only the pellets underwent mechanical and chemical lysis to protect free DNA. The lysates and supernatants were then combined, and the extracted DNA further purified using the FastDNA Spin Kit.

The library preparation and sequencing of the V3/V4 region of the 16S rRNA gene were conducted following the protocol for Metagenomic Sequencing Library preparation for the Illumina MiSeq System (Part # 15044223 Rev. B, MiSeq Reagent Kit v3). Amplicon polymerase chain reaction (PCR) with 45 cycles was performed, followed by index PCR using primers from the Nextera XT Index Kit (Illumina Inc., San Diego. CA, USA). The samples were pooled and normalised for the paired-end sequencing of 2 × 300 base pairs (Figure 3).

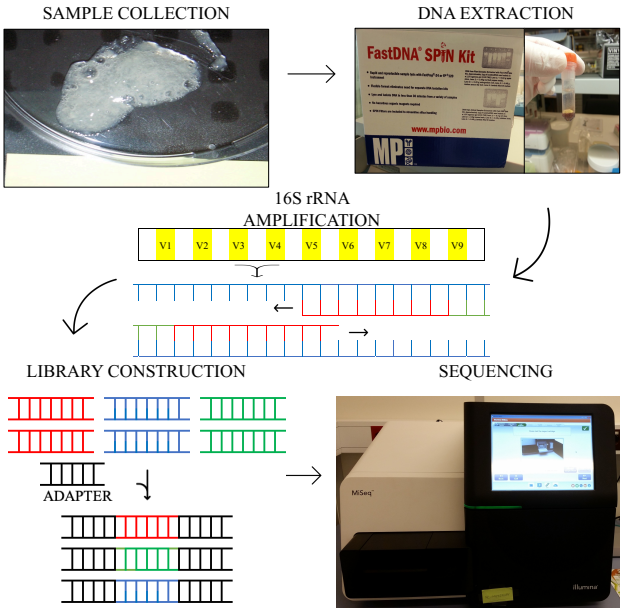

Figure 3: Illustration of processing for 16S rRNA sequencing. DNA extraction from samples in a multiple step procedure, PCR amplification and indexing, and paired-end 16S rRNA amplicon sequencing of the V3/V4 region.

8.2.6 Serum leukocyte counts

In paper III white blood cell counts (WBC) and absolute neutrophil counts (ANC) were reported from the *BCES*. Peripheral venous blood was processed for leukocyte counts at the Haukeland University Hospital clinical laboratory. WBC and ANC were categorised as high or low with cut-offs for *high* at 11.3 10^9/L, and 8.4 10^9/L respectively. These cut-offs represented the upper limit of normal + known analytical variance of 2.7% as given by the laboratory.

8.3. Bioinformatic processing – for those interested

8.3.1 Quantitative Insights Into Microbial Ecology 1 & 2

We chose the open-source bioinformatics pipeline QIIME (35, 55). QIIME 2 succeeded QIIME 1 as of 2018. Both versions of the pipeline make use of numerous

external software packages. Due to the significant advancements in bioinformatic methods over the past few years, there are certain differences between QIIME 1 & 2 (Figure 4).

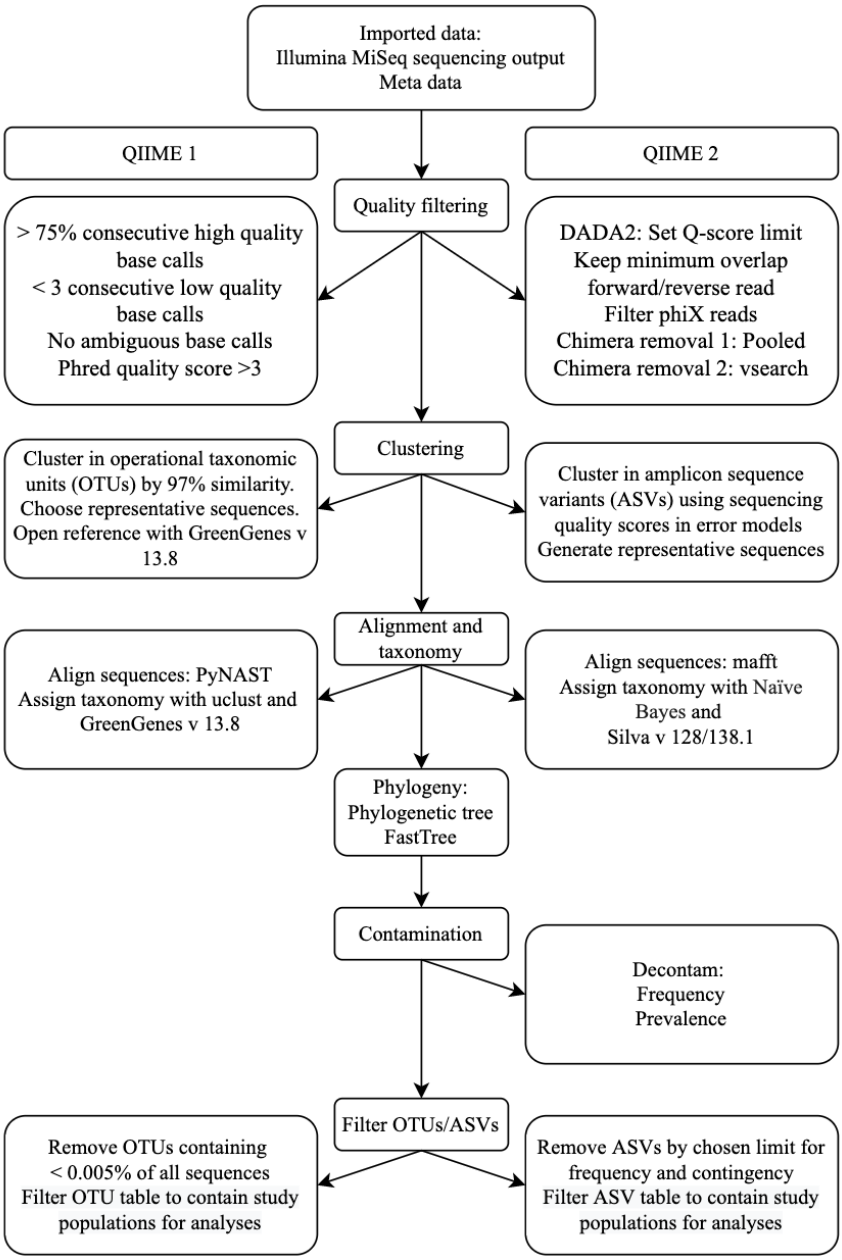

Figure 4: QIIME 1 & 2 workflow for microbiota analyses in the current thesis.

8.3.1.1 Quality filtering

In QIIME 1, we followed a recommended approach to prevent the inclusion of incorrect amplicon sequences in downstream analyses (56). The specific parameter boundaries can be found in Figure 4. For this strategy, we set the quality scores (phred scores) generated by Illumina MiSeq to low values. Chimeras were not removed using QIIME 1.

In QIIME 2, we used DADA2 to denoise and generate amplicon sequence variants (ASVs) (57). To initiate quality filtering, we examined the quality score graphs produced from the Illumina MiSeq phred scores. Forward and reverse reads from the paired-end sequencing runs were evaluated separately. The quality score graphs displayed the phred scores (Q-score) for each base throughout the length of the reads. If the Q-scores consistently dropped below 20, the likelihood of an incorrect base insertion was 1:100, and we trimmed the sequences accordingly. Primers containing ambiguous nucleotides were removed. Our forward and reverse primers were 17 and 21 nucleotides long, respectively. Through trimming, there was a risk of shortening the sequences to the point where the forward and reverse reads did not overlap by at least 20 base pairs, which is necessary for achieving acceptable quality upon joining the reads. Thus, a more stringent quality score cut-off could result in the loss of a significant number of reads. In this thesis, we set the phred score cut-off at 25 for sputum samples and 20 for BAL samples. Additionally, DADA2 discarded reads if the number of expected errors in a forward or reverse read exceeded 2. Finally, DADA2 removed so-called phiX reads, which are sequences originating from a small bacteriophage genome added to the samples for technical purposes.

We applied both DADA2 and VSEARCH (58) to filter out chimeras in our data, as manual examination of the sequences revealed some chimeric sequences remaining after using DADA2. Simply put, a chimera is identified when an amplicon sequence contains segments that are identical to those of other amplicon sequences present in higher quantities. Chimeras are formed during one of the PCR cycles and theoretically get duplicated fewer times than their "parents."

8.3.1.2 Clustering and establishing representative sequences

Amplicon sequences were categorised for analysis. In QIIME 1 amplicon sequences were grouped together based on a 97% similarity threshold, which was believed to correspond to the phylogenetic rank of species (59). These clusters were referred to as operational taxonomic units (OTUs). We used the *uclust* open-reference OTU picking method (55). Our data was compared to the GreenGenes ribosomal reference database, and any sequences resembling the ones found in GreenGenes were included in the clusters. If our sequences were not found in GreenGenes, they were added as de-novo OTUs. However, QIIME 2 suggests using ASVs instead of OTUs (35, 60). DADA2 is a QIIME 2 integrated tool that assign amplicon sequences to ASVs using denoising methods rather than clustering. Before adding our sequences to a given ASV, error models were built based on Illumina quality scores, and used to correct the sequences. As a result, DADA2 allowed for less variability within our ASVs compared with our OTUs in paper II. A single base difference was enough to categorise amplicon sequences into separate ASVs (60).

Certain bioinformatic processes required an input consisting of "*representative sequences*". For OTUs, this meant that each amplicon sequence had to be grouped with identical sequences within the same OTU. The representative sequence was chosen from the largest group of identical amplicon sequences. In contrast, for ASVs, the sequences were expected to be identical, and any single sequence could be chosen as the representative. The processes that required this subset of representative sequences included sequence alignment, taxonomic assignment, and the establishment of phylogenetic relationships between OTUs/ASVs.

8.3.1.3 Alignment and taxonomy

The alignment of amplicon sequences involves organising representative sequences based on their similarities. In QIIME 1, representative sequences were aligned using pyNAST, which is a python implementation of the Nearest Alignment Space Termination (NAST) algorithm. The GreenGenes ribosomal reference database served as a template for this alignment process (61, 62). However, with the introduction of

QIIME 2, pyNAST was replaced by the MAFFT algorithm, which performs a de-novo alignment of sequences based on the fast Fourier transform. Instead of using templates, MAFFT aligns sequences by considering the differences between the representative sequences provided (63).

In QIIME 1, GreenGenes was the recommended reference database for assigning taxonomy to each representative sequence. Taxonomy information was added to the output files during uclust open-reference OTU clustering. However, as GreenGenes updates ceased for many years after a release in 2012, other ribosomal reference databases became preferable. In my case, I chose to use the Silva ribosomal RNA database (64). To obtain the data from Silva, I used two methods. First, I used the QIIME 2 provided taxonomy files, which were pre-processed to only include sequences resulting from V3/V4 sequencing, like my own data. Additionally, in paper IV, I made use of the RESCRIPt tool to collect and modify relevant taxonomic information from Silva (65). In QIIME 2, self-trained Naive Bayes classifiers were generated for taxonomy assignment. The taxonomy classification system follows a hierarchical structure of ranks, as explained earlier. While assigning taxonomy in both QIIME 1 and QIIME 2, it was possible to obtain information ranging from a known phylum to a specific species name. However, if a sequence could not be assigned any taxonomy or was only assigned up to the kingdom level, these sequences were excluded from analysis in the latter two of the three microbiota papers. Further examination of such "unassigned" sequences using the online service BLAST (66), revealed that they originated from humans.

8.3.1.4 Phylogeny

Phylogeny describes the evolutionary connections between OTUs or between ASVs. By aligning the representative sequences, sophisticated models can be created using tools like FastTree (67) to establish the evolutionary distances. In simple terms, a family tree is constructed where the common ancestor is represented by the root and trunk of the tree. Each OTU or ASV is depicted as a leaf or external node on this tree. Intersections of branches signify the internal nodes, representing the most recent

shared ancestor. OTUs/ASVs that are separated by short branch lengths are considered more closely related than those separated by long branch lengths. Using the available phylogenetic information, we can assess diversity not only in terms of the number and abundances of different OTUs/ASVs, but also in relation to their evolutionary connections. The larger the family tree, the greater the diversity.

8.3.1.5 Contamination

A challenge in microbiota research is the ubiquitous nature of bacteria. They are in our laboratory fluids, on our containers and instruments, leading to sample contamination despite our efforts to prevent it. Consequently, the accurate interpretation of our data is affected by the presence of contaminating DNA (68). Contamination was not considered in our QIIME 1 analyses, but we used the online tool Decontam in our QIIME 2 analyses (69). In the case of sputum, we measured the total DNA-load using Quant-iT™ PicoGreen™ from ThermoFisher Scientific Inc, which was available for 50 samples. Based on this data, Decontam was implemented with the frequency option, and we examined the data using various threshold settings to classify contaminants. Ultimately, we determined an appropriate threshold of 0.2. Raising the threshold would increase the algorithm´s identification of contaminants, resulting in a higher risk of false positives. In the case of BAL analyses, negative controls were available, and Decontam was used with the prevalence option with all the *MicroCOPD* samples pooled together. The threshold was set to 0.2.

8.3.1.6 Filter OTUs/ASVs

Bokulich et al. benchmarked quality control and filtering parameters for QIIME 1 and recommended removing OTUs containing <0.005% of all sequences in the data (70). A similar procedure has not been recommended in QIIME 2 where no or limited filtering of the smallest and rarest ASVs are considered sufficient.

8.4 Statistics

Statistical methods used in this thesis are presented in Table 3.

Table 3. Summary of statistical methods used in the current thesis.

Type	Name of analyses	Tool	Description	In paper
Agreement	Bland-Altman	Stata 12.0	Assessing differences between two measuring methods. Bland & Altman plotted 1) Differences between measurements against the mean, and 2) The mean difference to show measurement bias. If the bias = 0, the two methods yield equal measurement. The mean of all differences +/- 2 standard deviations = The 95% limit of agreement (LOA). 95% of data will be contained within the limits.	I & II
Distribution	Shapiro-Wilks	R	Statistical method used to assess whether data distribution is parametric.	III & IV
Comparative	Wilcoxon sign-rank test	Stata 12.0, R	Statistical method for paired and non-parametric data.	I, III & IV
Comparative	Kruskal Wallis test with Dunn´s test	Stata 12.0, R	Statistical method for non-paired, non-parametric data, allows variables with >2 levels. Dunn´s test only run if Kruskal Wallis p-values <0.05. Dunn´s test compares the variable levels pairwise.	I, III & IV
Comparative	Pearson chi-squared test	R	Statistical method for categorical data.	IV
Comparative	Welch two-sample t-test	R	Statistical method for parametric data with <3 level variables.	IV
Comparative	ANOVA with Levene and Tukey test	R	Statistical method for parametric data when variable >2 levels. Levene´s test of homogeneity of variance preceded the two-way ANOVA test. Tukey test if ANOVA p-value <0.05. Tukey tests compare the variable levels pairwise.	IV

Comparative	G test	QIIME 1	Log-likelihood ratio tests with Bonferroni corrected p-values due to multiple comparisons. Assumes normal distribution of data. Must be used with rarefied data.	II
Correlation	Spearman rank correlation test	Stata 12.0, R	Statistical method for non-parametric data.	I
Taxonomic composition, differential abundance	Yue-Clayton dissimilarity	Excel	Assessing proportional similarities of both present and missing taxa across pairs of samples or groups of samples. Yue-Clayton dissimilarities cut-off is set by the researcher. Yue-Clayton dissimilarity ranges from 0-1, where 0 represents perfect agreement.	II & III
Taxonomic composition, differential abundance	ANOVA-Like Differential Expression (ALDEx2)	R	ALDEx2: Centered log ratio (clr) transformation of compositional sequencing data. Zeros handled by estimating possible values for zero followed by significance testing of all estimates. The average significance test value is used for each taxon in each sample. Considers the effect of DNA extraction and sequencing, randomly selecting some but not all present DNA. ALDEx2 with Wilcoxon sign-rank test for paired, non-parametric data. P-values corrected with Benjamin-Hochberg method.	III & IV
Taxonomic composition, differential abundance	Analysis of Compositions of Microbiomes with Bias Correction (ANCOM-BC)	R	ANCOM-BC identifies taxa that are differentially abundant across groups of samples by first estimating the unknown sampling fractions, and then correcting for the bias the unknown sampling fractions introduced by use of a log linear regression model. Zeros are classified according to Kaul´s model from ANCOM-II.	IV
Rarefaction	Alpha-rarefaction	QIIME 1&2	Random subsampling procedure with no replacement in which a set number of sequences are chosen from the data to normalise sample size to avoid bias caused by sequencing depth. The subsampling procedure was repeated x 10.	II, III & IV

Alpha-diversity	observed OTUs; Chao1; Shannon diversity index, Pielous; Faith´s phylogenetic diversity (PD)	QIIME 1&2, R	Diversity metrics calculating diversity within samples. Ranges from the simplest count of different OTUs in samples, to non-phylogenetic indexes considering evenness and richness, to phylogenetic diversity using the sum of branch lengths in a rooted phylogenetic tree. Details in paper IV, Supplemental material.	II, III & IV
Alpha-diversity	Two-sample t-test	QIIME 1	QIIME 1: Non-parametric alpha-diversity measures analysed with a two-sample t-test with 1000 Monte Carlo permutations in *compare_alpha_diversity.py*.	II
Beta-diversity	Sørensen, Bray-Curtis; unweighted and weighted UniFrac, Aitchison	QIIME 1&2	Diversity measures calculating diversity between samples. The metrics can be divided with regards to being qualitative or quantitative, and with regards to being or not being phylogenetic. Aitchison takes compositionality into account. Details in paper IV, Supplemental material.	II, III & IV
Beta-diversity	Principal coordinates analyses (PCoA), non-metric multidimensional scaling (NMDS)	QIIME 1, R	Visualisation of distances between samples from compared groups. PCoA infers the distances between samples in a low-dimensional, Euclidean space. Dimensions is given by the inputted data (n-1). Has only 1 solution. NMDS visualises ranks of distances between samples in a low-dimensional, non-Euclidean space. Number of dimensions are set by the researcher. NMDS is iterative and has multiple solutions.	II
Beta-diversity	Procrustes analyses	R	Procrustes symmetric rotation and scaling opts to minimise the distance between paired samples from corresponding PCoA or NMDS plots. The sum of the squares of calculated distances (Procrustes M^2), represents the best fit between the matrices, ranging from 0-1. M^2 = 0: Identical communities. Symmetric Procrustes analyses are permuted using PROTEST in R package Vegan to estimate significance of the Procrustes statistics.	II & III

Beta-diversity	ANOSIM	QIIME 1	QIIME 1: Analysis of similarities (ANOSIM), a non-parametric method applied to distance matrices determines significance after permutation analyses. Results range from -1 to +1. Negative results if samples from different groups resemble each other more than samples from the same group.	II
Beta-diversity	Weighted UniFrac significance test	QIIME 1	QIIME 1: Pair-wise beta-diversity differences with UniFrac Monte Carlo significance tests. Samples are randomly re-assigned in a phylogenetic tree before distances between two and two samples are calculated for each random dataset. Bonferroni corrected p-values calculated from the fraction of the time that UniFrac distances in the inputted dataset is smaller than in the random datasets.	II

9. Summary of papers

9.1 Comparison of inflammatory markers in induced and spontaneous sputum in a cohort of COPD patients

Airway inflammation is commonly assessed by measuring concentrations of inflammatory markers in sputum. The effects sputum sampling methods may have on measured levels of inflammatory markers are understudied. We investigated for differences in concentrations of six inflammatory markers based on sputum sampling method after albumin correction of concentrations. In addition, we assessed the safety of sputum induction in exacerbating patients with COPD.

Patients were included from the *Bergen COPD cohort study* (*BCCS*) and the *Bergen COPD exacerbation study* (*BCES*) running from 2006-2010. All patients were diagnosed and classified with COPD according to the GOLD guidelines. A routine follow-up was scheduled every six months, and patients experiencing exacerbations came in for additional consultations. For spontaneously expectorating patients spontaneous sputum sampling preceded inhalation of nebulised 3% or 0.9% saline inducing patients to expectorate. Forty-five patients had delivered sputum samples by both methods at 60 consultations. Thus, 60 sputum pairs were available for this study. IL-6, IL-8, IL-18, IP-10, TNF-α, and MIG were measured by bead based multiplex immunoassay. Albumin was measured by enzyme immunoassay to allow concentration correction. Culturing for bacterial growth was performed on 24 samples. Stata was used for statistical analyses.

The study showed fair correlation between the levels of markers in induced and spontaneous sputum with correlation coefficients between 0.58 (IL-18) and 0.83 (IP-10). Meanwhile, the Bland-Altman limits of agreements between the two sampling methods were low for all six markers. An observation of higher TNF-α during exacerbations ($p = 0.002$) and trending higher at the steady state ($p = 0.06$) was made only in spontaneous sputum. Similarly, only in spontaneous sputum were IL-18 and MIG significantly higher in ex-smokers ($p < 0.05$), and IL-6 significantly lower with positive *Haemophilus influenzae* (HI) cultures. Induction with saline during

exacerbations did not cause a significant decline in FEV_1 regardless of COPD severity.

From this study we recommend that sampling method is given careful consideration when conducting and comparing studies on inflammatory markers in sputum. We found induction with hypertonic saline to be a safe sampling procedure during exacerbations, also in patients with very severe disease.

9.2 Comparing microbiota profiles in induced and spontaneous sputum samples in COPD patients

As with airway inflammation, the airway microbiota is commonly assessed in sputum. Whether sputum sampling methods influence on the composition of the airway microbiota is understudied. We compared the bacterial microbiota in sputum pairs consisting of induced and spontaneous sputum from patients diagnosed with COPD.

In this study, patients were included from the *BCCS* and *BCES* if we had both spontaneous and induced sputum from the same consultation. Thirty-six paired sputum samples from 30 patients were available for DNA sequencing from our biobank. DNA was extracted by enzymatic and mechanical lysis methods. The V3/V4 region of the 16S rRNA gene was PCR-amplified and prepared for paired-end sequencing with the Illumina MiSeq System. QIIME 1 and Stata were used for bioinformatics and statistical analyses.

The study showed that diversity was not associated with sampling method if comparing all induced with all spontaneous sputum samples. Looking at individual patients, nine had significant differences in beta-diversity linked to sampling method according to weighted UniFrac (WUF) significance tests ($p < 0.01$). OTUs were assigned taxonomy, and the composition and abundances of bacterial genera were investigated. Pair-wise comparisons of composition with Yue-Clayton dissimilarity and Bland-Altman agreement analyses showed considerable differences in several

patients (Yue-Clayton >0.2 for seven patients, Bland-Altman limit of agreement >0.1 for 13 patients). Fifteen genera were differentially abundant across sample types collected at stable state and/or at exacerbations (log-likelihood ratio tests $p < 0.05$). Among the genera with differential abundance were potential pathogens like *Haemophilus* and *Moraxella*.

From this study we recommend that sputum sampling methods are standardised upon conducting studies on the sputum microbiota to avoid sampling method biases.

9.3 Sputum microbiota and inflammation at stable state and during exacerbations in a cohort of chronic obstructive pulmonary disease (COPD) patients

Exacerbations of COPD are often attributed to infections. The dynamics of the airway microbiota in exacerbations are not well described. We investigated for differences in the microbiota and immune responses in induced sputum samples collected from patients with COPD at stable state and during exacerbations.

In this study, patients were included from the *BCCS* and *BCES* if we were able to obtain measures of inflammatory markers, and biological material for DNA sequencing from induced sputum. Thirty-six paired sputum samples from 36 patients were included. In addition, one patient who delivered sputum on 13 different occasions during the three-year study period were included for a longitudinal case study. DNA was extracted by enzymatic and mechanical lysis methods. The V3/V4 region of the 16S rRNA gene was PCR-amplified and prepared for paired-end sequencing with the Illumina MiSeq System. Sputum inflammatory markers (IL-6, IL-8, IL-18, IP-10, MIG, TNF-α) and AMPs (LL-37/hCAP-18, SLPI) were measured in supernatants, whereas target gene sequencing (16S rRNA) was performed on corresponding cell pellets. Serum leukocyte counts were performed at the Haukeland University Hospital clinical laboratory. The bioinformatic pipeline QIIME 2 and the statistics environment R were used to analyse the data.

The study showed that the sputum microbiota and inflammatory markers in induced sputum differed from stable states to exacerbations. Changes in the microbiota were more apparent within individual patients with COPD, than for the patient cohort as a whole. Except for SLPI, all measures of inflammation were higher during exacerbations than at stable state (significantly for IP-10, MIG, TNF-α, SLPI and LL-37, $p < 0.05$). We observed significant changes in taxonomic composition when examining individuals (Yue-Clayton >0.2), rather than all patients at stable state against all patients during exacerbations (ALDEx2, $p > 0.05$). The seven patients with high serum ANC during exacerbations had significantly higher Shannon diversity index in stable state compared with patients with non-elevated serum ANC. The bacterial composition in the case study spanning over 13 stable state/exacerbation visits was highly dynamic and without a consistent stable state equilibrium.

From this study we recommend that longitudinal airway microbiota studies include analyses of inter-individual differences as well as group analyses to better assess the microbiota. Further, our study indicates that establishing guidelines for treatment of microbiota dysbiosis can be challenging due to the inter- and intra-individual differences observed.

9.4 The lower-airway microbiota in COPD and healthy patients

The bacterial microbiota in the lower airways has been linked to COPD, smoking, and for patients with COPD, the use of ICS. We investigated for possible differences in the microbiota in protected BAL across all these factors in a large cohort of patients with COPD and healthy controls.

Participants were included from the *MicroCOPD* study if we had valid DNA sequencing results from BAL available. Ninety-seven patients with COPD and 97 controls were included. Participant characteristics were obtained through standardised questionaries and clinical measurements between 2012-2015. DNA was extracted by enzymatic and mechanical lysis methods. The V3/V4 region of the 16S rRNA gene

was PCR-amplified and prepared for paired-end sequencing with the Illumina MiSeq System. The bioinformatic pipeline QIIME 2 and the statistics environment R were used to analyse the data.

The study showed lower alpha-diversity in COPD compared with controls. The difference was explained by a loss of evenness rather than a loss of richness (Pielou evenness p = 0.004, Shannon diversity index p = 0.01, Observed ASVs, and Faith´s PD p >0.05). Comparing beta-diversity in BAL from smoking and non-smoking patients with COPD revealed a significant difference only when abundances and phylogenetic information both were considered with weighted UniFrac (permutational multivariate analysis of variance (PERMANOVA) R^2 = 0.04, p = 0.01). Differential abundance of taxa assessed with Analysis of compositions of microbiomes with bias correction (ANCOM-BC) was found for nine genera. The three genera enriched in COPD all belonged to the *Firmicutes* phylum. Among these were *Streptococcus*, a potential pathogenic bacterium. Smoking quantified by pack years was associated with a significant reduction in *Haemophilus* and *Lachnoanarobaculum* in healthy controls only. The only genera associated with smoking in the COPD cohort was *Oribacterium* with lesser abundances in patients still smoking. Neither diversity nor taxonomic abundances differed in BAL fluid with the use of ICS or with increased COPD severity.

From this study we conclude that differences observed in the lower airway microbiota in patients with COPD compared with controls seem independent of smoking, and unaffected by use of ICS among patients with COPD. We recommend conducting longitudinal studies to assess causality between the development of COPD and the development of a dysbiosis of the lower airway microbiota.

10. Discussion of methods

10.1 Study design and study populations

BCCS, *BCES* and *MicroCOPD* were all longitudinal, observational COPD cohort studies with control groups enrolled. Paper I-III can best be described as observational

cohort studies with a cross-sectional, and a “crossover” design. Paper III also include a longitudinal case report. Paper IV was an observational case-control cohort study with a cross-sectional design.

While the original studies included large numbers of participants and samples, the number of available sputum samples for my analyses were small. The reason for this is that the microbiota research was initiated in 2012, two years after the completion of the sampling period. Samples that were available in our biobank was thus not collected nor stored for the purpose of answering the research questions in this thesis. Meanwhile, when establishing the *MicroCOPD* study it became clear that the stored sputum cell pellets could be of value, and it was sought to analyse as many as possible. Overall, the success rate of sputum induction was around 50%, and in addition, quality assessments excluded more samples. Further, for my analyses I needed pairs of sputum samples, and thus the number of valid sample pairs were reduced accordingly.

A lack of statistical power is a shared shortcoming for all three sputum studies as the number of samples are small. It follows that the larger BAL study should be expected to have higher statistical power. Sample size calculations were not performed preceding enrolment for the *MicroCOPD* study. The benefits of calculating sample sizes needed to answer research questions are known, but established methods for microbiota research seemed to lack in 2012. Ferdous et al. proposed a comprehensive and complex set of methods taking the different elements in microbiota research into account in 2022 (71). Hopefully, such calculations will be easier to implement in the future. However, an observational study such as *MicroCOPD* has several aims, and thus several potential “ideal” sample sizes depending on which aim one examines. Thus, the sample size for the *MicroCOPD* was also determined by what was possible to collect within a reasonable timeframe.

My sputum studies were designed such that confounding was reduced. For paper I and II all participants were represented both in the induced and spontaneous sputum groups eliminating confounding from static variables like sex. Also, the samples were

collected the same day preventing the effect of passed-time which otherwise could influence on for example accumulated pack years, COPD severity and age. Stratification was challenging for my small sputum studies as sub-groups would become very small. For the larger BAL study stratification was used for variables like age, sex, and smoking.

10.2 Airway sampling and sample processing

When conducting research on living human participants, it is not feasible to obtain entire lungs for investigation. Therefore, we must use alternatives such as sputum and BAL to represent the overall lower airway inflammation and microbiota. However, it is important to note that this simplified approach cannot fully capture the complexity of these entities along the entire bronchial tree, and lung.

Sputum and BAL differ significantly, not only in terms of the invasiveness of the sampling procedures but also in terms of potential oral contamination. While oral contamination could not be avoided in the case of sputum, measures were taken during bronchoscopy to minimise oral contamination of BAL. These measures likely reduced contamination significantly (26).

Determining the best sample type for predicting clinical outcomes is challenging. Intuitively, one would think that lower airway samples verified through bronchoscopy, and with minimal contamination from the upper airways would be preferable to sputum samples. However, in clinical practice, sputum samples are far more readily obtainable. Should we analyse and examine the lower airways using samples that are practical in a clinical setting rather than ones that are limited to research protocols? Or will important pathogenetic signals be lost in the presence of abundant oral microbiota? These are questions that undoubtedly need further investigation.

In terms of the validity of my data, the first factor to consider is selection biases. All participants included in the comparison of induced and spontaneous sputum had to be

able to expectorate. Therefore, observations of inflammation and microbiota may not be representative of all the patients with COPD who do not expectorate. The external validity of the actual measurements may be weakened by the selection criteria. The same issue arises in the third study, as the requirement for presence of exacerbations may render the measurements not applicable to the general COPD population. Selection biases in the BAL study (Under peer-review at the present time, available in the supplement of this thesis) were based on safety precautions that eliminated participants with certain features (48).

In my studies, as in so many others, internal validity may have been compromised with each step of the sample processing, despite following stringent protocols. It is also important to keep in mind that my studies aimed to determine whether sampling methods affected measurements, rather than what the measurements were in sputum from patients with COPD.

Lastly, it is worth mentioning that all samples in all four studies were taken as single samples. Ideally, we would have divided the original sputum samples into several fractions and analysed them in parallel, but this was not possible. For the BAL study, we discussed the possibility of amplifying and sequencing the extracted DNA in triplets, but tripling the processing expenses was not feasible. This limitation affects our ability to discuss whether variation in measurements resulted from processing, natural variability, or the variables of interest for analyses.

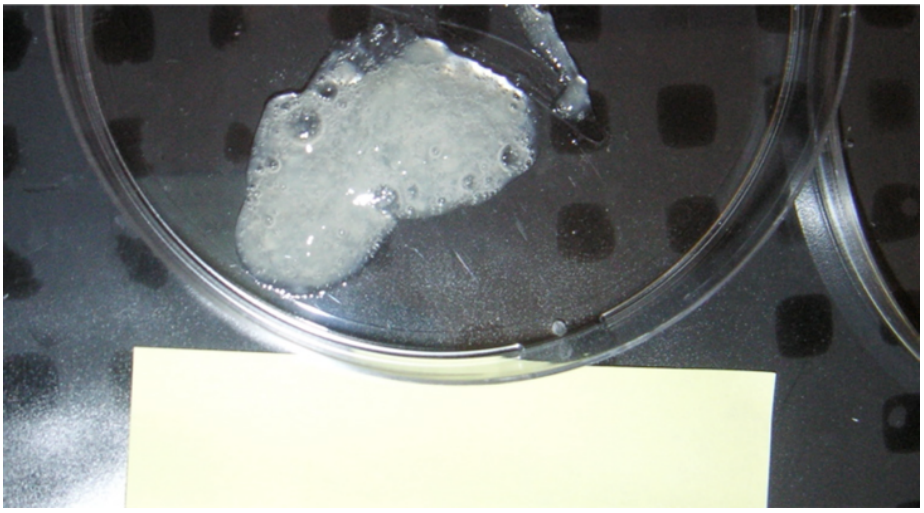

Illustration 5. Freshly delivered induced sputum sample, potentially diluted, and polluted by nebulised saline in a potentially contaminated petri dish, ready for processing with contaminated laboratory fluids, potentially affecting inflammatory marker concentrations.

10.2.1 Inflammation.

The impact of using PBS and DTT on recovery of cytokines has been demonstrated (72, 73). For paper I all samples were processed according to the same protocol, and the effect of chemicals on our measurements should be similar regardless of sputum sampling methods. However, the use of these chemicals should be kept in mind if comparing our data with other studies processing samples with different chemicals. The inflammatory markers and AMPs were measured in simplex due to cost, which possibly introduced measurement errors greater than if measured in duplex.

To address the potential diluting effect of saline inhalation (74), the ratio of albumin between induced and spontaneous samples was used as a correction factor as suggested by a reviewer for the journal (Respiratory Research). Applying this method affected the interpretation of three inflammatory markers. IL-18 measured during exacerbations was significantly lower in induced than spontaneous sputum before albumin correction, but not after. TNF-α and MIG were both significantly lower in induced than spontaneous sputum at stable state before correction, but not after. It

should be noted that this concentration correction method has not been validated, but in accordance with the journal request, adjusted values were reported in our publication.

Measurements of AMPs were carried out by colleague Persson (47). In her paper she discussed how sputum composition can influence on immunoassay-based measurements of cathelicidin. We did not examine AMPs across sputum types, but of course the composition of induced sputum might vary and thus influence AMPs measurements. We had no means to investigate for this comparing sputum AMPs across disease states.

As mentioned earlier, it is difficult to determine which sample type generally has the greatest internal validity. However, the induction method allows for the inclusion of individuals who do not spontaneously produce sputum, which may improve the external validity of airways studies. This also allows for larger sample sizes, leading to enhanced statistical power and enabling meaningful stratification.

10.2.2 Microbiota

To obtain data for the microbiota analysis we started with the isolation of DNA from the samples. However, DNA extraction may decrease internal validity due to its varying efficiency in extracting DNA from different types of bacteria. To mitigate this issue, it is imperative to ensure that the process aligns with the diversity in bacterial wall structures. Hence, we initiated the procedure by following the recommended enzyme step as described (75). The removal of DNA in supernatants before bead beating and chemical lysis was carried out to minimise any potential shearing or distortion of DNA. This step was deemed significant because the low biomass airway samples necessitated a high number of PCR cycles to obtain a meaningful signal. It was hypothesised that sheared and distorted DNA, amplified over numerous PCR cycles, could lead to an increase in chimeric sequences. The DNA extraction process consisted of several steps and involved laboratory solutions containing significant

amounts of contaminating bacterial DNA, which could also interfere with the interpretation of the microbiota (68). Further discussion on contamination will be provided later.

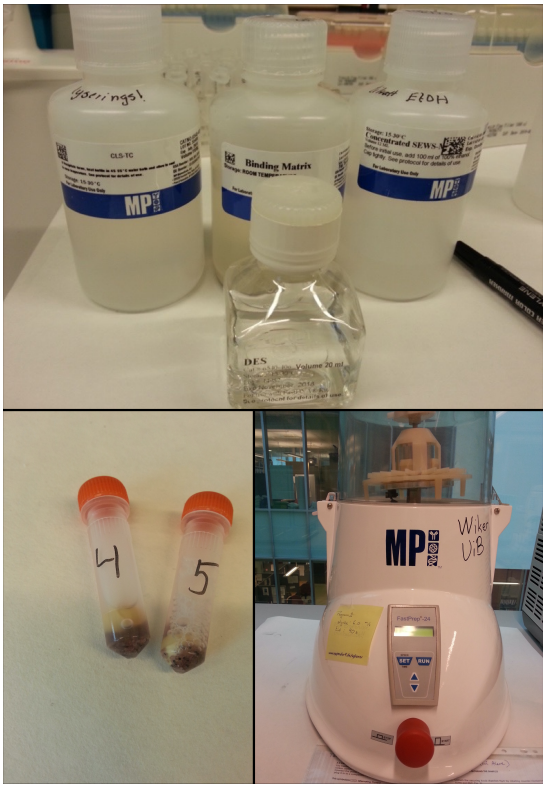

Illustration 6. Top picture: DNA-extraction laboratory fluids, the habitat of contaminants. Bottom pictures: Bead-beating equipment, the likely shearer of DNA.

DNA extraction does not discriminate between human and bacterial DNA. This is first possible in polymerase chain reactions (PCR). Primers designed to adhere to the bacterial 16S rRNA gene ensured that bacterial DNA was multiplied in each PCR cycle. The choice of primers has been shown to impact the results in microbiota studies (76). One study on faecal microbiota and mock communities points to primers designed for the V3/V4 regions of the 16SrRNA gene to perform better than other primers (77). In our own laboratory Drengenes compared V3/V4 sequencing against

V4 sequencing with different workflows and showed that V3/V4 resulted in more ASVs, and the best taxonomic resolution (76). Index PCR in which primers are attached to the amplicon primers allow for further, though not perfect discrimination between human and bacterial DNA.

To obtain an adequate amount of sequencing material from low biomass samples such as BAL, and to some extent sputum, it is necessary to increase the number of PCR cycles. However, this comes at the expense of a potential amplification of contaminating and/or chimeric sequences (78-80). These errors are likely to have diminished the internal validity of my data. Moreover, studies with high biomass samples using fewer PCR cycles are less likely to be impacted by the variable numbers of the 16S rRNA gene present in bacteria compared to studies using a high number of PCR cycles.

Sequencing of the PCR products represented the last step in which biological material was processed. Errors from sequencing includes erroneous insertion of bases. The median sequencing error rate for the Illumina MiSeq has been estimated by Stoler et al. to be 0.47% (81). Bioinformatical tools dealing with such sequencing errors, chimeric sequences, and contaminants have been introduced and improved continuously during the decade I have researched microbiota. So has my knowledge about both the issues and possible solutions.

Measurement differences can exist between samples processed in different batches, resulting in what is known as a *batch effect*. To mitigate this effect, we processed samples collected from the same participants together. Bar charts were used to visualise the relative abundances of various genera, and paired samples were compared. A notable dissimilarity was observed between the induced and spontaneous sputum samples from seven participant, which raised concerns about the processing of these samples. To investigate further, these seven pairs, along with three pairs that appeared similar, underwent PCR and sequencing twice to determine if the results would be consistent. The second run yielded consistent results for all but one pair, and

19 out of the 20 samples. However, due to the small number of samples, this investigation into variability caused by PCR/sequencing has limited value.

10.3 Statistical and bioinformatic analyses

Starting in 2013, and wrapping up 10 years later, this thesis hopefully illustrates not only the evolution of bioinformatic tools for microbiota research, but also an evolution of personal knowledge. Statistical methods used in this thesis will be discussed for each paper separately.

10.3.1 Comparison of inflammatory markers in induced and spontaneous sputum in a cohort of COPD patients

10.3.1.1 General considerations

Data distribution was assessed visually from bar-charts for all analyses, and statistical methods chosen accordingly.

10.3.1.2 Agreement

I examined the correlation between inflammatory markers sampled by induction and by spontaneous expectoration and observed a fair correlation. Bland and Altman criticised the use of correlation in such comparisons, and proposed that measuring methods should yield similar rather than just correlating results (82). To address this concern, they introduced *The Bland-Altman plot* and *limits of agreement* (LOA) (Table 3). Log transformation of my non-parametric data was recommended before plotting, with the upper and lower limits of agreement interpreted after calculating the antilogs (83). Whether the limits are acceptable or not is decided by the researcher and must be contextualised and evaluated before the analyses are run. For instance, while a variability of +/- 50 g may be acceptable when weighing an average person, it would not be appropriate when weighing gold. Therefore, the choice of a cut-off for inferring differences must be based on existing knowledge and the researcher´s discretion.

Regrettably, paper I lacks a pre-discussion of LOA due to my unfamiliarity with this method and its significance in setting a cut-off prior to calculating the results.

10.3.2 Comparing microbiota profiles in induced and spontaneous sputum samples in COPD patients

10.3.2.1 General considerations

Visual assessment of bar-charts was employed to evaluate the data distribution for all analyses, and statistical methods were selected accordingly, except for one situation that will be elaborated on later. The paired study design represented a challenge in finding appropriate statistical methods.

QIIME 1 was chosen for analyses primarily due to the availability of courses designed for researchers venturing into microbiota research, as well as the compatibility of our data with the pipeline. Lacking proficiency in bioinformatics, I relied on the tools provided by QIIME 1 for data analysis, while employing Stata and worksheets as alternatives.

10.3.2.2 Chimera and contamination

In QIIME 1 the possibility of removing chimera was acknowledged, but concerns were raised about its potential impact on alpha-diversity estimates (84). While attending our first QIIME course held in New York, USA in 2015, our research group was advised against removing chimera. Instead, it was emphasised that the implementation of benchmarked filtering steps was sufficient (56). Furthermore, the issue of contamination was disregarded in paper II. I lacked negative control samples containing saline used for induction nor did I consider processing a sterile PBS sample along with the biological samples to check for laboratory contamination. The latter has been confirmed problematic by my fellow researcher Drengenes (68). Consequently, this represents a clear shortcoming. However, the confounding from potential contamination in the laboratory was reduced by treating all samples identically, and all pairs with the same laboratory kits.

10.3.2.3 Diversity – measures of that within and that between

In paper II, rarefaction was employed to normalise samples before conducting diversity analyses (Table 3). To ensure that no samples were lost, we set the rarefaction limit equal to the number of sequences in the least rich sample. Rarefaction removed 2.67 million sequences, reducing the number of OTUs in the dataset only by two, from 1004 to 1002. Rarefaction may have heightened the discrepancy between the diversity in the actual airway microbiota and the diversity analysed statistically.

For paper II we chose three alpha-diversity metrics provided as default by QIIME 1 (84). Of these, *observed OTUs* were the simplest alpha-diversity metric, Chao1 attempted to estimate what should have been present in the samples but were not due to incomplete sampling (85), and Faith´s PD implemented phylogenetics. With this choice of alpha-diversity metrics we lacked information about the evenness of OTUs. In hindsight Shannon diversity index or Pielou evenness could have been included for a fuller picture of the alpha-diversity. For statistical analyses the integrated solution in QIIME 1 was a non-parametric two-sample t-test which was chosen (Table 3). In retrospect, a better solution would have been to export the alpha-diversity metrics from QIIME 1 and to use the Wilcoxon signed rank test based on the *paired*, non-parametric nature of the data.

Beta-diversity could also be calculated with different methods in QIIME 1, but UniFrac was recommended (55). UniFrac integrate phylogenetic information from a phylogenetic tree, and can be both qualitative and quantitative (86). For paper II we restricted beta-diversity analyses to UniFrac, thus we lacked information on sputum sampling methods effects on non-phylogenetic beta-diversity. For statistical analyses we ran *beta_significance.py* with weighted UniFrac test (Table 3) This was the only available test for pairwise comparisons of UniFrac that I could apply at the time, but it later became obsolete as the datasets grew. Its limitation can be illustrated with our 36 pairs of sputum, resulting in a staggering 2556 comparisons, of which only 36 were of

interest as they pertained to the paired samples. Bonferroni correction of p-values was used to counter the multiple testing issue.

The Procrustes transformation of principal coordinates analysis (PCoA) plots continues to be a useful visual method for comparing pairwise beta-diversity (Table 3). These plots are advised to be considered appropriate only when the sum of variation explained by the first three axes >50% (87). Like the Bland-Altman limits of agreement, the Procrustes transformation lacks a predetermined threshold for the outputted M^2, making the reporting and interpretation of results more complex. Consequently, it was easier to draw conclusions from the analysis of similarities (ANOSIM) between induced and spontaneous sputum, as ANOSIM in QIIME 1 provided p-values with a significance level of $p < 0.05$. However, it is important to note that the ANOSIM test in QIIME 1 were not designed to handle paired data.

10.3.2.4 Taxonomy

QIIME 1 made several online taxonomic databases available, but I opted for the recommended GreenGenes (61, 84). Selecting a different database could have resulted in different assigned taxonomy (88). This discrepancy in sequence classification across databases still has the potential to impact comparisons between research papers. While taxonomy was assigned down to the species level, limitations in the 16S rRNA gene sequencing prevented us from investigating more detailed taxonomic ranks beyond genera (39). Some OTUs were not assigned taxonomy, and for some, only higher taxonomic ranks such as family, class, and order were assigned. When revisiting the GreenGenes taxonomic assignment from 2015, I discovered that 6% of sequences lacked taxonomic information beyond the kingdom rank. These sequences were not given special attention and were analysed as *Unassigned;Other;Other;Other;Other;Other* (5.7% of sequences in induced sputum, 6.1% of sequences in spontaneous sputum) and *k__Bacteria;Other;Other;Other;Other;Other* (0.01% of sequences in both induced and spontaneous sputum). It later became apparent that these sequences likely originated from human DNA.

To evaluate the effect of sputum sampling methods on taxonomy, Bland-Altman analyses were used, as described in paper I. In paper II, I calculated the range of limits of agreement (LOA) to facilitate interpretation. It became clear that an a-priori cut-off needed to be established, and this was discussed within our study group. The chosen limit allowed for a 10% difference in the relative abundance of taxa between induced and spontaneous samples. We recognised that larger differences indicated greater variability than what we expected would arise from processing and natural variation. Inspired by Bassis et al. (89) I also employed Yue-Clayton dissimilarities (90) to express beta-diversity calculated on OTU/ASV data when comparing groups. We found this method helpful in quantifying differences in the relative abundances of phyla and genera within individual sample pairs. Establishing a cut-off for what could be considered different using Yue-Clayton dissimilarities was a challenge, and it was discussed with both Bassis and within our research group. This situation exemplifies a scenario where running fractioned samples through the laboratory would have been valuable. Instead, our solution was to have three members of the research group independently classify taxonomic bar plots as visually similar or dissimilar. We determined that the Yue-Clayton dissimilarity cut-off needed to align with the a-priori visual classification. A cut-off at 0.2 proved sufficient in this regard, although it remained somewhat arbitrary. Additionally, we were unable to create comprehensive plots including all 106 taxa, so the cut-off was established in relation to the 11 visualised taxa containing at least 1% of all sequences. Extrapolating the two within-pair similarity estimates to other studies and populations is therefore challenging.

To assess the differential abundance of taxa between all induced and spontaneous samples, we performed rarefaction using the log-likelihood ratio test provided in the QIIME 1 command *group_significance* (Table 3). We did this although our data had a non-parametric distribution, and the statistical test assumed normality. In fact, QIIME 1 did not offer any statistical methods accepting paired, non-parametric data. The development and improvement of tests for differential abundance are ongoing, both within the QIIME pipeline and in available packages in R. Consequently, the analyses of taxonomic differences in paper II are now somewhat outdated.

10.3.3 Sputum microbiota and inflammation at stable state and during exacerbations in a cohort of chronic obstructive pulmonary disease (COPD) patients

10.3.3.1 General considerations

After completion of numerous statistics courses, statistical assessment of data distribution was conducted using the Shapiro-Wilks test for all analyses in paper III. The selected statistical analyses for the evaluation of inflammatory markers are assumed to be well-known and will not be discussed further (Table 3).

In January 2019, QIIME 1 was replaced by QIIME 2, which provides a more comprehensive bioinformatic and statistical toolbox. Additionally, I attended an R class by Professor Pat Scloss (*Crashcourse in R Workshop for Microbial Ecologists*) in 2016, giving me the opportunity to analyse data for paper III in R. However, the paired study design posed a significant challenge in identifying suitable statistical methods and analyses for the microbiota data.

10.3.3.2 Chimera and contamination

During the time when paper III was being worked on, the recommendations in QIIME 2 were to eliminate both chimeric sequences and contaminants as outlined in the Methods. It was generally believed that DADA2 was effective in removing chimeras, but our manual curation in the laboratory revealed otherwise. Additionally, Drengenes highlighted the potential for chimeric sequences due to shearing of DNA during bead beating and the high number of PCR cycles in our protocol (91). While it is possible that the heightened elimination of chimeras could have affected the data by excluding legitimate biological sequences, I maintain the belief that it ultimately improved the internal validity of my findings.

In 2018, the introduction of the R package *Decontam* provided us with a tool to identify contamination in our data, despite the absence of negative controls (68, 69). We had picogreen measurements of the total DNA-load in 50 samples and applied the frequency option in *Decontam* on these samples. The analyses likely enhanced the

internal validity though 22 samples could not contribute to the identification of contaminants.

For QIIME 1, Bokulich et al. conducted a benchmark study on a quality filtering procedure (56), but this has not yet been done for QIIME 2. Therefore, the question of whether ASVs should be filtered based on abundances and contingency was discussed among our research group and with the QIIME 2 developers. Considering the high number of PCR cycles, I made the decision to eliminate ASVs that had less than 10 sequences across the 72 samples. Additionally, I removed sequences that were not found in at least five samples, which corresponds to three participants due to the paired design. The features filtered out in this process were extremely rare and low abundant, making it difficult to attribute biological meaning to any of them. As a result of this filtering, the number of ASVs decreased from 1439 to 408, with an average number of sequences per removed ASV of 197. This represented only 0.007% of the total number of sequences. It is important to acknowledge that this filtering did potentially impact the richness and differential abundance analyses, but I did not investigate to which extent it affected the data.

10.3.3.3 Diversity – measures of that within and that between

The developer´s recommendation for rarefaction in diversity analyses remained unchanged after the transmission to QIIME 2. To avoid losing samples, I set the rarefaction limit to match the sample with the lowest number of reads.

For paper III, I computed Faith´s PD and Shannon diversity index, the later including consideration of the evenness of features. These analyses were conducted using the Vegan package in R, and the Wilcoxon signed rank test was used to handle both paired and non-parametric data, which was an improvement from our previous paper.

The evaluation of beta-diversity was expanded to include calculations of both phylogenetic and non-phylogenetic matrices, as well as Aitchison distances that treated microbiota data as compositional (34). The introduction of Aitchison distances was directly linked to the availability of differential abundance analyses, which will

be discussed later (34). Additionally, the use of the Vegan package allowed me to perform PERMANOVA for both paired and non-parametric data, without violating the assumptions of the statistical test (92). As a result, the analysis of beta-diversity was also enhanced compared to our previous paper.

After learning more about ordination plots, I decided to replace principal coordinate analysis (PCoA) with non-metric, multidimensional scaling plots (NMDS). This decision was made because I calculated five different distance matrices that were unlikely to satisfy the assumption of linearity in PCoA. Furthermore, the sum of variation explained by the first three axes in PCoA was less than 50% for both Bray-Curtis and unweighted UniFrac distances generated in QIIME 2. By using NMDS, I forced the data into two dimensions without considering stress, which is a measure of how well the data is represented by the plot. Unfortunately, at that time, I had not yet realised that ordinating the data in three dimensions would result in a more acceptable level of stress (87, 92). Procrustes transformation of NMDS ordinations were performed as described for PCoA in paper II.

10.3.3.4 Taxonomy

QIIME 2 could make use of several taxonomic databases, and I chose to use Silva for paper III as GreenGenes had not been updated since 2012. After conducting a BLAST investigation on the ASVs that were lacking taxonomic annotation, I removed these ASVs before conducting any statistical analyses. This step likely contributed to enhancing the internal validity of the data in paper III.

To assess the impact of disease state on taxonomy, I employed Yue-Clayton dissimilarities for each participant. I chose to maintain the cut-off value from paper II. The samples used in both papers underwent the same processing procedures, with 28 samples being included in both studies.

When it came to differential abundance analyses for non-parametric data from a paired study design, my options were limited to ANCOM and ALDEx2. The statistical method incorporated into ANCOM for paired data was Friedeman´s test,

which assumes a minimum of three repeated samples per participant. ALDEx2 was the only test I could find that did not have its assumptions violated by my data. In addition to analysing the abundances of phyla and genera, I used ALDEx2 for ASVs. In comparison to the methods employed in QIIME 1, I consider ALDEx2 to be an improvement.

10.3.4 The lower airway microbiota in COPD and healthy controls

10.3.4.1 General considerations

The data for paper IV is a sub-selection from the greater *MicroCOPD* study. Several studies have been published by my fellow researchers describing methods and statistical challenges in depth (26, 28, 48, 54, 68, 76, 93, 94). I limited the data to BAL samples after having discussed the option of also including oral wash collected directly prior to the bronchoscopic procedure. This would have allowed me to compare the oral microbiota with the lower airway microbiota. We did find it necessary to limit the study to only one sample type to answer the main research question within the word limits set by the publisher.

10.3.4.2 Chimera and contamination

The process of removing chimeric sequences from all 2448 samples in the *MicroCOPD* study was performed in the same manner as described in paper III. However, for the *MicroCOPD* data, *Decontam* was used with the prevalence-based approach since we had negative control samples available (69). As a result, I can confidently conclude that the handling of contaminants in my research has progressively improved.

With regards to the filtering of ASVs based on abundances and contingency, our research group again debated on this topic. Initially, ASVs comprising less than 10 sequences throughout the entire *MicroCOPD* dataset were eliminated after the removal of chimeric and contaminating sequences. Subsequently, we implemented the strategy outlined in Figure 5. Any ASVs lacking taxonomic annotation were excluded,

likely increasing the internal validity. Samples with <500 sequences were not included. Both rarefaction for diversity analyses and ANCOM-BC for differential abundance analyses required a higher sequencing depth for inclusion of samples.

10.3.4.3 Diversity – measures of that within and that between

With data from *MicroCOPD* it was not feasible to prevent loss of samples by setting the rarefaction limit equal to the sparsest sample as I did for the sputum studies. Therefore, the rarefaction level of 2200/sequences per sample was chosen as a compromise between the goal of keeping as many samples as possible and the aim of avoiding an underestimation of diversity.

For paper IV, I included four alpha-diversity metrics and four distance matrices. A detailed description of these is provided in the supplementary materials of paper IV. Using this approach, I was able to examine different aspects of the data and identify the most prominent differences between the groups.

10.3.4.4 Taxonomy

Silva was again chosen for taxonomy in paper IV, despite that the research group otherwise had used the Human Oral Microbiome Database (HOMD) for *MicroCOPD* data (28, 93, 94). I wanted to use Silva for continuity, and I personally worried HOMD could have limitations with regards to the lower airway microbiota. The database comprises of sequencing data from the oral cavity, nasal passages, sinuses, pharynx and oesophagus, but not from the airways distal to the vocal cords (95).

For differential abundance, ANCOM-BC was the only one used in paper IV. ANCOM-BC builds on the previous ANCOM and ANCOM-II and claims to handle the excessive number of zeros, allow for non-normal distribution of data and the bias correction is said to deal with biased sampling. It is vulnerable for false positives due to the many comparisons made which is why we present adjusted p-values, or q-values. It is important to acknowledge that every differential abundance analysis tool has its own strengths and weaknesses. Consequently, limiting the differential abundance analyses to only ANCOM-BC, could represent a shortcoming.

11. Discussion of results

11.1 Comparing induced and spontaneous sputum samples and assessing the safety of hypertonic saline for sputum induction

The first two studies aimed to examine how the different methods of sputum sampling, induced and spontaneous, affected the measurements of sputum inflammation and microbiota. This inquiry arose from our intention to investigate the microbiota and associated inflammation in stable and exacerbated COPD. We wished to use sputum samples from the *BCCS* and the *BCES* from which both sputum types were stored. We noticed that there was limited knowledge regarding the interchangeability of induced and spontaneous sputum for measuring inflammation and microbiota. Therefore, we felt it was necessary to investigate this before conducting the comparative analyses of stable states versus exacerbations in COPD. We also wanted to know more about the safety of the induction procedure performed in the *BCCS* and the *BCES* for future references. The first two studies would have a direct practical application by allowing a more knowledge-based selection of samples for paper III.

I observed discrepancies in the measurements of inflammatory markers and microbiota in consecutively collected induced and spontaneous sputum samples obtained from patients diagnosed with COPD (52, 96). While the observed differences were not statistically significant when the samples were grouped by type, they became evident when each participant was examined individually. It was interesting to observe that statistically significant associations between inflammatory markers and clinical variables like smoking and *H. Influenzae*-positive cultures were indeed not uniformly observed across the two sputum types (52).

Henderson et al. assessed mucus hydration in induced and spontaneous sputum (74). They observed that hypertonic saline had a diluting effect on the induced sputum, which could lead to lower concentration measurements. I have previously discussed the implementation of concentration corrections with albumin and its impact. Henderson et al. did not identify any significant differences in inflammatory markers

when comparing all sputum samples across the two sampling methods in stable state COPD. However, it is important to note that they did not assess the impact of sampling techniques on their measured inflammatory markers for the 18 individual participants with consecutively sampled induced and spontaneous sputum (74).

Earlier studies have suggested that induction can allow for the collection of sputum distal to spontaneous samples (97, 98). Additionally, discrepancies in immune cells and the microbiota along the airway tractus have been observed (25, 26, 99). Induced sputum samples may thus represent a slightly different environment with a different pattern of immunity and microbiota compared with spontaneous samples, even if both sputa originate from proximal rather than distal airways (24).

Mayhew et al. included both types of sputum in a comprehensive, longitudinal study across stable state and exacerbated COPD (100). Comparisons of sputum types were not the main target with their study, and therefor naturally limited. They concluded with a lack of discrepancies between induced and spontaneous sputum based only on UniFrac distances, and a visualisation of percent abundances for three phyla within 65 participants. The changes in percent abundances of *Haemophilus* and *Streptococcus* however seem non-consistent between the plotted panels, and the beta-diversity analyses might not be powered to capture differences between individual patient´s sputum types. The comparisons were made between samples collected from different consultations and with the patients being in different disease states. Thus, it is difficult to use these results to advocate for the inclusion of both sputum types without other confirmatory studies. Diver et al. acknowledged the potential limitation of using both types of sputum in their study and they mentioned that the impact of the sampling method had been assessed (101). Unfortunately, the paper they referred to focused on inflammation and lacked comparisons for microbiota.

In a 2019 review conducted by Ditz et al. a summary of studies on the sputum microbiota in COPD was provided. Based on this review, it can be inferred that the microbiota analysed in sputum samples, whether induced, spontaneous, or analysed collectively, may be linked to several clinical outcomes (102). This in turn, reinforces

the utilisation of sputum as a readily accessible and non-invasive sample for studying the airway microbiota in COPD. In clinical settings it is highly unlikely that bronchoscopic sampling will be routinely performed on admitted patients with COPD exacerbations. It is also unlikely that routine bronchoscopic sampling at a stable state aimed at mapping the microbiota and inflammation for individual patients will be feasible. Consequently, increasing our understanding of the microbiota and inflammation in sputum may prove more useful, as recognised associations and patterns observed in sputum can be more easily applied in a clinical setting. Ditz et al. also discussed the strengths and limitations of different sampling methods (102).

To the best of my knowledge, large-scale studies aimed at investigating the effects of sputum sampling techniques on microbiota and inflammation are still lacking. It is important for future research to acquire a better understanding of the impact of sample processing methods, the natural variability of inflammation and microbiota in sputum, and the specific role of sampling techniques. This knowledge will enable a more accurate interpretation of the associations between measurements of inflammation and microbiota and the clinical variables studied. Deciding on a preferred sample type for microbiota studies would further depend on which sample type correlate strongest with clinical variables.

Taking adequate safety precautions is essential for the researcher as well as the clinician. Ethical considerations in research include strict demands to avoid using methods potentially putting study participants at risk for adverse events. The risks and burdens of induced sputum sampling for study participants have been found acceptable in several studies (21-23).

In our cohort, using hypertonic saline to induce not only stable state but also exacerbated patients including those with very severe COPD, did not result in any severe adverse events (52). Tolerance for the combination of hypertonic saline and the bronchodilator salbutamol might not surprise clinicians as it is offered some hospitalised exacerbated patients with excessive sputum. Studies of the usefulness of hypertonic saline inhalations for patients with COPD has been conflicting, but also

conducted under very different patient settings. This might partially explain why different studies conclude differently (103, 104). A key point is to always protect patients against bronchoconstriction by administrating bronchodilators before or together with the hypertonic saline. Our results on the safety of sputum induction align with the previous studies and support the use of induced sputum for research purposes.

Summarising both the discussion of sputum sampling methods and results from paper I and II I would advocate for the use of induced sputum over spontaneous sputum when studying the lower airway microbiota and inflammation.

11.2 Comparing stable state and exacerbated COPD

My goal for paper III was to identify specific patterns of change in inflammation and microbiota in sputum from patients with COPD sampled both during stable and exacerbated states. Paper I and II resulted in the selection of induced sputum samples exclusively.

I discovered that over 2/3 of patients had substantial changes in the composition of their sputum microbiota from stable to exacerbated state. However, the changes observed were not uniform across patients. Furthermore, the longitudinal case study demonstrated that even within the same individual, there were no clear patterns of change between stable states and exacerbations. For measurements of sputum inflammation clearer patterns were observed for the entirety of the cohort, with three cytokines significantly elevated during exacerbations. Interestingly, their role in viral infections and against intracellular bacteria might be more prominent than their role in defeating extracellular bacterial infections (105-107).

Both stability and variability of the sputum microbiota in patients with COPD have been previously observed, both among different groups of patients, and within individual patients (100, 108-110). Establishing whether the extent to which changes in the microbiota are causally linked to the transition from a stable state to exacerbations is a challenging question that needs to be addressed.

In a comprehensive sputum study published in 2018, Mayhew et al. demonstrated several interesting findings (100). They analysed the microbiota in multiple sputum samples collected over time, both at stable state and exacerbations. They observed a certain level of consistency in the microbial composition within patients, allowing for differentiation between participants based on their sputum microbiota. Furthermore, they demonstrated that exacerbations were more closely linked to changes in the abundance of taxa rather than the acquisition or loss of taxa. These results align with our own Procrustes findings, as the matrices sensitive to abundances (Bray-Curtis and weighted UniFrac) exhibited higher M^2 values compared to their abundance-insensitive counterparts (46).

Mayhew et al. also observed significantly higher relative abundances of *Moraxella* in samples collected during exacerbations, and that this bacterial genus had the strongest association with the frequency of exacerbations (100). Ideally, this would support targeted therapy directed at *Moraxella* in patients with COPD. Alas, it is also described that only a minority of exacerbated patients had a substantial increase in *Moraxella.* Our case study aligns with this observation. Figure 9 in paper III demonstrates substantial fluctuations in *Moraxella* abundances within the COPD participant both during exacerbations and periods of stability. *Moraxella* was both completely absent and the most abundant genera in both states of COPD (46). Previous analysis of our sputum samples revealed that *Moraxella* was substantially more common in spontaneous sputum compared to induced sputum. However, there was no discernible correlation with exacerbations when examining all patients collectively. Similarly, Wang et al. did not find any correlation either, but they did note that higher abundances of *Moraxella* were negatively correlated with Shannon diversity index (111).

The overall absence of significantly different taxa in my sputum data could be attributed to the small number of participants. But it could also be a result of the fact that changes in taxa abundances were opposite in different patients. This observation can be seen in Figure 4 of paper III for several genera. Our findings suggest that employing a single treatment approach for all patients will be ineffective in COPD.

The lack of consistent changes in the microbiota across disease states in my study could be partly explained by the multifactorial nature of COPD. This introduces the need for sub-group analyses of larger data to find COPD phenotypes in which there are consistent patterns. While my study was too small to run sensible sub-group analyses incorporating inflammation measurements, Wang et.al., presented such a study in 2021 (112). They classified COPD sputum samples according to leukocyte patterns and further sub-classified neutrophilic sputum according to the microbiota composition (balanced, *Haemophilus*). They reported distinct associations between the sub-groups and several inflammatory markers from sputum and serum. They also

observed that patients with neutrophilic-balanced stable state sputum were more likely to experience larger changes in the sputum microbiota during exacerbations. Interestingly, in my study, large sputum microbiota changes during exacerbations were associated with a depletion of sputum SLPI. Negative associations between sputum neutrophil percentages and SLPI were discussed by Persson et al., who also pointed at studies linking SLPI with bacterial infections (47, 113, 114). Links between a dysbiotic sputum microbiota in exacerbated COPD and neutrophilic-balanced sputum at stable state, and a depletion of SLPI during exacerbations, could indicate a potentially harmful interplay between the lower airway microbiota and inflammatory system.

The evaluation of serum inflammatory markers in our data was limited. We did observe that the seven participants classified with high serum absolute neutrophil counts during exacerbations had significantly higher Shannon diversity index when sampled at stable state. This contradicts the result from Lonergan et al. on stable state COPD showing that high blood neutrophil counts were associated with significantly lower Shannon diversity index (115). Wang et al. found a joint association between microbiota dysbiosis and blood eosinophilia on the one hand, and a significantly more reduced FEV_1 at exacerbations compared with stable state on the other (111). Dicker et al. were able to observe associations between blood eosinophils and relative abundance of *Proteobacteria/Haemophilus* (negatively associated) and *Firmicutes*/*Streptococcus* (positively associated) in sputum samples collected at stable state of COPD (116). From our data, though not reported in paper III, we did not observe any associations between eosinophils in blood and the microbiota in sputum either at stable state or during exacerbations. The correlation between blood inflammatory markers and the lower airway microbiota remains understudied.

In summary, there is a need to view the lower airway microbiota in concert with both local and systemic inflammation. Expanding analyses to include the effect of fungi and viruses is likely necessary to paint a fuller picture of what drives the inflammation in COPD airways and lung tissue. To effectively customise treatment for patients with

COPD we may depend on identifying well-defined inflammatory-microbial phenotypes for which targeted treatment can be prescribed.

11.3 Comparing patients with COPD and healthy controls

My goal for paper IV was to find potentially important differences in the lower airway microbiota between a large group of patients with COPD and a similarly sized group of healthy participants. For this paper, the microbiota was measured in BAL.

The remodelling of airways and lung tissue in COPD, once established, cannot be reversed. Therefore, it is crucial to prevent the damage from occurring in the first place. If it is too late to prevent the disease, the next best objective is to halt its further development. Ultimately, our research goal is to uncover a causal relationship between a treatable bacterial dysbiosis in healthy smokers and COPD. Additionally, we aim to identify a causal relationship between an equally treatable bacterial dysbiosis and disease progression and exacerbations on behalf of those already diagnosed with COPD. In the following discussion, I will present our observations from paper IV in the context of relevant and available publications regarding the lower airway microbiota in COPD and healthy controls.

We found that BAL microbiota in COPD had lower evenness than in healthy controls. The most prevalent phyla and genera found in patients with COPD were also the most prevalent in healthy controls. Still, nine genera differed significantly across COPD and controls. One phylum was enriched in COPD and that was *Firmicutes*. In fact, only genera belonging to the phylum *Firmicutes* were significantly more abundant in COPD compared with controls.

Alpha-diversity has previously been reported to be decreased in airway samples from patients with COPD compared with controls, and to be further lowered with disease progression (27, 117-120). In paper IV, we observed that the richness and the phylogenetic diversity within BAL from patients with COPD were comparable to that observed in healthy controls. What separated COPD and control BAL was the

evenness of the different microbiota members present in the samples. If we dissect the studies listed in Table 1 and Table 2 we find that Ramsheh et al. observed a COPD-associated reduction of all alpha-diversity characteristics (richness, evenness, and phylogenetic diversity) in their comprehensive study of 339 patients with COPD and 207 healthy controls (120). The remaining studies on alpha-diversity across COPD and health are far smaller. Sze et al. studied the microbiota in lung tissue and compared alpha-diversity between patients with COPD and controls with two different mathematical methods (118, 121). The reported lower Shannon diversity index in lung tissue from patients with COPD compared with controls was observed only at one of six sampled positions within the nine lungs that were examined. The Simpson index calculated in their study published in 2012 showed no difference in alpha-diversity between COPD and controls. This was also the case for both Pragman et al., and Zakharkina et al. who both chose Shannon diversity index to describe alpha-diversity (122, 123). One can suspect that some of these studies have been too small to detect a difference between COPD and control samples.

Together with the study of Ramsheh et al., and Opron et al. we have the largest patient population sampled from the distal airways. So why are both the richness and phylogenetic diversity sustained in our, but not in Ramsheh et al.´s, COPD population when compared to controls? Disease severity could have been a suggested cause, as Mayhew et al. observed that the severity of COPD was negatively associated with alpha-diversity (100). But the percent cases of severe and very severe COPD were 31 in our study and only 12 in Ramsheh et al.´s study population. The discrepancy between our and Ramsheh et al.´s study is thus not in accordance with the observations made by Mayhew et al. The lower airway microbiota is proposed to be dynamic with bacterial influx through inhalation and micro-aspiration (124), and removal through mucociliary clearance, cough, and immune responses. In addition, one can postulate that local replication and survival will vary between the different microbiota members (125). One possible explanation for the loss of evenness but sustained richness and phylogenetic diversity observed in our samples could be that it

reflects a regular influx of the upper airway microbiota, combined with local factors in the lower airways causing some taxa to thrive.

The lack of difference in beta-diversity between BAL from patients with COPD and controls agrees with previous studies (120, 126, 127). In sputum, Wang et al. observed a significant difference in weighted UniFrac distances between COPD and controls, but with an effect size of only 3% (R^2), thus having disease explain very little of the variation in the model (119). Opron et al. observed significant associations between beta-diversity measures and both COPD assessment scores (CAT scores) and bronchodilator response. Meanwhile, the significant p-values from the PERMANOVA analyses were accompanied by very low R^2 scores (1% - 2%), thus the effect sizes of CAT scores and bronchodilator response were also very small.

In COPD, the shift from a state of good health to chronic obstruction occurs gradually, rather than abruptly. Consequently, it is conceivable that alterations in the microbiota of the lower airways could also exhibit a continuum of change, making it difficult to distinguish between the healthy and the obstructive airways. Additionally, considering the continual supply of microorganisms from the upper airway and the dynamic nature of the lower airway microbiota, these factors might account for the absence of a correlation between beta-diversity and COPD or health.

Diving into the taxonomy, we observed several taxa with differential abundance across COPD and controls, as would have been expected from the alpha-diversity results. Differential abundance of taxa associated with disease was more pronounced than that associated with other demographics considered in our study (Paper VI). In COPD, we observed an increase only in genera belonging to the *Firmicutes* phylum: *Granulicatella*, *Gemella,* and *Streptococcus*. For the potentially pathological genera *Streptococcus*, there was an additional, successive increase in relative abundances with increased disease severity. Opron et al., and Wang et al. observed the same pattern in BAL and sputum, respectively (119, 126). Ramsheh et al. also observed an increase of *Gemella* and *Streptococcus* in patients with COPD compared with their control group. Furthermore, they observed an enrichment of *Prevotella,* and

Leptotrichia (120). Thus, all three bronchoscopy studies comparing the microbiota in COPD with healthy controls observe that the distal airways of patients with COPD harbours significantly more *Streptococcus* than the airways of healthy controls. The result of Dicker et al. is interesting in this context as they observed associations between *Streptococcus* in induced sputum in COPD and increased blood eosinophil levels. Sputum enriched with *Streptococcus* was also associated with frequent exacerbations (116). Opron et al. found another link between *Streptococcus* and host responses as neutrophils in BAL were positively associated with abundances of *Streptococcus* (126). This indicate that *Streptococcus* trigger both local and systemic inflammatory activity and is putting patients at risk for more frequent exacerbation events, which again is linked to declining quality of life and life expectancy and increasing disability and need for hospitalisation. Whether the lower airway in COPD has an impaired inflammatory profile in which *Firmicutes* are allowed to multiply, or if *Firmicutes* multiplies with a subsequent change in inflammation should be investigated further in longitudinal studies.

Irrespective of a diagnosis of COPD, the correlations between sex and age and the lower airway microbiota were few. Neither sex or age were found associated with diversity measures, and although there was a decrease in *Bacteroidota* with age among healthy controls, this trend was not evident at the genus level. Therefore, it is challenging to draw any definitive biological conclusions from this observation. Additionally, there were no significant associations between the severity of COPD and frequency of exacerbations in our COPD patients and the BAL microbiota. In fact, none of the observed associations were consistent enough to propose a clear biological significance (paper IV).

Considering the well-established role of cigarette smoking in the development of COPD, it is reasonable to expect a connection between smoking and the composition of the lower airway microbiota. However, we found minimal evidence of such a relationship in the BAL samples of our study participants. In terms of alpha-diversity, there were no significant variations in any of the four alpha-diversity metrics included

in our study. This remained true even when analysing COPD patients and controls individually. Regarding beta-diversity, PERMANOVA analysis confirmed a discrepancy in weighted UniFrac between smoking and non-smoking COPD patients, but the effect size was only 4%. Smoking thus remain weakly linked to the diversity of the lower airway microbiota, as has been observed in other large studies as well (100, 119, 126).

Is it possible that certain members of the microbiota could be affected by smoking while the overall diversity remained unchanged? Our findings indicate that smoking, measured in terms of pack years, is significantly linked to a decrease in *Haemophilus* and *Lachnoanarobaculum* in healthy controls (paper IV). This finding aligns with the research conducted by Pfeiffer et al. which examined the impact of smoking on the airway microbiota in healthy subjects (128). Within our COPD cohort, *Oribacterium* was decreased in smoking patients. Common for *Lachnoanarobaculum* and three genera significantly differentially abundant in all smokers versus non-smokers is a low abundance and presence in few samples. The research on smoking and the lower airway microbiota has so far not identified convincing links between the two (126, 129).

In paper IV we further assessed if ICS could influence the airway microbiota in our COPD cohort. This is an interesting question as frequent exacerbators with elevated blood eosinophils have been identified as a COPD sub-group who can benefit from ICS, while others should better avoid ICS due to the known increase in the risk of pneumonia and infectious exacerbations (6, 130, 131). Keir et al. concluded a review of ICS and the lung microbiota by stressing that ICS use should be founded on leucocyte endotypes to avoid prescription to patients at a higher risk of *Streptococcus* and *Haemophilus* overgrowth (132). Interestingly, of the studies included in this review, only two on COPD in humans showed increased *Streptococcus* abundances, and a third study showed decreased abundances of the taxon in case of ICS use (133-135). Contradictive observations for *Haemophilus* are also apparent *(135, 136)*. With Dicker showing that both blood eosinophiles and frequent exacerbations were

associated with increased *Streptococcus* in sputum samples (116), it is difficult to fit the microbiota observations with the long-accepted observations of an adverse association between ICS and pneumonia risk. In my study, 60 of the 97 patients with COPD used ICS. In stable COPD, we did not find evidence of a significant influence of ICS on the lower airway microbiota, including the abundances of *Streptococcus* and *Haemophilus.* This is supported by the observations made by both Opron and Ramsheh (120, 126).

Leitao Filjo et al. studied the lower airway microbiota and ICS using bronchial brush samples from 56 patients with COPD randomised into two different ICS treatment and one non-ICS treatment groups (136). Both lower richness and Shannon diversity index were observed in patients receiving ICS. However, the visualisation of within-individual differences in alpha-diversity between sampling time-points resembles that of our paper I (46). Whether these within-individual differences are linked to ICS or exacerbations respectively, or if it represents what to be expected between repeated samples is unknown. Leitao Filjo et al. observed no associations between ICS and *Streptococcus*, while *Haemophilus* was reduced after the 12-week treatment period with fluticasone (136). The current evidence of a link between ICS treatment and unfavourable changes in the lower airway microbiota in patients with COPD thus remains rather unclear.

To summarise our results, there were convincing differences between the BAL microbiota measured in our group of patients with COPD and that measured from the group of healthy controls. The main cause of COPD, smoking, could not be associated any closer to the microbiota than other characteristics, and neither could ICS even though it is an established risk factor for infectious COPD exacerbations and pneumonia. The lower airway microbiota is likely to play a part in COPD and other lung diseases, but to unveil how its conducting its influence proves difficult. Future research will call for complex study designs integrating host inflammatory activity and using state of the art bioinformatic processing and statistical analyses.

12. Conclusion

I conducted the two first studies with the intention to understand the influence of different methods of sputum sampling on the measurement of lower airway inflammation and microbiota and to use the results to direct the sample selection for my third study.

Both studies showed there were differences between induced and spontaneous sputum in terms of inflammatory markers and microbiota profiles in COPD patients. These findings emphasise the importance of understanding the variations between these two types of sputum samples. Further research is warranted to fully explore the clinical implications and potential applications of this knowledge. Meanwhile, based on the findings from paper I and II, I advocate for the use of induced sputum over spontaneous sputum when studying the lower airway microbiota and inflammation.

Next, my goal was to identify specific patterns of change in inflammation and microbiota in induced sputum from patients with COPD enrolled if sampled both during stable and exacerbated states.

The study revealed significant variability in the alterations of sputum microbiota during the progression from a stable to an exacerbated state. This variation was also observed within the same individual over a three-year period. Considering previous research findings, our results support that while the bacterial microbiota may play a role in both stable and exacerbated COPD, its impact likely intertwines with host inflammation and potentially other microorganisms such as fungi and viruses. To tailor treatments for COPD patients, it is crucial to identify precise inflammatory-microbial phenotypes that can guide the prescription of targeted therapies including antibiotics, corticosteroids, bacteriophage therapy and pro/prebiotics.

Lastly, I aimed to find potentially important differences in the lower airway microbiota between a large group of patients with COPD and a similarly sized group of healthy participants, using bronchoalveolar lavage (BAL).

Our findings indicate significant differences between the lower airway microbiota in COPD compared to in healthy controls. Despite smoking being a major cause of COPD, we were unable to establish a stronger association between smoking and the microbiota compared to other factors. Similarly, although ICS are known to increase the risk of airway infections, we could not find a link between ICS and the microbiota. Understanding the impact of the lower airway microbiota on COPD and other respiratory conditions continues to present significant challenges.

13. Implications and future perspectives

Microbiota research has the potential to improve our understanding of COPD and pave the way for new treatment strategies. While earlier research has focused on the role of inflammation and genetics in COPD, recent studies have begun to unravel the influence of the lung microbiota on disease development, progression, and exacerbations. Understanding the complex interplay between the lower airway microbiota, host immune responses, and disease progression could lead to innovative therapeutic strategies including probiotics, prebiotics, or bacteriophage therapy. Additionally, manipulation of the lung microbiota could potentially halt or slow down disease progression, transforming COPD management into a more proactive approach. Meanwhile, significant challenges lie ahead, and a selection of them are listed below.

1. Standardisation of methods: One major challenge is the lack of standardisation of methods for studying the lower airway microbiota. Researchers use different techniques, sample types, and analytical methods, making it difficult to compare and replicate findings. The high-dimensional, complex data generated through microbiota research require advanced computational tools for analysis and interpretation. There is a need for standardised bioinformatics pipelines and data sharing platforms facilitating collaborative research and meta-analyses, and for standardised protocols and methodologies allowing for better comparison of results.

2. Biological and clinical heterogeneity: COPD encompasses a heterogeneous group of patients, with different clinical phenotypes and patterns of disease progression. The microbiota composition and its relationship with disease severity, exacerbations, and response to therapy varies among individuals. Understanding this heterogeneity and identifying microbiota signatures associated with specific subgroups will aid in personalised treatment approaches. There is a need to supply the present day single-kingdom studies with broader examinations including all microbiological kingdoms inhabiting the lower airways. Studies integrating information from the bacterial, fungal, and archaeal communities and viruses in the lower airways will provide a more holistic view on the interplay between host inflammation and microbes.

3. Longitudinal studies: COPD is a chronic and progressive disease, and the microbiota composition changes over time. Longitudinal studies tracking the dynamics of the lower airway microbiota and its association with disease development, progression and exacerbations are needed. Large-scale cohort studies with long follow-up periods would provide valuable insights into the temporal relationship between the lower airway microbiota and COPD.

4. Therapeutic interventions: Microbiota-targeted interventions may hold promise for COPD management, but the development of novel therapies requires rigorous research involving larger patient cohorts and long-term follow-up assessing the safety and efficacy of new treatments.

5. Causality and mechanisms: Establishing a causal relationship between altered microbiota and disease development or progression is challenging. With the limitations of amplicon sequencing techniques, whole genome sequencing represents a promising method for improved characterisation and understanding of the microbiota. A true understanding of the role of the microbiome requires assessment of its functionality. With whole genome sequencing, species resolution of the full (bacterial and fungal) microbiota can be used to calculate dysbiosis indexes based on which species are differentially abundant between groups, potentially making for new disease biomarkers. Further, the functionality of the lower airway microbiome can be assessed from whole genome sequencing. To elucidate the mechanistic pathways through which microbiota influences COPD pathogenesis and exacerbations researchers must now move beyond the 16S rRNA sequencing.

In conclusion, microbiota research in COPD holds promise for the future. By unravelling the intricate connections between the microbiota and disease pathogenesis, we may unlock novel therapeutic options and develop personalised treatment regimens for COPD patients. With further research, we can anticipate a future where microbiota-based interventions are part of the management of COPD.

14. Errata

In paper I, I stated that sputum leucocyte particle count (LPK) viability >30% was considered sufficient, when in fact it had to be >70%.

15. References

1. O'Donnell R, Breen D, Wilson S, Djukanovic R. Inflammatory cells in the airways in COPD. Thorax. 2006;61(5):448-54.
2. Pando-Sandoval A, Ruano-Ravina A, Candal-Pedreira C, Rodriguez-Garcia C, Represas-Represas C, Golpe R, et al. Risk factors for chronic obstructive pulmonary disease in never-smokers: A systematic review. Clin Respir J. 2022;16(4):261-75.
3. Strange C. Alpha-1 Antitrypsin Deficiency Associated COPD. Clin Chest Med. 2020;41(3):339-45.
4. Lokke A, Lange P, Scharling H, Fabricius P, Vestbo J. Developing COPD: a 25 year follow up study of the general population. Thorax. 2006;61(11):935-9.
5. Crotty Alexander LE, Shin S, Hwang JH. Inflammatory Diseases of the Lung Induced by Conventional Cigarette Smoke: A Review. Chest. 2015;148(5):1307-22.
6. Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global Initiative for Chronic Obstructive Lung Disease 2022,2023 2023.
7. Sapey E, Stockley RA. COPD exacerbations . 2: aetiology. Thorax. 2006;61(3):250-8.
8. Herath SC, Normansell R, Maisey S, Poole P. Prophylactic antibiotic therapy for chronic obstructive pulmonary disease (COPD). Cochrane Database Syst Rev. 2018;10(10):CD009764.
9. Husebø G. Systemic inflammatory markers as predictors of longitudinal outcomes in COPD Results from the Bergen COPD Cohort Study. https://bora.uib.no/bora-xmlui/handle/11250/2756623: University of Bergen; 2021.
10. Baker JR, Donnelly LE. Leukocyte Function in COPD: Clinical Relevance and Potential for Drug Therapy. Int J Chron Obstruct Pulmon Dis. 2021;16:2227-42.
11. Bafadhel M, Haldar K, Barker B, Patel H, Mistry V, Barer MR, et al. Airway bacteria measured by quantitative polymerase chain reaction and culture in patients with stable COPD: relationship with neutrophilic airway inflammation, exacerbation frequency, and lung function. Int J Chron Obstruct Pulmon Dis. 2015;10:1075-83.
12. Berg G, Rybakova D, Fischer D, Cernava T, Verges MC, Charles T, et al. Microbiome definition re-visited: old concepts and new challenges. Microbiome. 2020;8(1):103.
13. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. Nature. 2007;449(7164):804-10.
14. Hilty M, Burke C, Pedro H, Cardenas P, Bush A, Bossley C, et al. Disordered microbial communities in asthmatic airways. PLoS One. 2010;5(1):e8578.
15. Sethi S, Evans N, Grant BJ, Murphy TF. New strains of bacteria and exacerbations of chronic obstructive pulmonary disease. N Engl J Med. 2002;347(7):465-71.
16. Paggiaro PL, Chanez P, Holz O, Ind PW, Djukanovic R, Maestrelli P, et al. Sputum induction. Eur Respir J Suppl. 2002;37:3s-8s.
17. Pizzichini MM, Popov TA, Efthimiadis A, Hussack P, Evans S, Pizzichini E, et al. Spontaneous and induced sputum to measure indices of airway inflammation in asthma. Am J Respir Crit Care Med. 1996;154(4 Pt 1):866-9.
18. Bhowmik A, Seemungal TA, Sapsford RJ, Devalia JL, Wedzicha JA. Comparison of spontaneous and induced sputum for investigation of airway inflammation in chronic obstructive pulmonary disease. Thorax. 1998;53(11):953-6.
19. Wark PA, Simpson JL, Hensley MJ, Gibson PG. Safety of sputum induction with isotonic saline in adults with acute severe asthma. Clin Exp Allergy. 2001;31(11):1745-53.
20. Pizzichini E, Pizzichini MM, Leigh R, Djukanovic R, Sterk PJ. Safety of sputum induction. Eur Respir J Suppl. 2002;37:9s-18s.

21. Bathoorn E, Liesker J, Postma D, Koeter G, van Oosterhout AJ, Kerstjens HA. Safety of sputum induction during exacerbations of COPD. Chest. 2007;131(2):432-8.
22. Makris D, Tzanakis N, Moschandreas J, Siafakas NM. Dyspnea assessment and adverse events during sputum induction in COPD. BMC Pulm Med. 2006;6:17.
23. Gao P, Gibson PG, Zhang J, He X, Hao Y, Li P, et al. The safety of sputum induction in adults with acute exacerbation of COPD. Clin Respir J. 2013;7(1):101-9.
24. Alexis NE, Hu SC, Zeman K, Alter T, Bennett WD. Induced sputum derives from the central airways: confirmation using a radiolabeled aerosol bolus delivery technique. Am J Respir Crit Care Med. 2001;164(10 Pt 1):1964-70.
25. Cabrera-Rubio R, Garcia-Nunez M, Seto L, Anto JM, Moya A, Monso E, et al. Microbiome diversity in the bronchial tracts of patients with chronic obstructive pulmonary disease. J Clin Microbiol. 2012;50(11):3562-8.
26. Gronseth R, Drengenes C, Wiker HG, Tangedal S, Xue Y, Husebo GR, et al. Protected sampling is preferable in bronchoscopic studies of the airway microbiome. ERJ Open Res. 2017;3(3).
27. Mika M, Nita I, Morf L, Qi W, Beyeler S, Bernasconi E, et al. Microbial and host immune factors as drivers of COPD. ERJ Open Res. 2018;4(3).
28. Leiten EO, Nielsen R, Wiker HG, Bakke PS, Martinsen EMH, Drengenes C, et al. The airway microbiota and exacerbations of COPD. ERJ Open Res. 2020;6(3).
29. Bothwell LE, Greene JA, Podolsky SH, Jones DS. Assessing the Gold Standard--Lessons from the History of RCTs. N Engl J Med. 2016;374(22):2175-81.
30. Bosdriesz JR, Stel VS, van Diepen M, Meuleman Y, Dekker FW, Zoccali C, et al. Evidence-based medicine-When observational studies are better than randomized controlled trials. Nephrology (Carlton). 2020;25(10):737-43.
31. Rothman KJ. Epidemiology : an introduction. 2nd ed. New York, NY: Oxford University Press; 2012. viii, 268 p. p.
32. Dickson RP, Erb-Downward JR, Martinez FJ, Huffnagle GB. The Microbiome and the Respiratory Tract. Annu Rev Physiol. 2016;78:481-504.
33. Kaul A, Mandal S, Davidov O, Peddada SD. Analysis of Microbiome Data in the Presence of Excess Zeros. Front Microbiol. 2017;8:2114.
34. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome Datasets Are Compositional: And This Is Not Optional. Front Microbiol. 2017;8:2224.
35. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37(8):852-7.
36. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75(23):7537-41.
37. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; 2020.
38. Whittaker RH. Evolution and Measurement of Species Diversity. Taxon. 1972;21.
39. Johnson JS, Spakowicz DJ, Hong BY, Petersen LM, Demkowicz P, Chen L, et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. Nat Commun. 2019;10(1):5029.
40. Tamilarasan AG, Krishnananthan T. Faecal microbiota transplantation: what's beyond Clostridium difficile infection? Eur J Gastroenterol Hepatol. 2021;33(4):487-94.
41. Man WH, de Steenhuijsen Piters WA, Bogaert D. The microbiota of the respiratory tract: gatekeeper to respiratory health. Nat Rev Microbiol. 2017;15(5):259-70.

42. Eagan TM, Ueland T, Wagner PD, Hardie JA, Mollnes TE, Damas JK, et al. Systemic inflammatory markers in COPD: results from the Bergen COPD Cohort Study. Eur Respir J. 2010;35(3):540-8.
43. Eagan TM, Aukrust P, Ueland T, Hardie JA, Johannessen A, Mollnes TE, et al. Body composition and plasma levels of inflammatory biomarkers in COPD. Eur Respir J. 2010;36(5):1027-33.
44. Persson LJ, Aanerud M, Hiemstra PS, Michelsen AE, Ueland T, Hardie JA, et al. Vitamin D, vitamin D binding protein, and longitudinal outcomes in COPD. PLoS One. 2015;10(3):e0121622.
45. Waatevik M, Johannessen A, Gomez Real F, Aanerud M, Hardie JA, Bakke PS, et al. Oxygen desaturation in 6-min walk test is a risk factor for adverse outcomes in COPD. Eur Respir J. 2016;48(1):82-91.
46. Tangedal S, Nielsen R, Aanerud M, Persson LJ, Wiker HG, Bakke PS, et al. Sputum microbiota and inflammation at stable state and during exacerbations in a cohort of chronic obstructive pulmonary disease (COPD) patients. PLoS One. 2019;14(9):e0222449.
47. Persson LJ, Aanerud M, Hardie JA, Miodini Nilsen R, Bakke PS, Eagan TM, et al. Antimicrobial peptide levels are linked to airway inflammation, bacterial colonisation and exacerbations in chronic obstructive pulmonary disease. Eur Respir J. 2017;49(3).
48. Gronseth R, Haaland I, Wiker HG, Martinsen EM, Leiten EO, Husebo G, et al. The Bergen COPD microbiome study (MicroCOPD): rationale, design, and initial experiences. Eur Clin Respir J. 2014;1.
49. Leiten EO, Eagan TML, Martinsen EMH, Nordeide E, Husebo GR, Knudsen KS, et al. Complications and discomfort after research bronchoscopy in the MicroCOPD study. BMJ Open Respir Res. 2020;7(1).
50. Taube C, Holz O, Mucke M, Jorres RA, Magnussen H. Airway response to inhaled hypertonic saline in patients with moderate to severe chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2001;164(10 Pt 1):1810-5.
51. Efthimiadis A, Spanevello A, Hamid Q, Kelly MM, Linden M, Louis R, et al. Methods of sputum processing for cell counts, immunocytochemistry and in situ hybridisation. Eur Respir J Suppl. 2002;37:19s-23s.
52. Tangedal S, Aanerud M, Persson LJ, Brokstad KA, Bakke PS, Eagan TM. Comparison of inflammatory markers in induced and spontaneous sputum in a cohort of COPD patients. Respir Res. 2014;15(1):138.
53. Tjabringa GS, Vos JB, Olthuis D, Ninaber DK, Rabe KF, Schalkwijk J, et al. Host defense effector molecules in mucosal secretions. FEMS Immunol Med Microbiol. 2005;45(2):151-8.
54. 16S Amplicon PCR for the V3-V4 region for the MicroCOPD samples [Internet]. 2019. Available from: https://www.protocols.io/view/16s-amplicon-pcr-for-the-v3-v4-region-for-the-micr-2sygefw.
55. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7(5):335-6.
56. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, et al. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat Methods. 2013;10(1):57-9.
57. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13(7):581-3.

58. Rognes T, Flouri T, Nichols B, Quince C, Mahe F. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 2016;4:e2584.
59. Konstantinidis KT, Tiedje JM. Genomic insights that advance the species definition for prokaryotes. Proc Natl Acad Sci U S A. 2005;102(7):2567-72.
60. Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. 2017;11(12):2639-43.
61. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol. 2006;72(7):5069-72.
62. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics. 2010;26(2):266-7.
63. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30(14):3059-66.
64. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41(Database issue):D590-6.
65. Robeson MS, 2nd, O'Rourke DR, Kaehler BD, Ziemski M, Dillon MR, Foster JT, et al. RESCRIPt: Reproducible sequence taxonomy reference database management. PLoS Comput Biol. 2021;17(11):e1009581.
66. Sayers EW, Agarwala R, Bolton EE, Brister JR, Canese K, Clark K, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2019;47(D1):D23-D8.
67. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol. 2009;26(7):1641-50.
68. Drengenes C, Wiker HG, Kalananthan T, Nordeide E, Eagan TML, Nielsen R. Laboratory contamination in airway microbiome studies. BMC Microbiol. 2019;19(1):187.
69. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome. 2018;6(1):226.
70. Bokulich NA, Dillon MR, Bolyen E, Kaehler BD, Huttley GA, Caporaso JG. q2-sample-classifier: machine-learning tools for microbiome classification and regression. J Open Res Softw. 2018;3(30).
71. Ferdous T, Jiang L, Dinu I, Groizeleau J, Kozyrskyj AL, Greenwood CMT, et al. The rise to power of the microbiome: power and sample size calculation for microbiome studies. Mucosal Immunol. 2022;15(6):1060-70.
72. Bafadhel M, McCormick M, Saha S, McKenna S, Shelley M, Hargadon B, et al. Profiling of sputum inflammatory mediators in asthma and chronic obstructive pulmonary disease. Respiration. 2012;83(1):36-44.
73. Woolhouse IS, Bayley DL, Stockley RA. Effect of sputum processing with dithiothreitol on the detection of inflammatory mediators in chronic bronchitis and bronchiectasis. Thorax. 2002;57(8):667-71.
74. Henderson AG, Anderson WH, Ceppe A, Coakley RD, Button B, Alexis NE, et al. Mucus Hydration in Subjects with Stable Chronic Bronchitis: A Comparison of Spontaneous and Induced Sputum. COPD. 2018;15(6):572-80.
75. Yuan S, Cohen DB, Ravel J, Abdo Z, Forney LJ. Evaluation of methods for the extraction and purification of DNA from the human microbiome. PLoS One. 2012;7(3):e33865.

76. Drengenes C, Eagan TML, Haaland I, Wiker HG, Nielsen R. Exploring protocol bias in airway microbiome studies: one versus two PCR steps and 16S rRNA gene region V3 V4 versus V4. BMC Genomics. 2021;22(1):3.
77. Abellan-Schneyder I, Matchado MS, Reitmeier S, Sommer A, Sewald Z, Baumbach J, et al. Primer, Pipelines, Parameters: Issues in 16S rRNA Gene Sequencing. mSphere. 2021;6(1).
78. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol. 2014;12:87.
79. Sze MA, Schloss PD. The Impact of DNA Polymerase and Number of Rounds of Amplification in PCR on 16S rRNA Gene Sequence Data. mSphere. 2019;4(3).
80. Witzke MC, Gullic A, Yang P, Bivens NJ, Adkins PRF, Ericsson AC. Influence of PCR cycle number on 16S rRNA gene amplicon sequencing of low biomass samples. J Microbiol Methods. 2020;176:106033.
81. Stoler N, Nekrutenko A. Sequencing error profiles of Illumina sequencing instruments. NAR Genom Bioinform. 2021;3(1):lqab019.
82. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999;8(2):135-60.
83. Giavarina D. Understanding Bland Altman analysis. Biochem Med (Zagreb). 2015;25(2):141-51.
84. Navas-Molina JA, Peralta-Sanchez JM, Gonzalez A, McMurdie PJ, Vazquez-Baeza Y, Xu Z, et al. Advancing our understanding of the human microbiome using QIIME. Methods Enzymol. 2013;531:371-444.
85. Chao A. Non-parametric estimation of the number of classes in a population. Scandinavian Journal of Statistics. 1984;11(4):265-70.
86. Lozupone CK, R. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol. 2005;71(12):8228-35.
87. Buttigieg PL, Ramette A. A guide to statistical analysis in microbial ecology: a community-focused, living review of multivariate data analyses. FEMS Microbiol Ecol. 2014;90(3):543-50.
88. Dixit K, Davray D, Chaudhari D, Kadam P, Kshirsagar R, Shouche Y, et al. Benchmarking of 16S rRNA gene databases using known strain sequences. Bioinformation. 2021;17(3):377-91.
89. Bassis CM, Erb-Downward JR, Dickson RP, Freeman CM, Schmidt TM, Young VB, et al. Analysis of the upper respiratory tract microbiotas as the source of the lung and gastric microbiotas in healthy individuals. MBio. 2015;6(2):e00037.
90. Yue JC, Clayton MK. A similarity measure based on species proportions. Commun Stat Theory Methods. 2005;34(11):2123-31.
91. Drengenes C. Methodological issues in airway microbiome studies. https://hdl.handle.net/11250/2739784: University of Bergen; 2020.
92. Jari Oksanen FGB, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs, Helene Wagner. vegan: Community Ecology Package. R package version 2.5-2. 2018.
93. Knudsen KS, Lehmann S, Nielsen R, Tangedal S, Haaland I, Hiemstra PS, et al. The lower airways microbiome and antimicrobial peptides in idiopathic pulmonary fibrosis differ from chronic obstructive pulmonary disease. PLoS One. 2022;17(1):e0262082.

94. Nielsen R, Xue Y, Jonassen I, Haaland I, Kommedal O, Wiker HG, et al. Repeated bronchoscopy in health and obstructive lung disease: is the airway microbiome stable? BMC Pulm Med. 2021;21(1):342.
95. Chen T, Yu WH, Izard J, Baranova OV, Lakshmanan A, Dewhirst FE. The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. Database (Oxford). 2010;2010:baq013.
96. Tangedal S, Aanerud M, Gronseth R, Drengenes C, Wiker HG, Bakke PS, et al. Comparing microbiota profiles in induced and spontaneous sputum samples in COPD patients. Respir Res. 2017;18(1):164.
97. Gershman NH, Liu H, Wong HH, Liu JT, Fahy JV. Fractional analysis of sequential induced sputum samples during sputum induction: evidence that different lung compartments are sampled at different time points. J Allergy Clin Immunol. 1999;104(2 Pt 1):322-8.
98. Holz O, Jorres RA, Koschyk S, Speckin P, Welker L, Magnussen H. Changes in sputum composition during sputum induction in healthy and asthmatic subjects. Clin Exp Allergy. 1998;28(3):284-92.
99. Moodley YP, Krishnan V, Lalloo UG. Neutrophils in induced sputum arise from central airways. Eur Respir J. 2000;15(1):36-40.
100. Mayhew D, Devos N, Lambert C, Brown JR, Clarke SC, Kim VL, et al. Longitudinal profiling of the lung microbiome in the AERIS study demonstrates repeatability of bacterial and eosinophilic COPD exacerbations. Thorax. 2018;73(5):422-30.
101. Diver S, Richardson M, Haldar K, Ghebre MA, Ramsheh MY, Bafadhel M, et al. Sputum microbiomic clustering in asthma and chronic obstructive pulmonary disease reveals a Haemophilus-predominant subgroup. Allergy. 2020;75(4):808-17.
102. Ditz B, Christenson S, Rossen J, Brightling C, Kerstjens HAM, van den Berge M, et al. Sputum microbiome profiling in COPD: beyond singular pathogen detection. Thorax. 2020;75(4):338-44.
103. Khan SY, O'Driscoll BR. Is nebulized saline a placebo in COPD? BMC Pulm Med. 2004;4:9.
104. Valderramas SR, Atallah AN. Effectiveness and safety of hypertonic saline inhalation combined with exercise training in patients with chronic obstructive pulmonary disease: a randomized trial. Respir Care. 2009;54(3):327-33.
105. Waters JP, Pober JS, Bradley JR. Tumour necrosis factor in infectious disease. J Pathol. 2013;230(2):132-47.
106. Liu M, Guo S, Hibbert JM, Jain V, Singh N, Wilson NO, et al. CXCL10/IP-10 in infectious diseases pathogenesis and potential therapeutic implications. Cytokine Growth Factor Rev. 2011;22(3):121-30.
107. Egesten A, Eliasson M, Johansson HM, Olin AI, Morgelin M, Mueller A, et al. The CXC chemokine MIG/CXCL9 is important in innate immunity against Streptococcus pyogenes. J Infect Dis. 2007;195(5):684-93.
108. Huang YJ, Boushey HA. The Sputum Microbiome in Chronic Obstructive Pulmonary Disease Exacerbations. Ann Am Thorac Soc. 2015;12 Suppl 2(Suppl 2):S176-80.
109. Millares L, Perez-Brocal V, Ferrari R, Gallego M, Pomares X, Garcia-Nunez M, et al. Functional Metagenomics of the Bronchial Microbiome in COPD. PLoS One. 2015;10(12):e0144448.
110. Su J, Liu HY, Tan XL, Ji Y, Jiang YX, Prabhakar M, et al. Sputum Bacterial and Fungal Dynamics during Exacerbations of Severe COPD. PLoS One. 2015;10(7):e0130736.
111. Wang Z, Singh R, Miller BE, Tal-Singer R, Van Horn S, Tomsho L, et al. Sputum microbiome temporal variability and dysbiosis in chronic obstructive pulmonary disease exacerbations: an analysis of the COPDMAP study. Thorax. 2018;73(4):331-8.

112. Wang Z, Locantore N, Haldar K, Ramsheh MY, Beech AS, Ma W, et al. Inflammatory Endotype-associated Airway Microbiome in Chronic Obstructive Pulmonary Disease Clinical Stability and Exacerbations: A Multicohort Longitudinal Analysis. Am J Respir Crit Care Med. 2021;203(12):1488-502.
113. Parameswaran GI, Sethi S, Murphy TF. Effects of bacterial infection on airway antimicrobial peptides and proteins in COPD. Chest. 2011;140(3):611-7.
114. Parameswaran GI, Wrona CT, Murphy TF, Sethi S. Moraxella catarrhalis acquisition, airway inflammation and protease-antiprotease balance in chronic obstructive pulmonary disease. BMC Infect Dis. 2009;9:178.
115. Lonergan M, Dicker AJ, Crichton ML, Keir HR, Van Dyke MK, Mullerova H, et al. Blood neutrophil counts are associated with exacerbation frequency and mortality in COPD. Respir Res. 2020;21(1):166.
116. Dicker AJ, Huang JTJ, Lonergan M, Keir HR, Fong CJ, Tan B, et al. The sputum microbiome, airway inflammation, and mortality in chronic obstructive pulmonary disease. J Allergy Clin Immunol. 2021;147(1):158-67.
117. Erb-Downward JR, Thompson DL, Han MK, Freeman CM, McCloskey L, Schmidt LA, et al. Analysis of the lung microbiome in the "healthy" smoker and in COPD. PLoS One. 2011;6(2):e16384.
118. Sze MA, Dimitriu PA, Suzuki M, McDonough JE, Campbell JD, Brothers JF, et al. Host Response to the Lung Microbiome in Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2015;192(4):438-45.
119. Wang Z, Liu H, Wang F, Yang Y, Wang X, Chen B, et al. A Refined View of Airway Microbiome in Chronic Obstructive Pulmonary Disease at Species and Strain-Levels. Front Microbiol. 2020;11:1758.
120. Ramsheh MY, Haldar K, Esteve-Codina A, Purser LF, Richardson M, Muller-Quernheim J, et al. Lung microbiome composition and bronchial epithelial gene expression in patients with COPD versus healthy individuals: a bacterial 16S rRNA gene sequencing and host transcriptomic analysis. Lancet Microbe. 2021;2(7):e300-e10.
121. Sze MA, Dimitriu PA, Hayashi S, Elliott WM, McDonough JE, Gosselink JV, et al. The lung tissue microbiome in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2012;185(10):1073-80.
122. Pragman AA, Kim HB, Reilly CS, Wendt C, Isaacson RE. The lung microbiome in moderate and severe chronic obstructive pulmonary disease. PLoS One. 2012;7(10):e47305.
123. Zakharkina T, Heinzel E, Koczulla RA, Greulich T, Rentz K, Pauling JK, 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. 2013;8(7):e68302.
124. Pragman AA, Lyu T, Baller JA, Gould TJ, Kelly RF, Reilly CS, et al. The lung tissue microbiota of mild and moderate chronic obstructive pulmonary disease. Microbiome. 2018;6(1):7.
125. Mathieu E, Escribano-Vazquez U, Descamps D, Cherbuy C, Langella P, Riffault S, et al. Paradigms of Lung Microbiota Functions in Health and Disease, Particularly, in Asthma. Front Physiol. 2018;9:1168.
126. Opron K, Begley LA, Erb-Downward JR, Freeman C, Madapoosi S, Alexis NE, et al. Lung microbiota associations with clinical features of COPD in the SPIROMICS cohort. NPJ Biofilms Microbiomes. 2021;7(1):14.
127. Molyneaux PL, Mallia P, Cox MJ, Footitt J, Willis-Owen SA, Homola D, et al. Outgrowth of the bacterial airway microbiome after rhinovirus exacerbation of chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2013;188(10):1224-31.

128. Pfeiffer S, Herzmann C, Gaede KI, Kovacevic D, Krauss-Etschmann S, Schloter M. Different responses of the oral, nasal and lung microbiomes to cigarette smoke. Thorax. 2022;77(2):191-5.
129. Huang C, Shi G. Smoking and microbiome in oral, airway, gut and some systemic diseases. J Transl Med. 2019;17(1):225.
130. Johnsen RH, Heerfordt CK, Boel JB, Dessau RB, Ostergaard C, Sivapalan P, et al. Inhaled corticosteroids and risk of lower respiratory tract infection with Moraxella catarrhalis in patients with chronic obstructive pulmonary disease. BMJ Open Respir Res. 2023;10(1).
131. Heerfordt CK, Eklof J, Sivapalan P, Ingebrigtsen TS, Biering-Sorensen T, Harboe ZB, et al. Inhaled Corticosteroids in Patients with Chronic Obstructive Pulmonary Disease and Risk of Acquiring Streptococcus pneumoniae Infection. A Multiregional Epidemiological Study. Int J Chron Obstruct Pulmon Dis. 2023;18:373-84.
132. Keir HR, Contoli M, Chalmers JD. Inhaled Corticosteroids and the Lung Microbiome in COPD. Biomedicines. 2021;9(10).
133. Contoli M, Pauletti A, Rossi MR, Spanevello A, Casolari P, Marcellini A, et al. Long-term effects of inhaled corticosteroids on sputum bacterial and viral loads in COPD. Eur Respir J. 2017;50(4).
134. Singanayagam A, Glanville N, Cuthbertson L, Bartlett NW, Finney LJ, Turek E, et al. Inhaled corticosteroid suppression of cathelicidin drives dysbiosis and bacterial infection in chronic obstructive pulmonary disease. Sci Transl Med. 2019;11(507).
135. Wang Z, Bafadhel M, Haldar K, Spivak A, Mayhew D, Miller BE, et al. Lung microbiome dynamics in COPD exacerbations. Eur Respir J. 2016;47(4):1082-92.
136. Leitao Filho FS, Takiguchi H, Akata K, Ra SW, Moon JY, Kim HK, et al. Effects of Inhaled Corticosteroid/Long-Acting beta(2)-Agonist Combination on the Airway Microbiome of Patients with Chronic Obstructive Pulmonary Disease: A Randomized Controlled Clinical Trial (DISARM). Am J Respir Crit Care Med. 2021;204(10):1143-52.

16. Supplements

Paper I: Tangedal S, Aanerud M, Persson LJ, Brokstad KA, Bakke PS, Eagan TM (2014): Comparison of inflammatory markers in induced and spontaneous sputum in a cohort of COPD patients.

Paper II: Tangedal S, Aanerud M, Gronseth R, Drengenes C, Wiker HG, Bakke PS, Eagan TM (2017): Comparing microbiota profiles in induced and spontaneous sputum samples in COPD patients.

Paper III: Tangedal S, Nielsen R, Aanerud M, Persson LJ, Wiker HG, Bakke PS, Hiemstra PS, Eagan TM (2019) Sputum microbiota and inflammation at stable state and during exacerbations in a cohort of chronic obstructive pulmonary disease (COPD) patients.

S1 Text. More on bioinformatics and statistical methods. *Online supplement for paper III.*

Paper IV: Tangedal S, Nielsen R, Aanerud M, Drengenes C, Husebø G, Lehmann S, Knudsen K, Hiemstra PS, Eagan TM (2023) The lower airway microbiota in COPD and healthy controls.

Supplemental material. *Online supplement for paper IV.*

Tangedal et al. Respiratory Research 2014, 15:138
http://respiratory-research.com/content/15/1/138

RESPIRATORY RESEARCH

RESEARCH Open Access

Comparison of inflammatory markers in induced and spontaneous sputum in a cohort of COPD patients

Solveig Tangedal[1*], Marianne Aanerud[1], Louise JP Persson[1], Karl A Brokstad[2], Per S Bakke[2] and Tomas M Eagan[1,2]

Abstract

Background: Sputum induction is a non-invasive method for obtaining measurements of inflammation in the airways. Whether spontaneously sampled sputum can be a valid surrogate is unknown. The aim of this study was to compare levels of six inflammatory markers in sputum pairs consisting of induced and spontaneous sputum sampled on the same consultation either in a stable state or during exacerbations of chronic obstructive pulmonary disease (COPD).

Methods: 433 COPD patients aged 40–76, Global initiative for chronic Obstructive Lung Disease (GOLD) stage II-IV were enrolled in 2006/07 and followed every six months for three years. 356 patients were followed for potential exacerbations. Interleukin-6, interleukin-8, interleukin-18, interferon gamma-inducible protein-10, monokine induced by gamma interferon and tumor necrosis factor-alpha (IL-6, IL-8, IL-18, IP-10, MIG and TNF-α) were measured by bead based multiplex immunoassay in 60 paired sputum samples from 45 patients. Albumin was measured by enzyme immunoassay, for concentration correction. Culturing for bacterial growth was performed on 24 samples. Bland-Altman plots were used to assess agreement. The paired non-parametric Wilcoxon signed-rank test, the non-parametric Spearman's rank correlation test and Kruskal-Wallis test were used for statistical analyses. For all analyses, a p-value < 0.05 was considered significant.

Results: Agreement between the two measurements was generally low for all six markers. TNF-α was significantly higher in spontaneous sputum at exacerbations ($p = 0.002$) and trending higher at the steady state ($p = 0.06$). Correlation coefficients between the levels of markers in induced and spontaneous sputum varied between 0.58 (IL-18) to 0.83 (IP-10). In spontaneous sputum IL-18 and MIG were higher in ex-smokers ($p < 0.05$). The levels of all markers were higher in GOLD stage III & IV except for IL-6 in spontaneous sputum and IL-18 in induced sputum, compared with GOLD stage II, although not statistically significant. In spontaneous sputum the levels of IL-6 were significantly higher if Haemophilus influenzae (HI) was not cultured.

Conclusion: We observed a low agreement and significant differences in inflammatory markers between induced and spontaneous sputum, both at steady state and exacerbations. We recommend considering sampling method when reporting on inflammatory markers in sputum.

Keywords: COPD, Sputum sampling, Inflammatory markers

* Correspondence: stangedal@gmail.com
[1]Department of Thoracic Medicine, Haukeland University Hospital, Bergen, Norway
Full list of author information is available at the end of the article

Background

Chronic obstructive pulmonary disease (COPD) is a chronic inflammatory disease affecting both the airways and lung parenchyma [1]. The increased airway inflammation has been well described, but its role is yet controversial [2]. Obtaining reliable measurements of airway inflammation non-invasively can enable large cohort studies. Biomarkers sampled by methods like exhaled breath condensate and induced sputum have been compared recently [3]. Induced sputum sampling (ISS) is a non-invasive procedure, which has been standardized and used extensively the last 20 years [4]. Nebulized and inhaled saline increases sputum production in the lungs [4]. Induction has been reported to provide sputum samples of sufficient quality for analyses in more than 80% of asthma and COPD patients [5-8]. In patients with obstructive lung disease, ISS is usually performed in the steady state as it can induce bronchoconstriction [9,10]. However, at least one study has shown that it can be done safely also during exacerbations in patients with mild to moderate COPD [11].

An alternative to ISS is spontaneous sputum sampling (SSS). Levels of inflammatory markers and cell counts in spontaneous and induced sputum have been presented without discriminating between the two sampling methods in some studies [12-14]. Two studies have found that cell viability was higher in induced than spontaneous sputum in patients with asthma or COPD [15,16]. However, few studies have addressed whether induced and spontaneous sputum sampled from patients with COPD can actually be used interchangeably for analyses of inflammatory markers, as it was pointed to in a review article published as late as in 2013 [17]. More studies on the subject were recommended already in 2002 [4].

The aim of this study was to compare the levels of the six common inflammatory markers interleukin 6, 8 & 18 (IL-6, IL-8 IL-18), interferon gamma-inducible protein-10 (IP-10), tumor necrosis factor-alpha (TNF-α) and monokine induced by gamma interferon (MIG) in paired induced and spontaneous sputum samples collected from COPD patients in the stable state and/or during acute exacerbations. These markers were chosen for different roles in airways inflammation in COPD, as part of the analyses in the Bergen COPD Exacerbation Study. In addition, this study allowed for an assessment of the safety of sputum induction in COPD patients undergoing an exacerbation.

Methods and material

Study population

The Bergen COPD Cohort Study (BCCS) was a three year follow-up of 433 COPD patients from western Norway between 2006 and 2010, previously described in detail [18]. The patients were invited to our study centre every six months, and sputum induction was performed at nearly all visits. Of the 433 COPD patients, 356 patients living in a proximity that meant they belonged to the Bergen hospital district were offered concomitant participation in the Bergen COPD Exacerbation Study (BCES). Patients included in the BCES were given a laminated green-card with detailed instructions regarding potential symptoms of COPD exacerbations and a telephone number to our study nurse. The telephone was open 12 hours per day, seven days a week for the three years the study lasted. Once contact had been made, the study nurse determined whether immediate hospitalization was necessary, or whether a visit with a study physician could be scheduled the next working day. During that visit or at the ward the day after hospitalization, sputum induction was attempted if our study physician determined the event to be a clinical COPD exacerbation, with a formal assessment according to Wedzicha and Donaldsons's definition [19].

Spontaneous sputum samples were collected before the induced sputum sample at the same time point at occasions when the patients presented with abundant sputum. In total 60 sputum pairs of acceptable quality from 45 patients in the stable state (n = 31) or during COPD exacerbation (n = 29) were available for analysis. Classification into Global initiative for chronic Obstructive Lung Disease (GOLD 2007) stage and information on smoking habits, were based on the baseline visit in the BCCS. All patients provided written informed consent, and both studies were approved by the Norwegian Regional Ethical Committee.

Sputum sampling and processing

Inductions were performed using an ultrasonic wave nebulizer. Hypertonic saline (3%) was inhaled seven minutes times three, and sputum was attempted sampled after each inhalation. If however, the patient was evaluated by the study physician as being too clinically obstructive, or if the patient did not want to inhale an increased saline concentration, the physiological saline concentration of 0.9% was inhaled instead. Of the 60 sputum pairs evaluated, induction was done with 3% saline in 47 cases, 0.9% in ten cases, while for three inductions the concentration was not recorded. Spirometric evaluations (Vitalograph S-model Vitalograph Ltd., Buckingham, England at regular visits in the steady state, EasyOne model 2001 Ndd Medizintechnik AG, Zurich, Switzerland at exacerbation visits) were performed after inhalation of 200–400 ug salbutamol prior to induction with saline. Spirometry was then repeated after each inhalation of the saline. The procedure ended if FEV_1 declined 20% or more, if the patient's symptoms worsened, or if the patient did not wish to proceed. If the patient's post-bronchodilator oxygen saturation was <90%, induction was not performed.

For the SSS, patients were asked to expectorate in two different cups, and the most purulent sputum was processed. Both types of sputum samples were kept on ice until processed for quality control and storage, usually within 30 minutes. To break disulphide bonds in mucin, 4 ml dithiothreitol 0.1% (DDT) per gram sputum were added [20]. The samples were then homogenized using an Eppendorf homogenizer at 600 rpm for 15 minutes at a temperature of 4 degrees Celsius. Phosphate-buffered saline (PBS) was added, and the sample filtered to increase homogenization. Supernatants were removed after 15 minutes centrifugation at 4 degrees Celsius, 450 *g*, aliqouted in 0.5 ml tubes, and stored at −80 degrees Celsius. Trained personnel evaluated viability after staining with tryptan blue. For the sputum samples to be considered of acceptable quality there had to be > 1 million/mL cells, < 20% epithelial cells and the leucocyte viability had to be > 30%. After December 2006, all sputum samples were also cultured at the Department of Microbiology, Haukeland University Hospital.

The sputum samples were analysed for cytokines using the Luminex® xMAP® technology (Luminex Corporation, Austin, Texas). The cytokine assay used was made by combining standards from BioRad (Bio-Plex Pro Human Cytokine Standards Group I 27-Plex #171-D50001, Lot No 5022130. Bio-Plex Pro Human Cytokine Standards Group II 23-Plex #171-D10502 Lot No 5015357) and singleplex assays containing beads for analyses of IL-6, IL-8 IL-18, IP-10, TNF-α and MIG. Thus, all six markers were analyzed in simplex. The samples were processed on a Luminex 100 instrument and the results collected and stored by STarStation software version 2.0 (STarStation Software Version 2.0, Applied Cytometry, Sheffield, UK.) The procedure was performed according to the manufacturer's instructions on six separate days in September 2011.

For 58 of the 60 sputum pairs we also had enough material to perform an enzyme immunoassay of levels of albumin in duplex (Albumin Human ELISA kit, ab 108788, Abcam, Cambridge, UK). Albumin was used as a correction factor for concentration differences between the induced and spontaneous sample for each pair in the following way: The induced to spontaneous albumin ratio was calculated for each sputum pair, and the level of each of the six markers in each of the spontaneous sputum samples multiplied by the corresponding ratio. All later statistical analyses were performed both on "corrected" sputum levels and "uncorrected" sputum levels.

Statistical analyses

Stata 12.0 was used for the statistical analyses (StataCorp. College Station, Texas). Bland-Altman plots were made to assess agreement between the measured levels of the markers in induced and spontaneous sputum pairs. Bland & Altman advocates using the difference between the two measurements as the central measurement of bias, and the spread of the difference as a measure of limits of agreement [21]. Usually the difference between the measurements is plotted against the mean of the two measurements, with 2 standard deviations (SD) of the difference representing the 95% limits of agreement. However, sometimes the difference is dependent upon the size of the mean, in which Bland & Altman advocates plotting on a log scale [22]. This was the case for all six markers in our study.

The inflammatory markers were not normally distributed, hence the paired non-parametric Wilcoxon signed-rank test was used to compare levels of the markers and cell viability between spontaneous and induced sputum. For correlation analyses between spontaneous and induced samples the non-parametric Spearman's rank correlation test was used. For comparisons of the levels of inflammatory markers by clinical characteristics, Kruskal-Wallis test was used. For comparisons of FEV_1 decline between stable state and exacerbations during inductions, Wilcoxon signed-rank test was used. For all analyses, a p-value of less than 0.05 was considered significant.

Results

The characteristics of the study population are presented in Table 1. 60 sputum pairs were available from 45 patients, of which 15 of the patients were women. Of the 60 sputum pairs, 31 were sampled during the stable state and 29 during COPD exacerbations (Table 1).

Mean cell viability was 98% for both the induced and spontaneous sputum samples. Among the induced samples, 2 out of 60 samples had viability below 90%, for the spontaneous samples all were 90% viable or better.

Of the six inflammatory markers, TNF-α was significantly higher when measured in spontaneous sputum during exacerbations and almost reaching statistical significance in the steady state (Table 2). For the other markers, no clear trend was seen (Table 2).

Bland-Altman plots for all six inflammatory markers on the log scale are presented in Figure 1. To obtain the limits of agreement the antilog of the two standard deviations were calculated, and these are presented in Table 3 together with the Spearman's rank correlation coefficients. Although the correlation was fair, varying between 0.58 for IL-18 to 0.83 for IP-10, the agreement was quite low for all six inflammatory markers. Since the 95% limits of agreement were calculated on the log scale, the upper and lower limits represents ratios relative to one. Thus, based on the calculations presented in Table 3, one would expect the measurement of for instance IL-6 in spontaneous sputum to fall between 6 times higher or 8 times lower than that measured in induced sputum 95% of the time.

Even though agreement between individual measurements was low, there could be value to the spontaneous samples if the levels of the markers showed the same associations to clinical parameters in spontaneous as in the

Table 1 Characteristics of the study population

	n	%
Patients	45	
Age, mean (range)[*]	63.4 (46–74)	
Sex		
Women	15	33
Men	30	67
Smoking habits		
Ex	30	67
Current	15	33
GOLD (2007) stage[*]		
II	15	33
III	24	53
IV	6	13
Patients with one sputum pair[**]	36	80
Patients with multiple sputum pairs	9	20
Sputum pairs	60	
stable state	31	52
during exacerbation	29	48
H.influenza positive[†]		
No	12	
Yes	12	

[*]At inclusion.
[**]Consisting of one spontaneous and one induced sputum sample.
[†]Detected in induced and/or spontaneous sputum sampled from stable state visits, and/or exacerbations.

induced sputum samples. Potential associations between measured levels of the inflammatory markers in spontaneous and induced samples, and clinical variables are presented in Table 4. There was no consistent difference in levels of any of the six markers between current and ex-smokers. However, in the spontaneous samples the measured levels of IL-18 and MIG were significantly higher in ex-smokers, an association not found in the induced samples. For all markers except IL-6 and IL-18, there was a non-significant trend of higher levels in GOLD stage III & IV compared with GOLD stage II. Importantly however, the pattern was the same for both spontaneous and induced sputum samples. Finally, in the 24 sputum pairs where culture was obtained, we examined which impact Haemophilus influenzae (HI) had on the pattern of the sputum markers. In the spontaneous samples HI was associated with significantly lower levels of IL-6, a difference not found in the induced samples. In addition, we observed that in the spontaneous samples levels of MIG were lower in sputum with HI, whereas the opposite pattern was seen in the induced samples (Table 4).

To assess the safety of induction during exacerbations and the stable state we calculated the decline in FEV_1% predicted during induction for all COPD patients who underwent inductions both in the BCCS and BCES. For decline in FEV_1% predicted from post bronchodilation values during induction the relative fall was calculated (thus a fall from 30% predicted to 20% predicted will be presented as a 33% decline). To avoid repeated measurements from the same patient at steady state and/or at exacerbations only one registered induction at the two different disease states was selected for analyses per patient. 63 patients were induced during exacerbation. 33 of the patients were GOLD stage III or IV, while the remaining 30 were GOLD stage II. We found no significant difference in FEV_1% predicted decline caused by induction related to disease severity ($p = 0.07$) during exacerbations. When comparing patient groups in the stable state we found that patients with more severe COPD had a statistically larger decline related to induction, than patients with COPD GOLD stage II ($p < 0.001$). The relative fall was significantly higher during the stable state than during exacerbations ($p = 0.03$) (Table 5). However, no adverse events followed inductions regardless of disease state and severity, and all patients increased in FEV_1 after a rest period and a new inhalation of salbutamol.

Table 2 A comparison of inflammatory markers in induced and spontaneous sputum sampled from the COPD patients at the time; either during a COPD exacerbation or during the stable state

	During a COPD exacerbation n = 28			During the stable state n = 30		
	Induced sputum sample	Spontaneous sputum sample	p*	Induced sputum sample	Spontaneous sputum sample	p*
IL-6(pg/ml) median, IQR	10.0(4.9-26.6)	10.4(2.8-23.3)	0.29	20.5(6.11-43.2)	13.9(2.2-55.2)	0.46
IL-8(pg/ml) median, IQR	339.7(193.6-663.9)	344.9(173.9-812.2)	0.77	514.2(225.6-1257.2)	3709.9(171.7-976.7)	0.18
IP-10(pg/ml)	735.0(205.4-2099.6)	372.1(218.7-1416.2)	0.15	529.0(215.4-2554.8)	362.3(142.0-1393.1)	0.07
TNF-α (pg/ml) median, IQR	3.0(0.2-11.8)	6.3(2.0-34.0)	0.002	0.9(0-2.4)	1.3(0.2-4.3)	0.06
IL-18 (pg/ml) median, IQR	9.2(4.9-12.9)	10.6(3.6-27.3)	0.52	6.5(2.4-25.6)	9.6(0.8-18.2)	0.64
MIG (pg/ml) median, IQR	539.2(100.1-1496.0)	384.6(221.2-1997.0)	0.41	534.4(56.9-1450.1)	567.4(178.0-2031.7)	0.82

[*]Wilcoxon sign rank test.

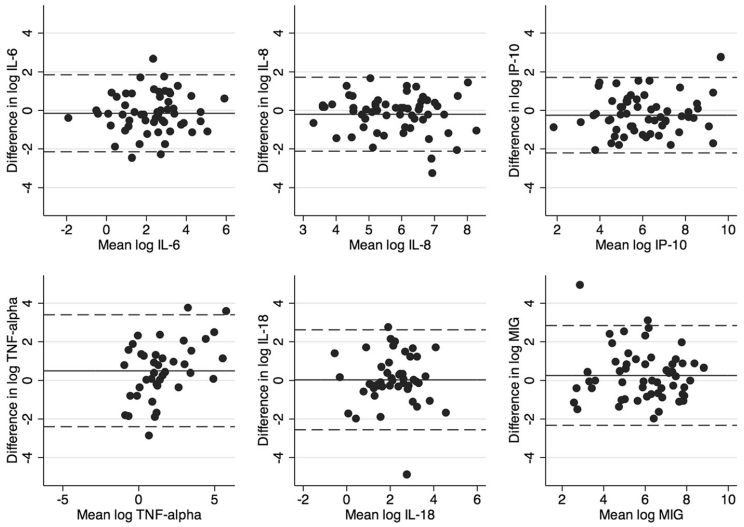

Figure 1 Bland-Altman plots of the agreement between measurements of six inflammatory markers in induced and spontaneous sputum samples.

Discussion

This study showed that for the six inflammatory markers, the correlation between levels measured in induced and spontaneous sputum pairs was fair, but the agreement was quite low. TNF-α was significantly higher in spontaneous sputum samples than in induced samples when measured during a COPD exacerbation. Further, there was a relationship between HI carrier state and IL-6, and smoking status and IL-18 and MIG, found only in spontaneous sputum samples.

There are some methodological issues to consider. Firstly, it has been shown that both PBS and DTT affect the recovery of some cytokines [12,23]. However, a strength of this study was that the exact same processing protocol was used for all sputum pairs, and this should thus not impact the measured levels differently between spontaneous and induced sputum samples. Secondly, all the inflammatory markers were measured in simplex, thus the potential measurement error is greater than if the markers were measured in duplex. The choice of analysing in simplex was due to cost, since this is part of a larger analysis of inflammatory markers in sputum. Most importantly however, all sputum pairs were analysed on the same plate, on the same day. Thus the measurement error should not differ between spontaneous and induced samples. Thirdly, we found associations between inflammatory markers and smoking, and inflammatory markers and colonization with HI only in spontaneous sputum. We found no association between inflammatory markers and GOLD stage in either type of sputum, but

Table 3 Rank correlation coefficients and the 95% limits of agreement between measurements of six inflammatory markers in induced and spontaneous sputum samples

	Correlation coefficient*	Bland & Altmans 95% limit of agreement†	
		Lower	Upper
IL-6 (pg/ml)	0.729	0.12	6.35
IL-8 (pg/ml)	0.695	0.12	5.59
IP-10 (pg/ml)	0.833	0.08	13.7
TNFα (pg/ml)	0.600	0.11	5.53
IL-18 (pg/ml)	0.583	0.09	29.99
MIG (pg/ml)	0.754	0.10	17.24

*Spearman's rank correlation test.
†±2 SD of the mean difference between the two measurements.

Table 4 Median (IQR) values of inflammatory markers in induced and spontaneous sputum samples by smoking, GOLD stage and H. influenzae carrier state

	Smoking status			GOLD stage			H. influenzae		
	Current smoker n = 15	Ex-smoker n = 29	p*	II n = 15	III+IV n = 29	p*	No n-12	Yes = n = 12	p*
IL-6(pg/ml)									
induced	17.2(5.7-36.2)	16.7(4.0-43.2)	0.63	16.7(4.5-67.8)	19.4(5.2-36.2)	0.78	20.7(10.9-49.0)	17.8(1.3-28.6)	0.16
spontaneous	10.2(3.0-29.5)	21.5(5.8-48.2)	0.22	15.6(2.2-64.9)	13.8(4.45.8)	0.79	32.2(22.4-75.38)	12.0(2.6-18.8)	0.01
IL-8(pg/ml)									
induced	505.4(161.3-805.9)	517.0(193.3-1257.2)	0.54	233.4(170.0-1257.2)	529.0(224.5-1156.0)	0.37	277.6(170.0-649.0)	655.49(274.70-832.70)	0.32
spontaneous	221.0(165.4-674.4)	405.0(173.2-891.8)	0.42	210.1(112.7-844.1)	546.3(174.7-891.8)	0.18	336.9(210.1-734.2)	570.4(265.8-1141.4)	0.23
IP-10(pg/ml)									
induced	216.5(130.9-826.2)	624.3(289.9-1844.9)	0.08	232.5(175.2-748.3)	600.8(216.7-1844.9)	0.20	477.5(175.2-2554.8)	696.1(56.0-1574.1)	0.85
spontaneous	345.6(124.9-716.5)	461.0(237.5-1361.7)	0.14	354.5(129.7-780.1)	448(237.5-1202.0)	0.33	491.9(285.2-2269.7)	582.7(180.0-1093.0)	0.42
TNF-α(pg/ml)									
induced	0.4(0-1.9)	1.5(0-4.8)	0.34	0.00(0.00-2.5)	1.6(0-4.2)	0.14	2.3(0.0-5.1)	3.2(0.7-24.3)	0.24
spontaneous	0.9(0-3.2)	3.3(0.3-7.4)	0.20	0.5(0.0-4.5)	2.4(0.8-7.4)	0.18	6.43.7-14.0	7.9(2.1-173.9)	0.69
IL-18(pg/ml)									
induced	9.0(0.9-19.3)	8.4(4.1-25.6)	0.58	10.6(2.4-14.9)	6.3(4.1-25.6)	0.91	5.04(0.0-13.6)	4.2(0.6-10.1)	1.0
spontaneous	2.9(0.5-14.8)	14.8(8.3-31.7)	0.01	8.8(1.0-15.1)	14.6(3.5-29.8)	0.24	10.4(5.0-13.9)	7.0(2.7-25.2)	0.45
MIG(pg/ml)									
induced	332.2(30.1-661.7)	806.3(110.5-1391.4)	0.15	121.8(41.2-897.0)	661.7(110.5-1314.3)	0.27	157.8(43.3-1714.8)	677.1(179.0-961.8)	0.54
spontaneous	208.5(31.7-385.8)	626.8(222.6-2031.7)	0.03	280.4(69.8-1537.1)	383.4(205.3-1941.6)	0.47	1473.0 (438.8-4392.9)	740.5(384.6-2088.3)	0.33

*Kruskal-Wallis test.

Table 5 Relative FEV_1 decline in % predicted during sputum induction

	Exacerbations n = 63	Steady state n = 390	p*
			0.004
Median (IQR)	12.64(5.56-21.79)	18.75(11.11-25)	
Mean (SD)	14.80(13.05)	18.51(11.44)	

*Wilcoxon sign rank test.

this may be due to lack of strength. Finally, the choice on whether to induce or not during an exacerbation was based on several subjective factors in addition to oxygen saturation; most importantly patients' willingness to be induced and the clinicians' evaluation regarding obstructivity. Thus, it is impossible from this design to conclude that sputum induction would be safe during all exacerbations.

Although more studies on the subject of whether spontaneous and induced sputum samples could be compared was recommended already in 2002 [4], few studies have yet been published. We have found one earlier report on levels of IL-8 in spontaneous versus induced sputum that showed no significant differences in IL-8 levels between the two sputum types in COPD patients in stable state [16]. Our study confirmed the results from this earlier study, but in addition we were able to show that this is true also during exacerbations. We have been unable to find earlier reports on the relationship between levels of inflammatory markers in spontaneous and induced sputum for the remaining five inflammatory markers. To our knowledge comparison of other inflammatory markers in induced and spontaneous sputum sampled on the same consultation has not been performed in patients with obstructive pulmonary disease.

It has previously been shown that the sputum sampled early during induction has a different consistency and cell composition than sputum sampled late in the induction [24,25]. It is likely that more central airways are sampled early, and would thus most resemble spontaneous sputum. Thus, induced sputum is likely to sample a more distal airways environment than spontaneous sputum. Central and distal airways differ by epithelial components [26], distribution of immune cells [27,28], and possibly respiratory microbiome [29]. Thus, it is theoretically rather likely that levels of inflammatory markers differ between spontaneous and induced sputum samples. However, one can argue that spontaneous sputum could be a favorable alternative to induced sputum when patients find induction uncomfortable, or the safety of the induction is uncertain, and enable sampling in primary healthcare settings where induction is rarely if ever performed to our knowledge. Cell viability in spontaneous sputum has in some studies been shown to be poorer than in induced sputum samples [15,16]. Such was not the case in our samples, where viability was as good in the spontaneous samples as in the induced samples. In our study the time from collection to processing was usually very short, which could explain the high viability.

Although agreement for individual measurements was low, measuring levels of inflammatory markers in spontaneous sputum could have value for instance in serial measurements of spontaneous sputum, something our study is not equipped to assess. Also, although comparisons of inflammatory markers between spontaneous and induced sputum is invalid for some markers, they may be valid for others.

There are still sparse data on the safety on induction in patients with severe COPD during exacerbations, and in several studies sputum induction is performed during exacerbation without the published reporting on potential adverse effects on the procedure [3,30,31].

In our study we found statistical differences in FEV_1% predicted decline between patients with moderate and severe/very severe COPD only during steady state, while disease severity did not affect the decline during exacerbations. No adverse events were registered during either the steady state or during exacerbations. This is in accordance with other reports [7,11], but we expand by including patients with severe/very severe COPD. However, it should be stressed that necessary precautions need to be taken such as having access to acute rescue medications, and that all inductions only should be performed by trained medical personnel [32].

The results from the current study point toward a necessity for reporting on sampling methods when considering inflammatory markers in sputum samples collected from COPD patients both during the steady state and during acute exacerbations as the agreement was generally low as assessed by Bland & Altman's 95% limits of agreement. Whether levels of inflammatory markers can be compared between spontaneous and induced sputum samples likely differ by each inflammatory marker in question, and should be addressed within each study. In cases where induced sputum sampling is impossible, spontaneous samples may have value if compared with other spontaneous samples.

Abbreviations

BCCS: Bergen COPD cohort study; BCES: Bergen COPD exacerbation study; COPD: Chronic obstructive pulmonary disease; DDT: Dithiothreitol; GOLD: Global initiative for chronic obstructive lung disease; HI: Haemophilus influenza; IL-6: Interleukin 6; IL-8: Interleukin 8; IL-18: Interleukin 18; IP-10: Interferon gamma-inducible protein-10; MIG, Monokine induced by gamma interferon; ISS: Induced sputum sampling; PBS: Phosphate-buffered saline; SSS: Spontaneous sputum sampling; TNF-α: Tumornecrosis factor-alpha.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions
Study concept and design: ST, TME, MA, PSB. Acquisition of data: ST, MA, LJPP, PSB, TME. Analysis and interpretation of data: ST, MA, KAB, TME. Drafting of the manuscript: ST, TME. Critical revision of the manuscript for important intellectual content: MA, LJPP, KAB, PSB. All authors read and approved of the final manuscript.

Acknowledgements
The authors wish to thank study nurse Margrete Klemmetsby, physiotherapist Marie Waatevik, microbiologist Dag Harald Skutlaberg, bioengineers Lene Svendsen and Tina Endresen-Vinsjevik for invaluable support during data collection, and Rune Grønseth for help with the Bland Altman analyses.

Author details
[1]Department of Thoracic Medicine, Haukeland University Hospital, Bergen, Norway. [2]Institute of Clinical Science, Faculty of Medicine, University of Bergen, Bergen, Norway.

Received: 14 April 2014 Revised: 26 September 2014
Accepted: 24 October 2014
Published online: 15 November 2014

References

1. **Global Strategy for the Diagnosis, Management and Prevention of COPD, Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2014.** [http://www.goldcopd.com/guidelines-global-strategy-for-diagnosis-management.html]
2. Brusasco V, Crimi E, Pellegrino R: **Airway inflammation in COPD: friend or foe?** *Am J Respir Crit Care Med* 2007, **176:**425–426.
3. Warwick G, Thomas PS, Yates DH: **Non-invasive biomarkers in exacerbations of obstructive lung disease.** *Respirology* 2013, **18:**874–884.
4. Paggiaro PL, Chanez P, Holz O, Ind PW, Djukanovic R, Maestrelli P, Sterk PJ: **Sputum induction.** *Eur Respir J Suppl* 2002, **37:**3s–8s.
5. Dragonieri S, Tongoussouva O, Zanini A, Imperatori A, Spanevello A: **Markers of airway inflammation in pulmonary diseases assessed by induced sputum.** *Monaldi Arch Chest Dis* 2009, **71:**119–126.
6. Wark PA, Simpson JL, Hensley MJ, Gibson PG: **Safety of sputum induction with isotonic saline in adults with acute severe asthma.** *Clin Exp Allergy* 2001, **31:**1745–1753.
7. Gao P, Gibson PG, Zhang J, He X, Hao Y, Li P, Liu H: **The safety of sputum induction in adults with acute exacerbation of COPD.** *Clin Respir J* 2013, **7:**101–109.
8. Tsoumakidou M, Tzanakis N, Siafakas NM: **Induced sputum in the investigation of airway inflammation of COPD.** *Respir Med* 2003, **97:**863–871.
9. Makris D, Tzanakis N, Moschandreas J, Siafakas NM: **Dyspnea assessment and adverse events during sputum induction in COPD.** *BMC Pulm Med* 2006, **6:**17.
10. Taube C, Holz O, Mucke M, Jorres RA, Magnussen H: **Airway response to inhaled hypertonic saline in patients with moderate to severe chronic obstructive pulmonary disease.** *Am J Respir Crit Care Med* 2001, **164:**1810–1815.
11. Bathoorn E, Liesker J, Postma D, Koeter G, van Oosterhout AJ, Kerstjens HA: **Safety of sputum induction during exacerbations of COPD.** *Chest* 2007, **131:**432–438.
12. Bafadhel M, McCormick M, Saha S, McKenna S, Shelley M, Hargadon B, Mistry V, Reid C, Parker D, Dodson P, Jenkins M, Lloyd A, Rugman P, Newbold P, Brightling CE: **Profiling of sputum inflammatory mediators in asthma and chronic obstructive pulmonary disease.** *Respiration* 2012, **83:**36–44.
13. Wallace J, D'Silva L, Brannan J, Hargreave FE, Kanaroglou P, Nair P: **Association between proximity to major roads and sputum cell counts.** *Can Respir J* 2011, **18:**13–18.
14. Ronchi MC, Piragino C, Rosi E, Amendola M, Duranti R, Scano G: **Role of sputum differential cell count in detecting airway inflammation in patients with chronic bronchial asthma or COPD.** *Thorax* 1996, **51:**1000–1004.
15. Pizzichini MM, Popov TA, Efthimiadis A, Hussack P, Evans S, Pizzichini E, Dolovich J, Hargreave FE: **Spontaneous and induced sputum to measure indices of airway inflammation in asthma.** *Am J Respir Crit Care Med* 1996, **154:**866–869.
16. Bhowmik A, Seemungal TA, Sapsford RJ, Devalia JL, Wedzicha JA: **Comparison of spontaneous and induced sputum for investigation of airway inflammation in chronic obstructive pulmonary disease.** *Thorax* 1998, **53:**953–956.
17. Koutsokera A, Kostikas K, Nicod LP, Fitting JW: **Pulmonary biomarkers in COPD exacerbations: a systematic review.** *Respir Res* 2013, **14:**111.
18. Eagan TM, Ueland T, Wagner PD, Hardie JA, Mollnes TE, Damas JK, Aukrust P, Bakke PS: **Systemic inflammatory markers in COPD: results from the Bergen COPD Cohort Study.** *Eur Respir J* 2010, **35:**540–548.
19. Seemungal TA, Donaldson GC, Bhowmik A, Jeffries DJ, Wedzicha JA: **Time course and recovery of exacerbations in patients with chronic obstructive pulmonary disease.** *Am J Respir Crit Care Med* 2000, **161:**1608–1613.
20. Efthimiadis A, Spanevello A, Hamid Q, Kelly MM, Linden M, Louis R, Pizzichini MM, Pizzichini E, Ronchi C, Van Overvel F, Djukanovic R: **Methods of sputum processing for cell counts, immunocytochemistry and in situ hybridisation.** *Eur Respir J Suppl* 2002, **37:**19s–23s.
21. Bland JM, Altman DG: **Statistical methods for assessing agreement between two methods of clinical measurement.** *Lancet* 1986, **1:**307–310.
22. Bland JM, Altman DG: **Measuring agreement in method comparison studies.** *Stat Methods Med Res* 1999, **8:**135–160.
23. Woolhouse IS, Bayley DL, Stockley RA: **Effect of sputum processing with dithiothreitol on the detection of inflammatory mediators in chronic bronchitis and bronchiectasis.** *Thorax* 2002, **57:**667–671.
24. Holz O, Jorres RA, Koschyk S, Speckin P, Welker L, Magnussen H: **Changes in sputum composition during sputum induction in healthy and asthmatic subjects.** *Clin Exp Allergy* 1998, **28:**284–292.
25. Gershman NH, Liu H, Wong HH, Liu JT, Fahy JV: **Fractional analysis of sequential induced sputum samples during sputum induction: evidence that different lung compartments are sampled at different time points.** *J Allergy Clin Immunol* 1999, **104:**322–328.
26. Rackley CR, Stripp BR: **Building and maintaining the epithelium of the lung.** *J Clin Invest* 2012, **122:**2724–2730.
27. Moodley YP, Krishnan V, Lalloo UG: **Neutrophils in induced sputum arise from central airways.** *Eur Respir J* 2000, **15:**36–40.
28. O'Donnell R, Breen D, Wilson S, Djukanovic R: **Inflammatory cells in the airways in COPD.** *Thorax* 2006, **61:**448–454.
29. Erb-Downward JR, Thompson DL, Han MK, Freeman CM, McCloskey L, Schmidt LA, Young VB, Toews GB, Curtis JL, Sundaram B, Martinez FJ, Huffnagle GB: **Analysis of the lung microbiome in the "healthy" smoker and in COPD.** *PLoS One* 2011, **6:**e16384.
30. Perera WR, Hurst JR, Wilkinson TM, Sapsford RJ, Mullerova H, Donaldson GC, Wedzicha JA: **Inflammatory changes, recovery and recurrence at COPD exacerbation.** *Eur Respir J* 2007, **29:**527–534.
31. Kersul AL, Iglesias A, Rios A, Noguera A, Forteza A, Serra E, Agusti A, Cosio BG: **Molecular mechanisms of inflammation during exacerbations of chronic obstructive pulmonary disease.** *Arch Bronconeumol* 2011, **47:**176–183.
32. Pizzichini E, Pizzichini MM, Leigh R, Djukanovic R, Sterk PJ: **Safety of sputum induction.** *Eur Respir J Suppl* 2002, **37:**9s–18s.

doi:10.1186/s12931-014-0138-6
Cite this article as: Tangedal *et al.*: Comparison of inflammatory markers in induced and spontaneous sputum in a cohort of COPD patients. *Respiratory Research* 2014 **15**:138.

Tangedal et al. Respiratory Research (2017) 18:164
DOI 10.1186/s12931-017-0645-3

Respiratory Research

RESEARCH Open Access

Comparing microbiota profiles in induced and spontaneous sputum samples in COPD patients

Solveig Tangedal[1,2*], Marianne Aanerud[1], Rune Grønseth[1], Christine Drengenes[1], Harald G. Wiker[2,3], Per S. Bakke[2] and Tomas M. Eagan[1,2*]

Abstract

Background: Induced and spontaneous sputum are used to evaluate the airways microbiota. Whether the sputum types can be used interchangeably in microbiota research is unknown. Our aim was to compare microbiota in induced and spontaneous sputum from COPD patients sampled during the same consultation.

Methods: COPD patients from Bergen, Norway, were followed between 2006/2010, examined during the stable state and exacerbations. 30 patients delivered 36 sample pairs. DNA was extracted by enzymatic and mechanical lysis methods. The V3-V4 region of the 16S rRNA gene was PCR-amplified and prepared for paired-end sequencing. Illumina Miseq System was used for sequencing, and Quantitative Insights Into Microbial Ecology (QIIME) and Stata were used for bioinformatics and statistical analyses.

Results: Approximately 4 million sequences were sorted into 1004 different OTUs and further assigned to 106 different taxa. Pair-wise comparison of both taxonomic composition and beta-diversity revealed significant differences in one or both parameters in 1/3 of sample pairs. Alpha-diversity did not differ. Comparing abundances for each taxa identified, showed statistically significant differences between the mean abundances in induced versus spontaneous samples for 15 taxa when disease state was considered. This included potential pathogens like *Haemophilus* and *Moraxella*.

Conclusion: When studying microbiota in sputum samples one should take into consideration how samples are collected and avoid the usage of both induced and spontaneous sputum in the same study.

Keywords: COPD, Sputum, Microbiota, High-throughput sequencing

Background

Chronic obstructive pulmonary disease (COPD) is characterized by a chronic inflammation of the lower airways, dominated by an influx of innate immune cells. Recent marker-gene studies indicate the existence of a pulmonary microbial flora (microbiota) present in both health and disease [1]. The chronic inflammation seen in COPD might be a consequence of a disrupted equilibrium between the pulmonary microbiota and the innate immune system. To explore this hypothesis, accurate measurements of the microbiota during both stable state and acute exacerbation of chronic obstructive pulmonary disease (AECOPD) are necessary.

The emerging gold standard for exploring the microbiota in the lower airways with minimal oral contamination is through bronchoscopy, but this is impossible during most AECOPD. Collecting induced sputum samples (ISS) is therefore a standardized sampling method of choice [2]. However, in several studies spontaneous sputum samples (SSS) have also been used since they are easier to retrieve [3, 4]. The validity of SSS with regard to microbiota studies is uncertain to date. Two previous studies have compared the microbial composition in ISS and SSS samples from cystic fibrosis (CF) patients, finding comparable results between the two methods [5, 6]. However, CF patients usually produce more sputum spontaneously, have a relatively high biomass in the

* Correspondence: stangedal@gmail.com; tomas.eagan@uib.no
[1]Department of Thoracic Medicine, Haukeland University Hospital, Bergen, Norway
Full list of author information is available at the end of the article

airways, and one of the cited studies used an earlier method of bacterial profiling (Terminal Restriction Fragment Length Polymorphism Profiling) [6], wheras the other had only 15 sputum pairs [5]. The validity of SSS with regard to 16S rRNA marker-gene based studies on non-CF patients is unknown to date.

The Bergen COPD Cohort Study (BCCS) and its adjunct Bergen COPD Exacerbation Study (BCES) offers an opportunity to address this issue in COPD patients as we have sampled sputum both induced and spontaneously in a number of our COPD patients repeatedly during follow-up. We have previously shown that levels of inflammatory markers differed between sputum types in a study from the same population [7]. In the present study we compared the taxonomic composition and diversity measures in 36 sputum pairs consisting of SSS and ISS sampled sequencially from COPD patients either during AECOPD or at the stable state.

Methods

Study design

The current study sample consisted of 36 sputum pairs collected from 30 COPD patients who participated in both the BCCS & BCES. The study design and sampling of the BCCS [8] and the BCES [7] has previously been described in detail. The COPD patients had a smoking history of ≥10 pack-years, and a post-bronchodilation FEV_1/FVC ratio < 0.7 and FEV_1 < 80% predicted. Active autoimmune diseases or cancer within the last 5 years were cause for non-inclusion. A study physician examined and undertook a structured interview of all patients upon inclusion and at half-yearly follow-up visits when the patients were in the stable state.

Patients were instructed to contact the study-staff at periods with worsening of symptoms (malaise, fever, airway symptoms). The study physician offered a clinical examination at the outpatient clinic, Dept. of Thoracic Medicine, Haukeland University Hospital within 24 h of contact, or on the first working day after the weekend. Hospitalized patients were examined by a study physician the first day after admission.

The study was approved by the regional ethical board (REK-Vest), case number 165.08.

Sputum sampling and processing

Both sputum sampling and immediate processing have been described in detail [7]. SSS was collected first from patients expectorating. If the patient's clinical state allowed it, induction with hypertonic saline (3%) was performed. Patients inhaled the saline for 7 min three times, and sputum was collected and pooled after each inhalation. Spirometric evaluations were performed before and after each inhalation during induction (Vitalograph S-model Vitalograph Ltd., Buckingham, England at regular visits in the steady state, EasyOne model 2001 Ndd Medizintechnik AG, Zurich, Switzerland at exacerbation visits). Sputum samples were kept on ice until undergoing quality control less than 30 min after sampling. For the sputum samples to be considered of acceptable quality there had to be >1 million/mL cells, <20% epithelial cells and the leucocyte viability had to be >30%. If the samples were of sufficient quality, they were further treated by standard protocol [7] to separate the supernatant from the cell pellet. All materials were aliqouted and frozen at −80 °C.

DNA extraction and 16S rRNA sequencing

The samples were thawed and treated with sputasol (Oxoid). They underwent pre-lysis with Lysostaphin (4000 U/mL), Lysozyme (10 mg/mL) and Mutanolysin (25,000 U/mL) (Sigma-Aldrich). To avoid shearing of free DNA each sample was centrifuged and supernatants and pellets separated. The pellets underwent mechanical and chemical lysis using the FastPrep-24 Instrument and reagents from the FastDNA Spin Kit (MP Biomedicals, LLC, Solon, OH, USA). Lysates and supernatants from each sample were recombined and the extracted DNA was further purified using the FastDNA Spin Kit. Library preparation and sequencing of the V3-V4 region of the 16S rRNA gene was carried out according to the protocol for Metagenomic Sequencing Library Preparation for the Illumina Miseq System (Part # 15044223 Rev. B, MiSeq Reagent Kit v3). Amplicon PCR was carried out with a total of 45 cycles and followed by Index PCR using primers from the Nextera XT Index Kit (Illumina Inc., San Diego. CA, USA). Pooled, normalized samples went through 2 × 300 cycles of paired-end sequencing. Each of the sample pairs were processed on the same day, and for all pairs we used the same reagent kits throughout DNA extraction, PCR and sequencing.

Bioinformatics analyses

FASTQ-files were computed using Quantitative Insights Into Microbial Ecology (QIIME) v.1.9.1 [9, 10]. First, forward and reverse reads were assembled, after which sequences that did not pass quality demands as advised by QIIME were removed [11]. The accepted sequences were clustered into operational taxonomic units (OTUs) through open reference OTU-picking using uclust v.1.2.22 [12] and the GreenGenes Database v.13_08 [13]. The latter was also used for taxonomic assignment with analyses performed on GreenGenes taxonomic level 6 (genus). The clustering was based on sequence similarity with a threshold of 97%, which is considered the conventional cut-off for 16S rRNA maker-gene surveys and representative for bacterial species [14]. For each OTU a representative sequence was aligned using PyNAST v.1.2.2 [15], and sequences not successfully aligned were omitted from further analyses. A phylogenetic tree was built using FastTree v.2.1.3 [16]. Counts of observations

(OTUs) on a per-sample basis were stored in Biological Observation Matrix (BIOM) format and OTUs containing less than 0.005% of the total number of sequences were removed according to QIIME guidelines [10, 11].

Statistical analyses

Comparisons of the taxonomic distribution between pairs were performed both by calculating the Yue-Clayton measure of dissimilarity (1-θ_{YC} - range 0 to 1; 0 indicates perfect similarity, 1 perfect dissimilarity) [17], and using limits of agreement (LOA) calculated from Bland-Altman plots [18, 19]. Both methods allow evaluation of quantitative differences within each pair.

The mean number of sequences allocated to each identified taxa in the 36 ISS was compared to that found in the 36 SSS, using log-likelihood ratio tests with Bonferroni corrected *p*-values due to multiple comparisons. The comparisons were made between samples normalized through rarefaction with random subsampling without replacement. Comparisons of alpha- and beta-diversity were performed on rarefied OTU-tables [20] with available statistical analyses incorporated in QIIME-scripts. Alpha-diversity (within-sample diversity) was estimated using Faith's phylogenetic diversity, Chao1 and counts of observed OTUs. Beta-diversity is a measure of diversity between samples. To evaluate differences in phylogenetic, quantitative beta-diversity pair-wise, weighted UniFrac (WUF) significance tests were applied [21]. All 72 samples were compared generating 2556 comparisons, for which Bonferroni corrected *p*-values were used.

Principal coordinates analysis (PCoA)-plots of WUF distances between sampling methods were used for visualization of distances in three-dimensional space using Procrustes analyses and transformations of principal coordinates 1-3 [22]. Analyses of similarities (ANOSIM), were used to compare differences in beta-diversity between ISS and SSS when samples were grouped by type [23], both considering WUF and its qualitative equivalent unweighted UniFrac distances (UWUF).

Stata 13.1 (StataCorp LP. 2013. College Station, TX) was used for generation of the Bland-Altman plots.

All relevant data were deposited at the Dryad Digital Repository (www.datadryad.org) and are referenced in the text using the following doi: http://dx.doi.org/10.5061/dryad.5gc82.

Results

We obtained a total of 36 high-quality pairs of sputum from 30 different COPD patients. Eleven patients were women; two thirds of patients were aged 55-64 years at inclusion. Patient characteristics are summarized in Table 1.

After processing of the raw data, 1004 different OTUs were identified with 2.5 of 4 million sequences belonging to samples delivered at exacerbations (25 of the 36 sputum pairs).

Table 1 Patient characteristics

	n (%)
Sex	
Women	11 (37%)
Men	19 (63%)
Age	
40-54 years	4 (13%)
55 - 64 years	18 (60%)
65 - 75 years	8 (27%)
Body composition	
Normal	21 (70%)
Obese	4 (13%)
Cachectic	5 (17%)
Smoking	
Ex	18 (60%)
Current	12 (40%)
GOLD stage	
II (FEV_1 50-80%)	14 (47%)
III (FEV_1 30-50%)	12 (40%)
IV (FEV_1 < 30%)	4 (13%)
Frequent exacerbator[a]	
No	20 (67%)
Yes	10 (33%)
Using inhaled steroids	
No	6 (20%)
Yes	24 (80%)
Using antibiotics[b]	
No	30 (100%)
Yes	0 (0%)

[a]>1 exacerbation last 12 months prior to inclusion
[b]At time of sampling

Taxonomy

The 1004 OTUs identified by QIIME were sorted into 106 different taxa by QIIME's taxonomic summary command. First, we calculated the Yue-Clayton measure of dissimilarity between the mean abundances of the most dominating OTUs (each containing ≥1% of all sequences) assigned to 11 different taxa in all ISS versus all SSS. This represents a group comparison and not a pair by pair comparison. The samples were then sorted with regards to disease state at time of sampling. The dissimilarity (1-θ_{YC}) measure was 0.04 when disease state was not considered. For exacerbation samples the dissimilarity (1-θ_{YC}) measure was also 0.04, and for stable state samples 0.03. Performing the same analyses including also low-abundance OTUs gave a dissimilarity

($1\text{-}\theta_{YC}$) measure of 0.04 when all samples were included, and $1\text{-}\theta_{YC}$ of 0.04, and 0.05 for exacerbation and stable samples respectively.

Taxonomic compositional differences within sample pairs were visualized as bar graphs for the same 11 dominating taxa (Fig. 1). As shown, there were obvious visual differences within some pairs. $1\text{-}\theta_{YC}$ was calculated both for dominating OTUs exclusively, and for all OTUs. When evaluating dissimilarities pair-wise for dominating OTUs and their associated taxa, $1\text{-}\theta_{YC}$ ranged from <0.01 – 0.92 (Fig. 1). $1\text{-}\theta_{YC}$ ranged from <0.01 – 0.58 when also including sparse OTUs and corresponding taxa (data not shown). With 0.2 as limit for acceptable within-pair $1\text{-}\theta_{YC}$, seven pairs were found dissimilar regardless of OTU-abundance (pairs 8, 11, 19, 20, 27, 34 and 36), while four pairs were found dissimilar only if filtering out low-abundance OTUs (pairs 6, 14, 24 and 28) or keeping low-abundance OTUs respectively (pairs 3, 13, 22 and 26).

To further assess differences in taxonomy between sample pairs, one Bland-Altman plot of the relative abundances of our 106 taxa was generated for each pair. From the upper and lower 95% LOA, the range is calculated (upper-lower/100) corresponding to a number between 0 and 1, where 0 indicates perfect agreement. Using this approach, we found ranges in LOA between

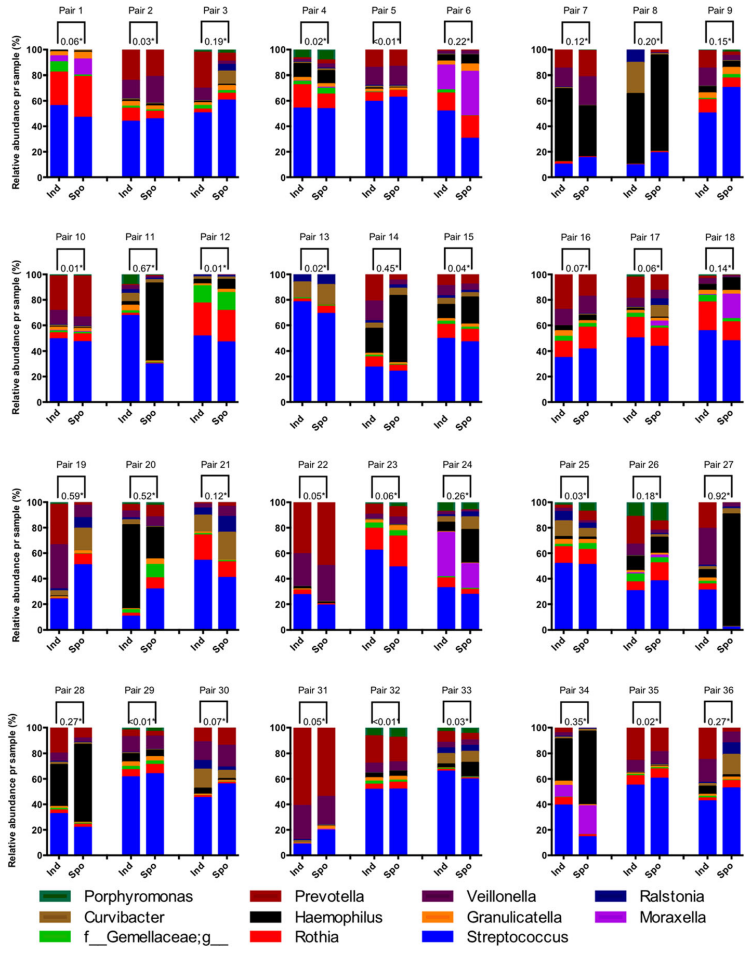

Fig. 1 Compositional taxonomic differences for each sample pair** on genus level for the 11 most dominating taxa***. * Yue-Clayton dissimilarity = Range 0-1. **Pair 1-11: Stable state Pair 12-36: Exacerbation. ***OTUs containing <1% of sequences were omitted from the data before performing taxonomic summaries (GreenGenes database level 6 = genus)

0.02-0.66 (Fig. 2). Setting an acceptable limit for LOA at 0.1 allows the relative abundance in each taxa to vary from ISS to SSS by 10%. With this limit 13 pairs could not be accepted as equal, including the seven pairs found too different by $1\text{-}\theta_{YC}$ regardless of OTU-abundance (Fig. 2).

There were significant differences between the mean abundances in induced versus spontaneous samples for 15 taxa in either the exacerbated or the stable state (Table 2). For instance for the well known pathogenic *Moraxella*, there were almost twice as many sequences in all spontaneous samples compared with all the induced samples both during exacerbations and in the stable state ($p < 0.001$, Table 2). Also *Haemophilus* was consistently more abundant in spontaneous than in induced samples.

Diversity

No statistically significant differences were found in alpha-diversity (Table 3).

However, we found statistically significant differences ($p < 0.01$, Bonferroni corrected due to multiple comparisons) in the pair-wise quantitative, phylogenetic beta-diversity as evaluated by weighted UniFrac for 9 pairs (Pair 3, 14, 17, 19, 26, 30, 32, 33 and 36).

The principal coordinates analysis (PCoA) plots are presented in Fig. 3.

Each dot represents the weighted UniFrac distance diversity measure for each sample, and lines illustrating the distance between paired sputum are shown (Blue line attaches to ISS, red to SSS). The greater the distance, the greater is the difference. Although this is a two-dimensional visualization of a three-dimensional calculation, Fig. 3 clearly shows that the distances between paired samples varied. A Monte Carlo simulation with 1000 permutations was applied giving $M^2 = 0.5$, confirming the visual interpretation (Identical plots: $M^2 = 0$, if completely dissimilar $M^2 = 1$).

Using analyses of similarities (ANOSIM), we found no significant differences in means of beta-diversity (UWUF and WUF) between ISS and SSS when samples were grouped by type. This was true both in stable state and at exacerbations ($p > 0.05$).

Discussion

This study on sputum samples collected sequentially using two different methodologies from COPD patients and treated equally by the same protocol shows that in approximately 1/3 of sputum pairs either taxonomical and/or diversity analyses differ significantly. Discordance between induced and spontaneous samples were seen both at exacerbations and during stable state.

The strength of the current study is the unique data material; including induced and spontaneous sputum samples collected simultaneously, treated by the same protocol [7, 8], both at the stable state and during exacerbations. However, there are some methodological issues to discuss. First, after either induction or through sampling of spontaneous sputum, sputum was kept in a clean collection dish, and material selected by trained technicians for further analyses. This is the standard approach [24], but entails a natural variation of sample selection. However, there is no reason to believe the judgment of the technician should differ between sample types, and all other processing was the same for both types of sputum.

Second, errors may occur during DNA extraction, PCR or sequencing steps. All pairs were run simultaneously for all steps in the laboratory protocol, including

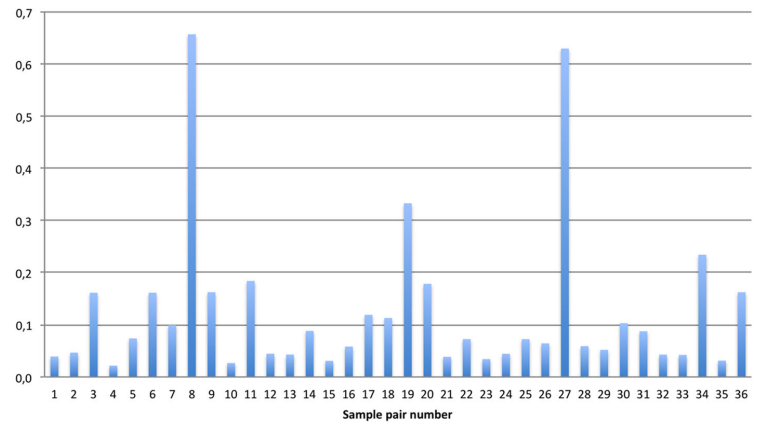

Fig. 2 Expected differences between relative abundances of taxa per sample pair with 95% Bland-Altman limits of agreement. *Range = 0-1. **Pair 1-11: Stable state Pair 12-36: Exacerbation

Table 2 Mean number of sequences per sample constituting the 15 taxa[a] found in significantly different amounts in induced and spontaneous sputum from COPD patients with and without respect to disease state

	All samples			Exacerbation			Stable state		
Taxonomy[b]	Induced	Spontaneous	p[c]	Induced	Spontaneous	p[c]	Induced	Spontaneous	p[c]
f_Prevotellaceae;g_Prevotella	2499.2	1760.3	<0.001	2633.1	1817.1	<0.001	2349.5	1763.0	<0.001
f_Pasteurellaceae;g_Haemophilus	1471.9	2440.3	<0.001	1308.2	2356.6	<0.001	1957.7	2859.8	<0.001
f_Moraxellaceae;g_Moraxella	259.2	542.3	<0.001	234.4	456.5	<0.001	353.9	798.0	0.001
f_Veillonellaceae;g_Veillonella	1488.3	1170.8	<0.001	1618.8	1173.1	<0.001			
f_Veillonellaceae;g_Megasphaera	179.9	100.0	<0.001	201.9	111.5	<0.001	134.8	79.4	0.01
f_Corynebacteriaceae;g_ Corynebacterium	44.9	12.0	<0.001	60.3	14.9	<0.001			
f_Oxalobacteraceae;g_Ralstonia				423.3	604.7	<0.001	331.8	213.3	<0.001
f_Comamonadaceae;g_Curvibacter				518.7	713.0	<0.001	430.7	269.5	<0.001
f_Leptotrichiaceae;g_Leptotrichia				171.8	95.2	<0.001			
f_Neisseriaceae;g_Neisseria							344.5	149.8	<0.001
f_Gemellaceae;g_Gemella							84.9	16.5	<0.001
f_Gemellaceae;g_							452.7	273.6	<0.001
f_Neisseriaceae;g_							79.5	17.5	<0.001
f_Leptotrichiaceae;g_Leptotrichia							123.6	217.6	<0.001
f_Actinomycetaceae;g_Actinomyces							350.1	256.0	0.001

[a]Rarefied OTU-tables: Sequences/Sample = 18,250 for All samples and Exacerbations, for Stable state: 19,743
[b]GreenGenes Level 6: f_ = Name of family level g_ = Name of genus level. One hundred six different taxa in total
[c]log-likelihood ratio test, Bonferroni corrected due to multiple comparisons

on the same flowcell in the Illumina MiSeq. However, random errors could be a factor, and based on the plots of the dominant taxa in Fig. 1, we chose the seven most visually dissimilar pairs (pairs 11, 14, 19, 20, 26, 27 and 36) and three visually similar pairs (pairs 2, 5 and 29) and redid the laboratory analyses. For only one of the 20 samples (pair 26, ISS) were the results convincingly different visually from the first to the second run. Since this was not a random selection, the likely error is much lower than 5%, and we do not believe our results are due to random laboratory error. For data analyses we chose to keep the sequences from run two for the ten re-run pairs.

Third, low biomass samples are prone to contamination from multiple sources during laboratory handling [25]. Approaches to handle the potential contamination include sequencing of known ("mock") communities, negative control samples, and manual curation of the sequencing output. A potential contaminator in our study is the saline used for induction. Unfortunately it was not stored at the time the procedures were performed, and so an important limitation to the current study is that we were not been able to examine the influence of negative saline controls on our samples. All samples were treated exactly similar at all steps of analyses, thus minimizing confounding from potential contamination. However, as the biomass and dilution of each sample in a pair may differ, we cannot exclude that samples could be differentially affected by contamination from saline. Finally, as in other studies comparing

Table 3 Mean within sample diversity (alpha diversity) in induced versus spontaneous sputum in COPD by different alpha diversity indices

	All samples			Exacerbation			Stable state		
Diversity Indices	Induced	Spontaneous	p[a]	Induced	Spontaneous	p[a]	Induced	Spontaneous	p[a]
Faith's Phylogenetic Diversity									
mean (std)	56.9 (9.3)	56.2 (8.6)	0.7	57.3 (9.5)	56.2 (8.8)	0.7	57.1 (8.5)	56.6 (8.4)	0.9
Chao1									
mean (std)	646.5 (116)	638.3 (107.5)	0.7	655.0 (118.5)	642.1 (105.1)	0.7	640.8 (110.4)	643.6 (105.3)	0.9
Observed OTUs									
mean (std)	543.2 (104.9)	528.8 (106.1)	0.6	552.9 (105)	534.3 (107.1)	0.6	531.6 (102.4)	527.7 (103.2)	0.9

[a]Non-parametric two-sample t-test using Monte Carlo permutations

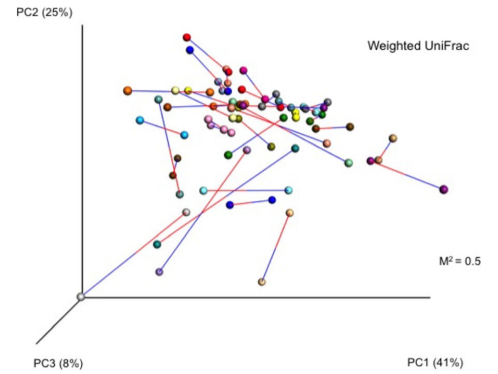

Fig. 3 PCoA plots modeling multi-dimentional distribution of OTUs derived from induced and spontaneous sputum samples*. *Sample pairs connected by bars - blue bars attach to induced sputum samples, red bars attach to spontaneous sputum samples

sampling methods' impact on microbiota [5, 26, 27], the number of samples is limited, and the statistical power therefore reduced.

One of the challenges in microbiome research is that the technological advancements develop faster than the establishment of statistical tools to assess results. What signifies a true compositional difference between two supposedly similar samples where each contains a large number of relative abundances of sequences is still an unsettled question. The cut-off for the two indices used for assessment of taxonomic differences, <0.2 for the Yue-Clayton dissimilarity index ($1\text{-}\theta_{YC}$) and <0.1 for LOA from the Bland Altman plots, are arbitrary, and no established consensus regarding these values exist. Similar for the Procrustes M^2 value there are no defined limit [28].

Finally, what constitutes a true clinically important difference is also an unsettled question. It could be that the entire ecological content of a sample is more relevant for disease, or it could be the presence of a few, perhaps even only one, low-abundant pathogen. If the latter is true, a cut off <0.2 for $1\text{-}\theta_{YC}$ and <0.1 for LOA will be too crude. With a sample size of 36 sputum pairs, this study did not have the power to evaluate whether ISS or SSS better correlated with clinical data. Future studies with larger sample sizes are needed to elucidate this question.

This study brings forward new information on the much used sputum samples in studies on COPD patients. Pair-wise comparisons of taxonomic composition on genus-level between ISS and SSS from lung patients have not previously been done to our knowledge. Neither have comparisons of alpha- or beta-diversity between ISS and SSS earlier been reported. Induced sputum sampling is an established protocol for studying COPD patients at stable state [29]. Common for both ISS and SSS is that they sample both lungs in contrast to bronchoalveolar lavage and biopsies, and are more easily accessible material. Spontaneous sputum is easier to collect during AECOPD when sputum production increase, and may be preferred by some for fear that induction may worsen airway obstruction. However, we have previously shown that induction can safely be performed during COPD exacerbations, at least with up to 3% hypertonic saline [7].

There are potential reasons why spontaneous and induced sputum samples would differ in their microbial content. Different airways regions have been shown to harbor different communities [30, 31], possibly partly due to different ventilation-circulation ratios in the lower and upper parts of the lungs, and possibly due to differences between proximal and distal airways. It has been shown before that sputum sampled early during induction has a different composition of cells than sputum sampled late during induction [32, 33]. Spontaneous samples may resemble proximal airways more than the distal sampled by induction, and possible differ in ability to sample upper and lower airways.

Abundant OTUs and correspondingly dominating taxa in different environments have been shown to be particularly important in their habitats [34]. However, sparse members of the microbiota have also been found to contribute in pathogenic processes in the lungs [35, 36]. With this in mind we chose to examine the identified taxa emphasizing both dominating and sparse OTUs. The group comparison of mean abundances of taxa by Yue-Clayton dissimilarity showed that pooling of observations can hide differences seen between individual sample pairs.

The strength of the $1\text{-}\theta_{YC}$ index is that it measures structural dissimilarity by calculating the proportions of both shared and unshared components in a community [17]. The number of pairs where ISS and SSS were considered too dissimilar to be accepted as good substitutes for each other ($1\text{-}\theta_{YC} > 0.2$) was the same regardless of focusing on taxonomic assignment of only abundant OTUs or accepting all OTUs. In both cases 1 of 3 pairs would render different results depending on which sample type was picked to represent the patient.

The Bland-Altman's LOA analyses confirmed the findings using Yue-Clayton's dissimilarity, in that ISS and SSS did not provide the same results in a significant fraction (13 of 36) of pairs when evaluating taxonomical composition in sputum from COPD patients.

Summarizing our findings on GreenGenes genus-level left 106 unique taxa. When comparing the mean abundance of sequences in each taxa between sample types, 8.5% of taxa were found in statistically significant different levels between sputum types during

exacerbations, and 11.3% in the stable state. Both of the known potential pathogens *Haemophilus* and *Moraxella* were significantly more abundant in spontaneous samples compared with induced samples, both in the stable state and during exacerbations. In this new era, where the whole composition of a microbiome may be relevant for disease, it may be that induced sputum samples better reflect presence of low-abundant species in the distal airways, which are masked by frequent colonization of genera like *Haemophilus* and *Moraxella* in spontaneous samples. However, presence of both *Haemophilus* [37, 38] and *Moraxella* [39] in stable state sputum samples have shown similar higher levels of inflammatory markers in the sputum samples indicating stimulation of the immune system. Thus either sampling method may have important value in research, but important differences in interpretation of the microbiota could result from using the sputum types interchangeably.

We could not find differences in alpha-diversity between sample types. This should perhaps not be surprising considering the shared route of delivery through the oral cavity and the samples not discriminating between right and left airways. It has been shown that diversity in sputum is higher than in explant lung samples, likely due to oral contamination [30].

There were no significant differences in mean phylogenetic beta-diversity between ISS and SSS, neither when considering absence/presence data, nor when emphasizing abundances (UWUF/WUF). However, when considering samples pair-wise we found differences in WUF in 1 of 4 pairs and for UWUF differences were found in 50% of the sputum pairs. With focus on quantitative data Procrustes transformation of PCoA-plots of WUF distances this pair-wise difference was confirmed, as the distances in multidimensional space were too large to ignore for several pairs. A defined limit for Procrustes M^2 to be considered too high to claim similarity does not exist, but levels >0.3 is indicative of influential differences.

Conclusions

In this study we found clear discrepancies in both taxonomic composition and beta-diversity between ISS and SSS collected concurrently from COPD patients in the stable state and during exacerbations when comparing samples pair-wise. For grouped analyses the differences were subtler, potentially masking important differences. The most prudent approach in studies using sputum for microbiota analyses is to only rely on either induced or spontaneous sputum. We advise that sampling method is always reported, and that comparisons are made and presented, if both sample types are used.

Abbreviations

AECOPD: Acute exacerbation of COPD; BCCS: Bergen COPD cohort study; BCES: Bergen COPD exacerbation study; CF: Cystic fibrosis; COPD: Chronic obstructive pulmonary disease; ISS: Induced sputum samples; LOA: Limits of agreement; OTU: Operational taxonomic unit; PCoA: Principal coordinates analysis; QIIME: Quantitative insights into microbial ecology; SSS: Spontaneous sputum samples; UFUW: Unweighted unifrac; WUF: Weighted unifrac

Acknowledgements

The authors wish to thank Lene Svendsen[1], Tina Endresen-Vinsjevik[1], Margrete Klemmetsby[1], Marie Wåtevik[1] and Eli Nordeide[1] for help in data collection, Inge Zwetzig[1] and Hilda Andersen[2] for performing the sputum processing, and Heidi Ellis Haraldsen[2] and Tuyen Hoang[1] for help with the DNA extraction, PCR and Illumina sequencing.

[1] Department of Thoracic Medicine, Haukeland University Hospital, Bergen, Norway.

[2] Department of Clinical Science, University of Bergen, Bergen, Norway. (Affiliations relate to positions at the time of the work).

Funding

The study was funded partially by the Department of Thoracic Medicine and the University of Bergen in terms of salaries. Analyses were funded by a grant from the Bergen Medical Research Foundation.

Availability of data and materials

The fastq files, metadata and command lines necessary for running the analyses presented in this article will be publicly available at DRYAD when the article is accepted for publishing. DOI: http://dx.doi.org/10.5061/dryad.5gc82.

Authors' contributions

ST performed DNA extraction, all bioinformatics analyses, and statistical analyses and drafted the manuscript. MA participated on data collection, statistical analyses and revision of the manuscript. RG participated in data collection, bioinformatics analyses, statistical analyses and revision of the manuscript. CD performed the PCR and Illumina sequencing analyses, and participated in revision of the manuscript. HGW planed all aspects of the sequencing analyses, and participated in the bioinformatics analyses and revision of the manuscript. PSB planed and participated in data collection and participated in the revision of the manuscript. TME planed and participated in data collection, designed the study, and participated in the bioinformatics, statistical analyses and revision of the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

The study was approved by the regional ethical board (REK-Vest Norway), case number 165.08. All participants signed a consent form upon inclusion.

Consent for publication

Not applicable.

Competing interests

ST: Nothing to declare.

MA: Within the last 5 years MA has received a lecture fee from Novartis.

RG: Reports grants from the Norwegian Association of Heart and Lung Patients and EXTRA funds from the Norwegian Foundation for Health and Rehabilitation as well as YaraPraxair during the conduct of the study, grants and personal fees from Boehringer Ingelheim, personal fees from AstraZeneca, and personal fees from GlaxoSmithKline outside the submitted work.

CD: Nothing to declare.

HGW: Nothing to declare.

PSB: Within the last 5 years PSB has received lecture fees or advisory board fees from AstraZeneca, GlaxoSmithKline, Boehringer-Ingelheim, Mundipharma, Chiesi, MSD. PSB is a member of the steering committee and scientific committee of the ECLIPSE study and is PI of the ECLIPSE extension study, both studies sponsored by GlaxoSmithKline.

TME: In the last 5 years TME has received unrestricted grants from Boehringer Ingelheim outside this work, speaker fees from AstraZeneca, Boehringer Ingelheim and GlaxoSmithKline, and support for travel to the AIR conference from InterMune.

Publisher's Note

Author details
[1]Department of Thoracic Medicine, Haukeland University Hospital, Bergen, Norway. [2]Department of Clinical Science, Faculty of Medicine and Dentistry, University of Bergen, Bergen, Norway. [3]Department of Microbiology, Haukeland University Hospital, Bergen, Norway.

Received: 16 January 2017 Accepted: 20 August 2017
Published online: 29 August 2017

References

1. Dickson RP, Erb-Downward JR, Huffnagle GB. The role of the bacterial microbiome in lung disease. Expert Rev Respir Med. 2013;7:245–57.
2. Paggiaro PL, Chanez P, Holz O, Ind PW, Djukanovic R, Maestrelli P, Sterk PJ. Sputum induction. Eur Respir J Suppl. 2002;37:3s–8s.
3. Sethi S, Anzueto A, Miravitlles M, Arvis P, Alder J, Haverstock D, Trajanovic M, Wilson R. Determinants of bacteriological outcomes in exacerbations of chronic obstructive pulmonary disease. Infection. 2016;44(1):65–76.
4. Bafadhel M, Haldar K, Barker B, Patel H, Mistry V, Barer MR, Pavord ID, Brightling CE. Airway bacteria measured by quantitative polymerase chain reaction and culture in patients with stable COPD: relationship with neutrophilic airway inflammation, exacerbation frequency, and lung function. Int J Chron Obstruct Pulmon Dis. 2015;10:1075–83.
5. Zemanick ET, Wagner BD, Robertson CE, Stevens MJ, Szefler SJ, Accurso FJ, Sagel SD, Harris JK. Assessment of airway microbiota and inflammation in cystic fibrosis using multiple sampling methods. Ann Am Thorac Soc. 2015;12:221–9.
6. Rogers GB, Skelton S, Serisier DJ, van der Gast CJ, Bruce KD. Determining cystic fibrosis-affected lung microbiology: comparison of spontaneous and serially induced sputum samples by use of terminal restriction fragment length polymorphism profiling. J Clin Microbiol. 2010;48:78–86.
7. Tangedal S, Aanerud M, Persson LJ, Brokstad KA, Bakke PS, Eagan TM. Comparison of inflammatory markers in induced and spontaneous sputum in a cohort of COPD patients. Respir Res. 2014;15:138.
8. Eagan TM, Ueland T, Wagner PD, Hardie JA, Mollnes TE, Damas JK, Aukrust P, Bakke PS. Systemic inflammatory markers in COPD: results from the Bergen COPD cohort study. Eur Respir J. 2010;35:540–8.
9. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7:335–6.
10. Navas-Molina JA, Peralta-Sanchez JM, Gonzalez A, McMurdie PJ, Vazquez-Baeza Y, Xu Z, Ursell LK, Lauber C, Zhou H, Song SJ, et al. Advancing our understanding of the human microbiome using QIIME. Methods Enzymol. 2013;531:371–444.
11. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA, Caporaso JG. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat Methods. 2013;10:57–9.
12. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26:2460–1.
13. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol. 2006;72:5069–72.
14. Drancourt M, Bollet C, Carlioz A, Martelin R, Gayral JP, Raoult D. 16S ribosomal DNA sequence analysis of a large collection of environmental and clinical unidentifiable bacterial isolates. J Clin Microbiol. 2000;38:3623–30.
15. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics. 2010;26:266–7.
16. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol. 2009;26:1641–50.
17. Yue JC, Clayton MK. A similarity measure based on species proportions. Commun Stat Theory Methods. 2005;34:2123–31.
18. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999;8:135–60.
19. Giavarina D. Understanding bland Altman analysis. Biochem Med (Zagreb). 2015;25:141–51.
20. He Y, Caporaso JG, Jiang XT, Sheng HF, Huse SM, Rideout JR, Edgar RC, Kopylova E, Walters WA, Knight R, Zhou HW. Stability of operational taxonomic units: an important but neglected property for analyzing microbial diversity. Microbiome. 2015;3:20.
21. Lozupone CA, Knight R. The UniFrac significance test is sensitive to tree topology. BMC Bioinformatics. 2015;16:211.
22. Gower JC. Generalized procrustes analysis. Psychometrika. 1975;40:33–51.
23. Clarke KR. Non-parametric multivariate analysis of changes in community structure. Aust J Ecol. 1993;18:117–43.
24. Efthimiadis A, Spanevello A, Hamid Q, Kelly MM, Linden M, Louis R, Pizzichini MM, Pizzichini E, Ronchi C, Van Overvel F, Djukanovic R. Methods of sputum processing for cell counts, immunocytochemistry and in situ hybridisation. Eur Respir J Suppl. 2002;37:19s–23s.
25. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol. 2014;12:87.
26. Bassiouni A, Cleland EJ, Psaltis AJ, Vreugde S, Wormald PJ. Sinonasal microbiome sampling: a comparison of techniques. PLoS One. 2015;10: e0123216.
27. Charlson ES, Diamond JM, Bittinger K, Fitzgerald AS, Yadav A, Haas AR, Bushman FD, Collman RG. Lung-enriched organisms and aberrant bacterial and fungal respiratory microbiota after lung transplant. Am J Respir Crit Care Med. 2012;186:536–45.
28. Schneider JW, Borlund P. Matrix comparison, part 2: measuring the resemblance between proximity measures or ordinat. J Am Soc Inf Sci Technol. 2007;58:1596–609.
29. Lacy P, Lee JL, Vethanayagam D. Sputum analysis in diagnosis and management of obstructive airway diseases. Ther Clin Risk Manag. 2005;1: 169–79.
30. Goddard AF, Staudinger BJ, Dowd SE, Joshi-Datar A, Wolcott RD, Aitken ML, Fligner CL, Singh PK. Direct sampling of cystic fibrosis lungs indicates that DNA-based analyses of upper-airway specimens can misrepresent lung microbiota. Proc Natl Acad Sci U S A. 2012;109:13769–74.
31. Erb-Downward JR, Thompson DL, Han MK, Freeman CM, McCloskey L, Schmidt LA, Young VB, Toews GB, Curtis JL, Sundaram B, et al. Analysis of the lung microbiome in the "healthy" smoker and in COPD. PLoS One. 2011;6:e16384.
32. Gershman NH, Liu H, Wong HH, Liu JT, Fahy JV. Fractional analysis of sequential induced sputum samples during sputum induction: evidence that different lung compartments are sampled at different time points. J Allergy Clin Immunol. 1999;104:322–8.
33. Holz O, Jorres RA, Koschyk S, Speckin P, Welker L, Magnussen H. Changes in sputum composition during sputum induction in healthy and asthmatic subjects. Clin Exp Allergy. 1998;28:284–92.
34. Mariadassou M, Pichon S, Ebert D. Microbial ecosystems are dominated by specialist taxa. Ecol Lett. 2015;18:974–82.
35. de Steenhuijsen Piters WA, Huijskens EG, Wyllie AL, Biesbroek G, van den Bergh MR, Veenhoven RH, Wang X, Trzcinski K, Bonten MJ, Rossen JW, et al. Dysbiosis of upper respiratory tract microbiota in elderly pneumonia patients. ISME J. 2016;10:97–108.
36. Arrieta MC, Stiemsma LT, Dimitriu PA, Thorson L, Russell S, Yurist-Doutsch S, Kuzeljevic B, Gold MJ, Britton HM, Lefebvre DL, et al. Early infancy microbial and metabolic alterations affect risk of childhood asthma. Sci Transl Med. 2015;7:307ra152.
37. Persson LJ, Aanerud M, Hardie JA, Miodini Nilsen R, Bakke PS, Eagan TM, Hiemstra PS. Antimicrobial peptide levels are linked to airway inflammation, bacterial colonisation and exacerbations in chronic obstructive pulmonary disease. Eur Respir J. 2017;49:1601328. https://doi.org/10.1183/13993003.01328-2016.
38. Tufvesson E, Bjermer L, Ekberg M. Patients with chronic obstructive pulmonary disease and chronically colonized with Haemophilus influenzae during stable disease phase have increased airway inflammation. Int J Chron Obstruct Pulmon Dis. 2015;10:881–9.
39. Parameswaran GI, Wrona CT, Murphy TF, Sethi S. Moraxella catarrhalis acquisition, airway inflammation and protease-antiprotease balance in chronic obstructive pulmonary disease. BMC Infect Dis. 2009;9:178.

PLOS ONE

RESEARCH ARTICLE

Sputum microbiota and inflammation at stable state and during exacerbations in a cohort of chronic obstructive pulmonary disease (COPD) patients

Solveig Tangedal[1,2]*, Rune Nielsen[1,2], Marianne Aanerud[2], Louise J. Persson[2], Harald G. Wiker[1,4], Per S. Bakke[1], Pieter S. Hiemstra[3], Tomas M. Eagan[1,2]

1 Dept. of Clinical Science, Faculty of Medicine, University of Bergen, Bergen, Norway, **2** Dept. of Thoracic Medicine, Haukeland University Hospital, Bergen, Norway, **3** Dept of Pulmonology, Leiden University Medical Center, Leiden, The Netherlands, **4** Dept. of Microbiology, Haukeland University Hospital, Bergen, Norway

* stangedal@gmail.com

OPEN ACCESS

Citation: Tangedal S, Nielsen R, Aanerud M, Persson LJ, Wiker HG, Bakke PS, et al. (2019) Sputum microbiota and inflammation at stable state and during exacerbations in a cohort of chronic obstructive pulmonary disease (COPD) patients. PLoS ONE 14(9): e0222449. https://doi.org/10.1371/journal.pone.0222449

Editor: Aran Singanayagam, Imperial College London, UNITED KINGDOM

Received: March 17, 2019

Accepted: August 30, 2019

Published: September 17, 2019

Data Availability Statement: The fastq files, metadata and command lines necessary for obtaining the dataset used in this article are available at DRYAD. DOI wet-lab protocol: dx.doi.org/10.17504/protocols.io.2sygefw. DOI fastq: https://doi.org/10.5061/dryad.5gc82. DOI metadata and command lines: 10.5061/dryad.2f25cj0.

Funding: The Bergen COPD cohort study was supported by the Centre for Clinical Research, Haukeland University Hospital. Cost of analyses for

Abstract

Background

Exacerbations of chronic obstructive pulmonary disease (COPD) are debilitating events and spur disease progression. Infectious causes are frequent; however, it is unknown to what extent exacerbations are caused by larger shifts in the airways' microbiota. The aim of the current study was to analyse the changes in microbial composition between stable state and during exacerbations, and the corresponding immune response.

Methods

The study sample included 36 COPD patients examined at stable state and exacerbation from the Bergen COPD Cohort and Exacerbations studies, and one patient who delivered sputum on 13 different occasions during the three-year study period. A physician examined the patients at all time points, and sputum induction was performed by stringent protocol. Only induced sputum samples were used in the current study, not spontaneously expectorated sputum. Sputum inflammatory markers (IL-6, IL-8, IL-18, IP-10, MIG, TNF-α) and antimicrobial peptides (AMPs, i.e. LL-37/hCAP-18, SLPI) were measured in supernatants, whereas target gene sequencing (16S rRNA) was performed on corresponding cell pellets. The microbiome bioinformatics platform QIIME2™ and the statistics environment R were applied for bioinformatics analyses.

Results

Levels of IP-10, MIG, TNF-α and AMPs were significantly different between the two disease states. Of 36 sample pairs, 24 had significant differences in the 12 most abundant genera between disease states. The diversity was significantly different in several individuals, but not when data was analysed on a group level. The one patient case study showed longitudinal dynamics in microbiota unrelated to disease state.

the Luminex analyses was provided from a personal grant from the Bergen Medical Research Foundation to T. E in 2009. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: Dr. Tangedal, Dr. Nielsen, Dr. Aanerud, Dr. Persson and Dr. Wiker have nothing to disclose. Dr. Bakke reports personal fees from Novartis, personal fees from AstraZeneca, personal fees from Chiesi, personal fees from Boehringer-Ingelheim, outside the submitted work. Dr. Hiemstra reports grants from Boehringer Ingelheim, grants from Galapagos, outside the submitted work. Dr. Eagan reports grants from Bergen Medical Research Foundation, during the conduct of the study; personal fees from Boehringer Ingelheim, outside the submitted work. This does not alter the authors' adherence to PLOS ONE policies on sharing data and materials.

Conclusion

Changes in the sputum microbiota with changing COPD disease states are common, and are accompanied by changes in inflammatory markers. However, the changes are highly individual and heterogeneous events.

Introduction

A myriad of bacteria and other microorganisms, collectively called the human microbiota, inhabits the human body. With modern marker-gene DNA-sequencing technology more knowledge of how bacteria affect the human host is rapidly being acquired. It was long believed that the lower airways were sterile, but recent studies have shown a present microbiota also in healthy subjects [1–3].

Chronic obstructive pulmonary disease (COPD) is characterized by chronic inflammation in the airways [4], and an increase in systemic inflammation [5, 6]. The cause of the inflammation has been unknown, but toxic effects of inhaled tobacco or other substances [7] and autoimmunity has been suggested [8].

A dramatic manifestation of COPD, the acute exacerbations [9] with potentially life-threatening airways obstruction, is most often seen in combination with symptoms of infection. Indeed, bacteria and viruses are believed to trigger most exacerbations [10, 11]. Traditionally this has been seen as single-agent infections, and one debate has been whether any such agent was acquired by contagion or an upswing of pre-existing colonizing agents [12]. Although most exacerbations are likely due to infections, it is suspected that environmental factors like air-borne pollution and air-temperature can trigger these episodes [10]. Thus, single-agent infections are unlikely explanations for all or the entire COPD exacerbation event.

We suggest that the chronic inflammation of COPD reflects a chronically distorted microbiota. And, that the COPD exacerbations may reflect an acutely imbalanced respiratory ecosystem, with an accompanying inflammatory response to this imbalance.

However, little information exists to date on the dynamics of the airways microbiota in COPD patients shifting from a steady state to a COPD exacerbation [13]. In the current study we examined the microbiota in 36 COPD patients from whom we had induced sputum samples collected both during the stable state and during COPD exacerbations. And in one particular patient prone to experience frequent exacerbations we assessed the temporal changes of the sputum microbiota over 36 months in six samples from stable state visits, and seven collected during exacerbations.

Methods

Study population

The Bergen COPD Exacerbation Study (BCES) included all COPD patients from the Bergen COPD Cohort Study (BCCS) that belonged to the Haukeland University Hospital district for emergency care (356 out of 433 COPD patients in the BCCS). Detailed descriptions of study design and inclusion for the BCCS and the BCES has been published [6, 14]. Only induced sputum samples were used in the current study, not spontaneously expectorated sputum. A flowchart depicting the selection of the study sample is presented in Fig 1. Of the 356 included patients, 154 had one or more examined exacerbation events. Sputum induction was attempted unless the patient declined or in some instances when we did not have available technicians to process the sputum fresh after the induction. A total of 36 patients had induced

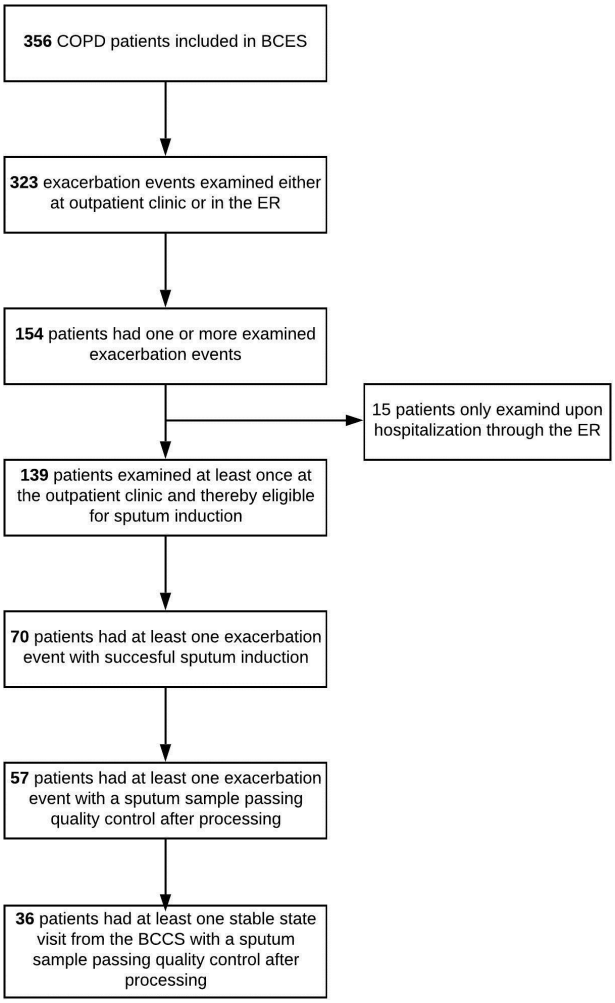

Fig 1. Flowchart depicting selection of the 36 patients in the current study, from the Bergen COPD Exacerbation Study (BCES) and the Bergen COPD Cohort Study (BCCS).

https://doi.org/10.1371/journal.pone.0222449.g001

sputum of acceptable quality available both from a stable state and an exacerbation visit, and these 36 sputum pairs define the current study population (Fig 1).

All patients provided written informed consent, and the Norwegian Regional Ethical Committee approved the study (REK-Vest, case number 165.08).

Data collection

A trained study physician examined all patients both at regular BCCS-visits and during BCES-exacerbation visits. Classification of airways obstruction was according to Global initiative for

chronic obstructive lung disease (GOLD) guidelines [15]. Body composition was determined with bioelectrical impedance measurements, and patients categorized as normal, obese or cachectic [16]. COPD exacerbation history was taken by the study physician at the baseline visit of the study, based on patient recall. An exacerbation was defined as a worsening of symptoms requiring treatment with either antibiotics or oral steroids. Induced sputum sampling was performed depending on patients' cooperation and availability of study technicians trained in sputum processing.

Laboratory analyses

Sputum samples had to fulfill quality measures ensuring lower airway sampling. The details of sputum induction and processing are previously published [14, 17].

Sputum processing was performed immediately after sampling. After the filtering step, samples were centrifuged at 4˚C for > 15 minutes at 450 g. The resulting supernatants and cell pellets were frozen separately at -80˚C. DNA was extracted from cell pellets using the FastPrep-24 Instrument and reagents from the FastDNA Spin Kit (MP Biomedicals, LLC, Solon, OH, USA). Amplicon PCR (45 cycles) and index PCR were run using primers from the Nextera XT Index Kit (Illumina Inc., San Diego, CA, USA). Paired-end sequencing (2 x 300 cycles) of the V3-V4 region of the 16S rRNA gene followed the protocol for Metagenomic Sequencing Library Preparation for the Illumina Miseq System (Part # 15044223 Rev. B, MiSeq Reagent Kit v3).

The inflammatory markers interleukin-6 (IL-6), interleukin-8 (IL-8), interleukin-18 (IL-18), interferon gamma-inducible protein-10 (IP-10), tumor necrosis factor-alpha (TNF-α) and monokine induced by gamma interferon (MIG) in sputum supernatants were processed using bead-based multiplex assays and the Luminex® xMAP® technology (Luminex Corporation, Austin, Texas). The data on sputum levels of LL-37 (a cathelicidin peptide derived from human hCAP-18) and secretory leucocyte protease inhibitor (SLPI) derived from previously unfrozen, aliquots of the same sputum supernatants by enzyme immunoassays, were derived from a previous analysis [18, 19].

Bioinformatics analyses

The amplicon sequences were quality and chimera filtered through the microbiota pipeline Quantitative Insights Into Microbial Ecology 2 (QIIME2) (v.2017.9 – v.2018.11) [20], using the Divisive Amplicon Denoising Algorithm 2 (DADA2) [21]. Laboratory-made sequences (chimeras) were removed first through DADA2 [21] and then VSEARCH [22]. Negative controls were unavailable, so to filter contaminants we used the total DNA-load measurements (Quant-iT™ PicoGreen™, ThermoFisher Scientific Inc) and the Decontam algorithm in R [23]. Amplicon sequence variants (ASVs) created by DADA2 were assigned taxonomy, using a self-trained Naïve Bayes classifier and the Silva database [24]. ASVs that could not be assigned taxonomy beyond kingdom level were omitted. After de-novo alignment, FastTree was used to build a phylogenetic tree for diversity analyses [25].

Statistical analyses

To compare inflammatory markers and antimicrobial peptides in sputum during the stable state and during exacerbations, Wilcoxon signed rank test was used to account for the paired design. To compare taxonomic composition between pairs of samples we calculated the Yue-Clayton measure of dissimilarity ($1-\theta_{YC}$) [26]. This was performed at the genus level, after omitting ASVs containing < 1% of the total amount of sequences. Differential abundances of taxa between disease states were analysed using an ANOVA-like differential expression procedure (Aldex2) in R [27]. Diversity analyses were performed after sub-setting all samples at the

number of sequences of the sparsest sample (rarefaction). Beta-diversity visualized as non-metric multidimensional scaling plots (NMDS), were analysed with permutation tests of multivariate homogeneity of variances, permuted analysis of variance (PERMANOVA) and Procrustes analyses in the *Vegan* package in R [28]. For analyses of clinical data relative to measurements from biological samples StataSE (StataCorp LP. Release 14. College Station, TX) was used. Further details on bioinformatics and statistical methods are available in the online supplement S1 Text. More on bioinformatics and statistical methods.

Results

Table 1 shows the patient characteristics for the 36 included COPD patients.

Eleven patients had experienced two or more exacerbations the last 12 months before inclusion. At inclusion 28 participants used inhaled corticosteroids. No patients used antibiotics or oral corticosteroids at stable state, whereas at exacerbation visits, one patient used antibiotics, one used oral corticosteroids, and one patient used both (Table 1). For 26 of the 36 sputum pairs, the stable sputum was collected prior to an exacerbation event, and vice versa for the other 10 pairs. The median number of days between the two collections were 257 days.

Inflammatory markers and antimicrobial peptides

Levels of the two AMPs and three of the measured inflammatory markers (IP-10, MIG, TNF-α) differed significantly in sputum sampled between disease states (Fig 2), with levels of all

Table 1. Patient characteristics at inclusion in the Bergen COPD Cohort Study.

	n (%)
Sex	
Women	15 (42%)
Men	21 (58%)
Age	
40–54 years	4 (11%)
55–64 years	21 (58%)
65–75 years	11 (31%)
Body composition	
Normal	27 (75%)
Obese	6 (17%)
Cachectic	3 (8%)
Smoking	
Ex	21 (58%)
Current	15 (42%)
GOLD COPD stage	
II (FEV1 50–80%)	18 (50%)
III (FEV1 30–50%)	14 (39%)
IV (FEV1 <30%)	4 (11%)
Frequent exacerbator*	
No	24 (67%)
Yes	11 (31%)

* >1 exacerbation last 12 months prior to inclusion. One patient missing information. GOLD: Global Initiative for Chronic Obstructive Lunge Disease. COPD: Chronic obstructive lunge disease. FEV1: Forced expiratory volume 1st second

https://doi.org/10.1371/journal.pone.0222449.t001

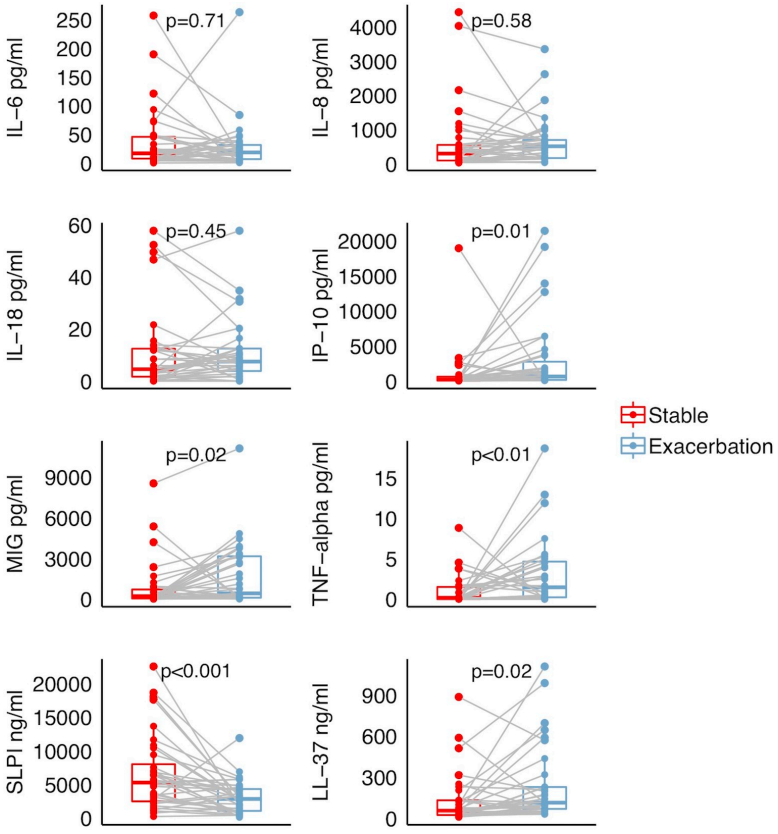

Fig 2. Inflammatory markers and antimicrobial peptides in induced sputum collected from a COPD cohort at stable state and during exacerbation. Interleukin-6 (IL-6), Interleukin-8 (IL-8), Interleukin-18 (IL-18), Interferon Gamma-Induced Protein 10 (IP-10), Monokine induced by gamma interferon (MIG): n = 35. Secretory Leukocyte Protease Inhibitor (SLPI): n = 36. LL-37/hCAP-18: n = 34. Boxes show the interquartile range (IQR = 75^{th} percentile-25^{th} percentile), with medians marked by the horizontal line within each box. Samples collected from the same patient at different disease states are connected by lines. Wilcoxon signed rank test was applied based on the paired, non-parametric data.

https://doi.org/10.1371/journal.pone.0222449.g002

mediators being higher during exacerbation except for SLPI. One patient had no measurement of inflammatory markers and two patients had no measurement of LL-37/hCAP.

Taxonomy

Of 15 phyla identified, *Firmicutes*, *Proteobacteria*, *Bacteroidetes* and *Actinobacteria* were the most abundant, containing 97% of all sequences at both disease states. *Proteobacteria* was relatively more dominating in samples collected during exacerbations compared to stable state. *Streptococcus*, *Rothia*, *Prevotella 7*, *Veillonella*, and *Haemophilus*; which altogether contained 68% of all sequences at both disease states were the most abundant genera (Fig 3).

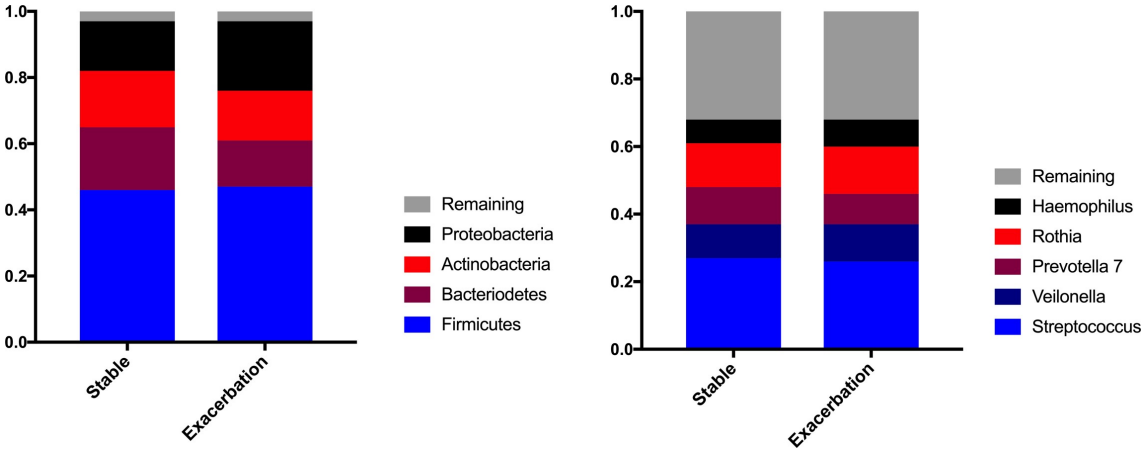

Fig 3. The four most abundant phylae and the five most abundant genera found in induced sputum samples from COPD patients during the stable state, and during exacerbations.

https://doi.org/10.1371/journal.pone.0222449.g003

The differential abundances of different taxa (often designated "features" in bioinformatics analyses) between disease states were tested at Silva's phyla and genus level, and for each ASV. Differential abundances in features between the stable state-sample group and the exacerbation-sample group were not found (FDR-corrected, effect size cut off 0.5. Wilcoxon $p>0.05$ for all taxa at all three levels, all data available in S1 Table.).

The taxonomic composition and $1\text{-}\theta_{YC}$ of the 36 sputum pairs are shown in Fig 4. The Yue-Clayton measure is 0 with perfect similarity and 1 with perfect dissimilarity. To evaluate the similarity within each sputum pair, 0.2 was set as the Yue-Clayton limit for acceptable within-pair similarity. With this cut-off, 26 patients had sputum pairs considered dissimilar.

The ten patients with low 1-θYC, and thus similar taxonomic composition across disease states, did not differ significantly from the other participants with regards to sex, age, body composition, smoking status, COPD stage, exacerbation frequency or use of inhaled corticosteroids ($p>0.05$, results not shown). Considering levels of inflammatory markers and AMPs at both disease states, only levels of SLPI during exacerbations were significantly lower in patients with dissimilar sputum pairs (Fig 5), whereas IL-8 trended towards higher levels in patients with dissimilar sputum during exacerbations ($1\text{-}\theta_{YC}<0.2$: Median IL-8 200.5 pg/ml, IQR (59.4–659.1) $1\text{-}\theta_{YC}\geq0.2$: Median IL-8 614.0 pg/ml, IQR (199.9–812.0), Kruskal Wallis $p = 0.053$).

Diversity

Rarefaction curves of alpha-diversity (within-sample diversity) showed asymptote at 1000 sequences/sample (Fig 1 in S1 Text. More on bioinformatics and statistical methods). Faith's phylogenetic diversity (PD) and Shannon's non-phylogenetic diversity (non-PD) indices showed no significant differences in alpha-diversity between the two disease states (Table 1 in S1 Text. More on bioinformatics and statistical methods).

Changes in individual alpha-diversity between disease states are visualized in Fig 6. There were inconsistent directionality and magnitude of alpha-diversity change between patients. Faith's PD was higher at stable state in 15 patients and for Shannon's non-PD this was the case in 17 patients (Fig 6A).

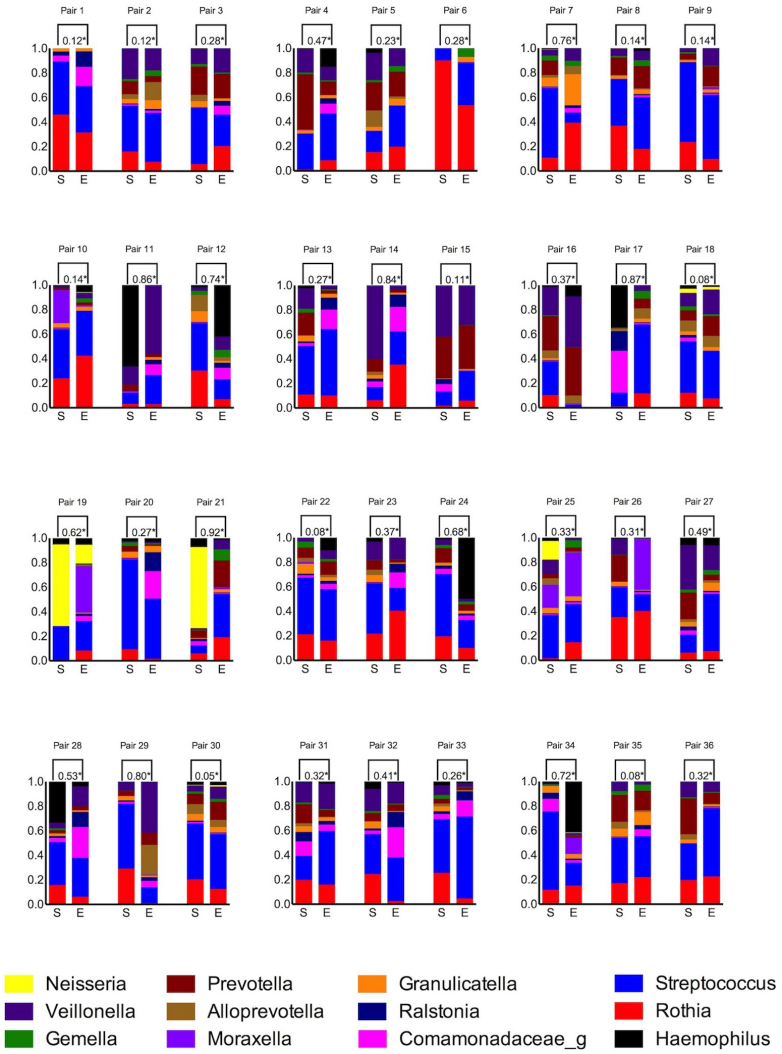

Fig 4. Comparison of bacterial composition in pairs of induced sputum samples (stable state and exacerbation) from 36 patients suffering from chronic obstructive lung disease. Presenting level 6 taxonomy (genus) provided by Silva database for amplicon sequence variants containing at least 1% of all sequences. *Yue-Clayton dissimilarity (1-θ_{YC}) Range 0 to 1; 0 = perfect similarity, 1 = perfect dissimilarity. S: Stable state E: Exacerbation.

https://doi.org/10.1371/journal.pone.0222449.g004

Changes in Faith's PD by disease state were not related to levels of white blood cell counts (WBC) or absolute neutrophil counts (ANC), while Shannon's non-PD was lower at stable state among patients whose ANC did not become elevated during exacerbations (Kruskal Wallis, p = 0.04) (Fig 6B).

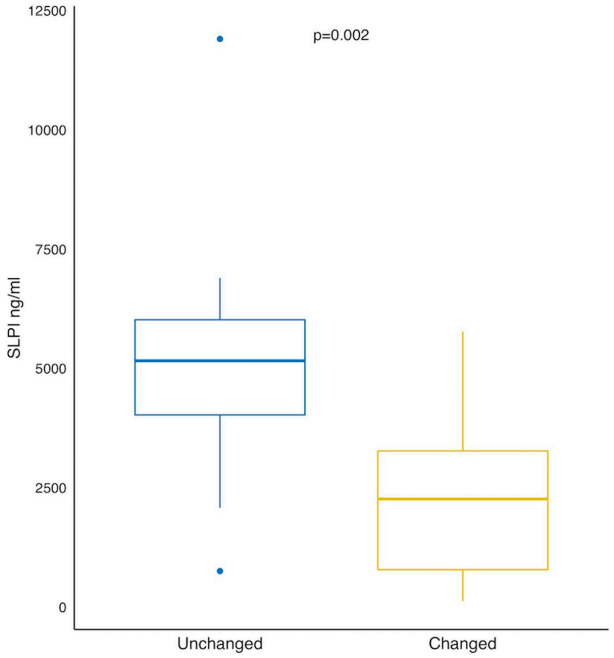

Fig 5. Comparing Secretory Leukocyte Protease Inhibitor (SLPI) measured in induced sputum in exacerbated COPD patients with regards to microbial composition alterations between disease states. Unaltered = Yue-Clayton dissimilarity index <0.2 (n = 10), Altered = Yue-Clayton dissimilarity index >0.2 (n = 26). Boxes show the interquartile range (IQR = 75th percentile-25th percentile), with medians marked by the horizontal line within each box. Kruskal Wallis test used due to non-parametric data.

https://doi.org/10.1371/journal.pone.0222449.g005

We did not find clustering by disease state when we examined different ordinance plots of beta-diversity (between-sample diversity) (Fig 7). With the PERMANOVA test to compare the average community value (centroid) between disease-states, significant differences were found only for non-phylogenetic matrices (Bray-Curtis p = 0.017, and Sørensen p = 0.004), however the corresponding R^2 values were only 0.02 for both.

To investigate beta-diversity within individuals, one distance matrix was created for each disease state. Overlaying stable state and exacerbation ordinance plots after Procrustes transformation showed ample distance within several sample pairs (Fig 8). M^2 values > 0.3 indicate that the samples delivered at the different disease states have poor resemblance. Information on which pairs have the least similar samples is given in Fig 2 in S1 Text. More on bioinformatics and statistical methods.

Longitudinal case study

One patient (NN) delivered induced sputum samples from six stable state visits and seven exacerbations. NN was a 66-year old ex-smoker, diagnosed with COPD stage IV at inclusion. NN continued being a frequent exacerbator the three years the study lasted.

The taxonomic composition (including ASVs consisting of $\geq$ 1% of all sequences) for each of the 13 samples are shown in Fig 9A. Of the six dominating genera, *Streptococcus*,

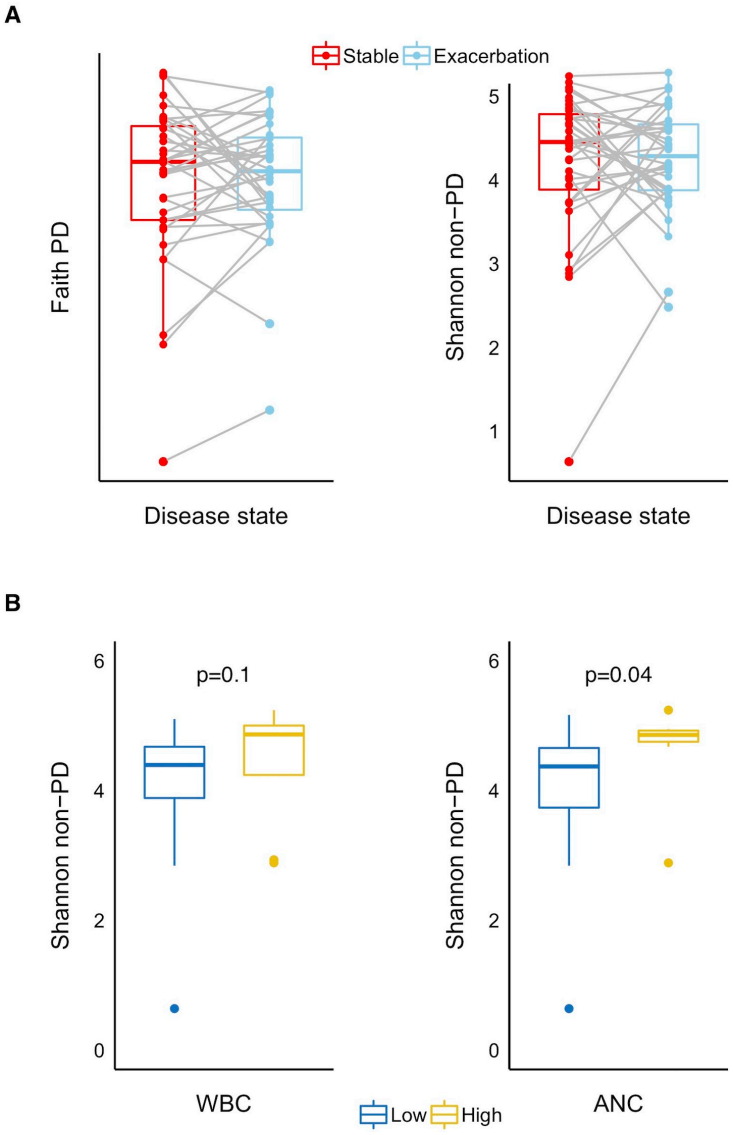

Fig 6. Alpha-diversity in induced sputum collected from 36 COPD patients. A: Comparison of phylogenetic (Faith PD) and non-phylogenetic (Shannon non-PD) alpha-diversity in sputum collected at stable state and during exacerbation. Lines connect samples from the same individual. Wilcoxon signed rank test was applied based on the paired, non-parametric data. B: Relations between Shannon's alpha-diversity at stable state and serum inflammatory markers during exacerbations. WBC: White Blood Cell counts high >11.3 10^9/L (n = 8) ANC: Absolute Neutrophil Count high >8.4 10^9/L (n = 7). Kruskal Wallis test used due to non-parametric data. Boxes show the interquartile range (IQR = 75^{th} percentile-25^{th} percentile), with medians marked by the horizontal line within each box.

https://doi.org/10.1371/journal.pone.0222449.g006

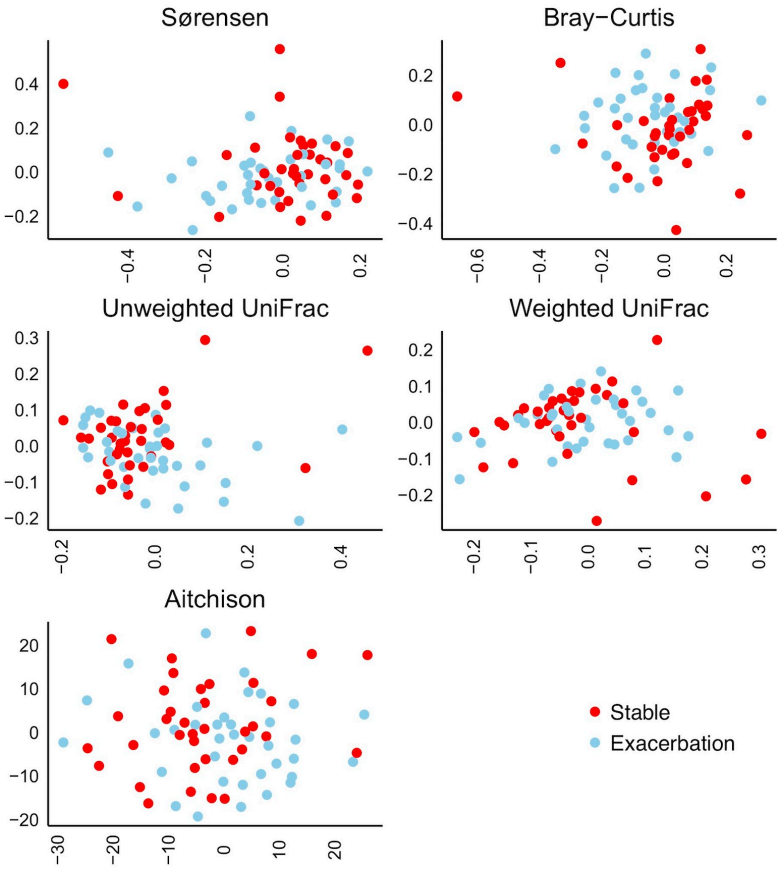

Fig 7. Beta-diversity in induced sputum collected from 36 chronic obstructive pulmonary disease sufferers both at stable state and during exacerbations, presented with non-metric multidimensional scaling (NMDS) ordinations. The X- and Y-axes display the first and second NMDS dimension respectively. Distance matrices: Sørensen and Bray Curtis: Both non-phylogenetic; qualitative and quantitative information respectively. Unweighted and weighted UniFrac: Both phylogenetic; qualitative and quantitative information respectively. Aitchison: Compositional interpretation of sequence counts.

https://doi.org/10.1371/journal.pone.0222449.g007

Ralstonia and *Comamonadaceae* were seen in all samples. *Rothia*, *Moraxella* and *Gemella* were the other genera found to dominate, though not consistently seen at each sampling occasion.

Variability in phylogenetic diversity measures are displayed in Fig 9B. Alpha-diversity changed between sampling time points, but there was no consistent pattern in directionality between the stable state samples and the samples collected during exacerbations. When comparing beta-diversity, there was a trending increase in distances over time with unweighted UniFrac. However, this could not be seen for weighted UniFrac distances, which also varied over time unrelated to disease state.

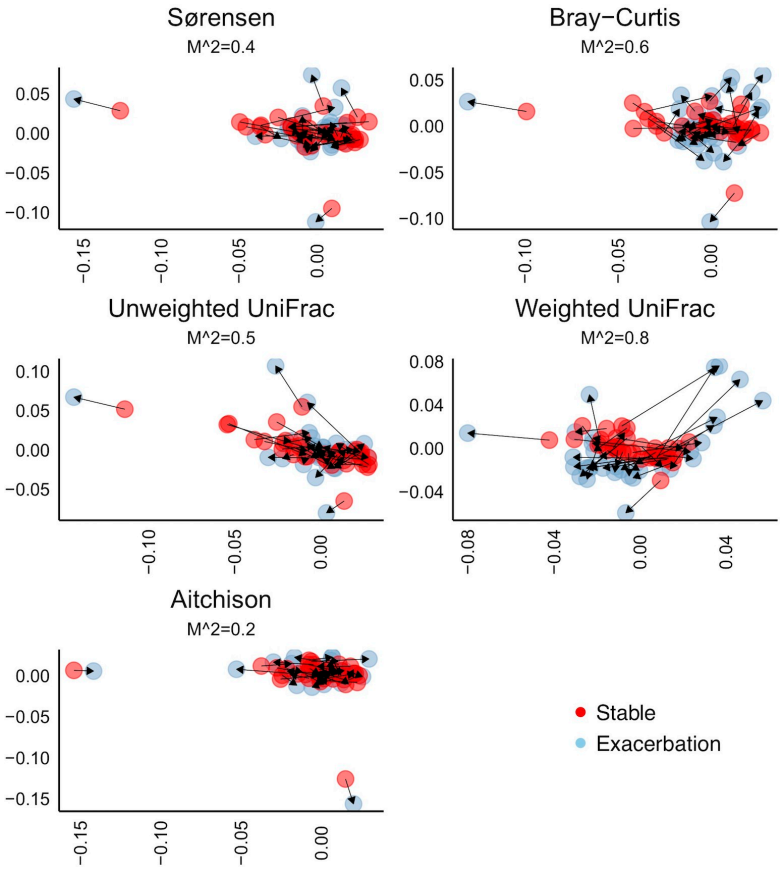

Fig 8. Non-metric multidimensional scaling plots after symmetric Procrustes transformation, illustrating differences in microbiota as distance between paired samples collected at stable state and during exacerbations of COPD. Distance matrices: Sørensen and Bray Curtis: Both non-phylogenetic; qualitative and quantitative information respectively. Unweighted and weighted UniFrac: Both phylogenetic; qualitative and quantitative information respectively. Aitchison: Compositional interpretation of sequencing data. M^2 = Summed squares of distances. The significance of M^2 was tested for all comparisons, with $p<0.05$ for all but WUF (PROTEST $p = 0.25$).

https://doi.org/10.1371/journal.pone.0222449.g008

Discussion

This study has shown that individual COPD patients had evident changes in the sputum microbiota from stable state to exacerbation, in parallel with significant changes in sputum inflammatory markers. The individual's changes in microbiota were to some extent camouflaged when analyses were run on groups of patients. Considerable shifts in bacterial composition were seen in the case study over 13 repeated stable state/exacerbation samples, but without a consistent stable state equilibrium.

COPD exacerbations are heterogeneous events, differing in length, symptom burden and need for treatment. In the current study, only patients who met the clinical criteria for an exacerbation, defined by the Wedzicha and Donaldsons' definition [29] and the judgment of

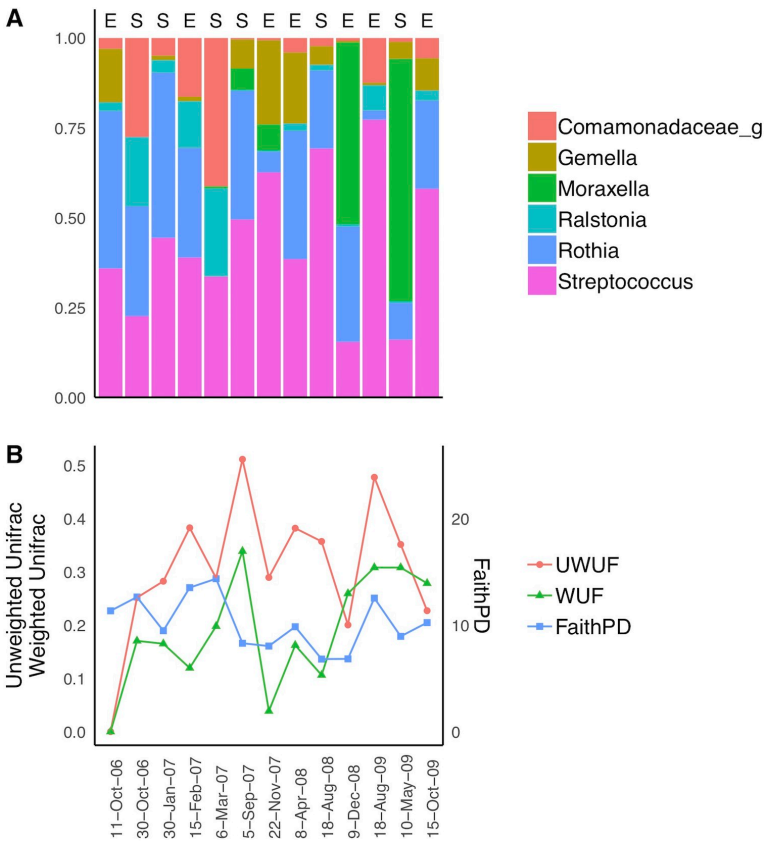

Fig 9. Taxonomic composition and diversity in 13 induced sputum samples collected from the same patient (chronic obstructive pulmonary disease) at different consultations. A: Presenting taxonomic composition at level 6 taxonomy (genus), provided by Silva database for amplicon sequence variants containing at least 1% of all sequences. Numbers are given as relative abundances per sample. B: Phylogenetic alpha- and beta-diversity. Alpha-diversity measured by Faith phylogenetic diversity (right y-axis). Non-quantitative and quantitative beta-diversity measured by UniFrac (UWUF, WUF respectively, left y-axis). Except from the first sampling time point beta diversity is calculated between consecutive samples. A+B: Disease state given in A by S = Stable state, E = Exacerbations. Samples are ordered chronologically and collection dates are given in B.

https://doi.org/10.1371/journal.pone.0222449.g009

an experienced study physician were included. All patients came to the outpatient clinic by themselves, and only those patients deemed not in need for hospitalization were considered for induced sputum sampling. Thus, all exacerbations were moderate at the time of sampling. Still, the sputum inflammatory markers confirmed an altered local immune state during these events, showing both that the exacerbation state was truly different from the stable state, and also that microbiota likely was affecting, or affected by, the airways inflammation.

We observed significantly higher levels of TNF-α, IP-10 and MIG during exacerbations. TNF-α is an upstream inflammatory cytokine with a wide range of effects. It has an important role in Th1-mediated immune responses, augmenting both IP-10 and MIG signaling downstream [30]. These are cytokines induced by interferon-gamma (IFN-γ) as part of a

Th1-mediated immune response [30]. All three cytokines have been shown to play a role against viral infections, intracellular bacteria and to some extracellular bacteria [31–33].

The AMPs are part of the innate immune response against a wide variety of microbes including bacteria, fungi and viruses. In a previous study from the BCCS and BCES, we have shown the same disease state related pattern of change as found in the current study [18]. In patients where the composition of the taxa in sputum changed with disease state, SLPI was significantly lower during exacerbations compared to those patients where the sputum composition was unchanged. Presumably this is a response to the microbial shift, for instance by degradation of SLPI by host and microbial proteases. However, in theory it could also be opposite; that during an exacerbation the immune response leads to changes in taxonomic compositions. In vitro studies are likely necessary to elucidate specific mechanisms. For the other markers, we could not find an association with shifts in taxonomic composition. Low sample size is perhaps the most likely explanation for this, in addition to the inherent heterogeneity of the COPD exacerbations.

The four most abundant phyla in our samples were coherent with previous studies on COPD sputum microbiota [34, 35]. It was the same four phyla dominating the samples independently of disease state, though we did see a shift involving increases in *Proteobacteria* during exacerbations, and a parallel decrease in *Bacteroidetes*. In the previous study by Mayhew et al [34], it was further shown that the fraction of *Proteobacteria* increased with increasing exacerbation severity, something the current study did not have power to examine. However, the current study adds to the other studies by showing an accompanied immune response with the shifts in microbial profiles.

Another important difference between our study and previous studies is that the current study only included induced sputum samples. We have previously shown that induced and spontaneous sputum collected during the same visits will not be sufficiently similar in microbial composition to allow them to be used interchangeably [17].

The most abundant genus belonging to the *Proteobacteria* phylum in our cohort was *Haemophilus*. This was the case for both stable state and exacerbation, and there were no significant changes in its abundance across disease states. Even though several studies have found *Haemophilus* to be of importance related to inflammation and exacerbation risk [36, 37], we could not find that *Haemophilus* discriminated between disease states when measured in induced sputum. This could reflect the sample size in the current study, and should not be interpreted as changes in *Haemophilus* being without importance. An imperative consideration when evaluating taxonomic composition is that increasing levels of one taxon invariably will result in decreases in others, since the sum total is 100%. We have used the Yue-Clayton index ($1\text{-}\theta_{YC}$) in an attempt to quantify the difference, but the cut-off value of 0.2 is arbitrary and no established consensus on what constitutes a biologically meaningful cut-off value exist. If the entire ecological content of a sample, or the overabundance of one low-abundant pathogen is more relevant to exacerbation risk, then a cut-off of 0.2 may be too high.

With that caveat, a very important finding in the current study was that there appeared to be significant changes in taxonomic composition when we examined individual (paired-samples) changes ($1\text{-}\theta_{YC} > 0.2$, $n = 26$) again confirming findings by Mayhew et al. However, we did not find significant changes in composition when all samples were pooled by disease state (Aldex2 analyses $p > 0.05$). Thus, paired analyses are necessary to evaluate changes in taxonomic compositions, and they confirm heterogeneity among patients.

With an infectious exacerbation, where the compositional taxonomy changes, one would imagine that the diversity would change as well. If one pathogenic organism dominated, it would presumably displace others completely (leading to a loss in richness) or skew the distribution significantly (leading to a loss of evenness).

In this study, we could not find a significant difference in alpha diversity between disease states on the group level with either non-phylogenetic or phylogenetic indices. However, we did detect higher diversity at stable state in patients with elevated ANC during exacerbations, indicating that reduced diversity can impair systemic immune responses. The plot showing individual changes revealed that alpha-diversity takes on all directionalities with changing disease states, thus larger numbers would be needed to look at sub-types in more detail.

For beta diversity, using several indices we again saw no convincing change in diversity from stable state to exacerbation with group comparisons, while such changes were supported when looking at diversity with paired analyses. Further, the larger change detected with weighted UniFrac than unweighted, could imply that we see predominately a change in pre-existing bacteria rather than addition or loss of new species.

Some methodological shortcomings need to be considered. First, we lack negative controls of the fluids used in the sputum induction in our study. We used the Decontam algorithm in R to identify likely contaminants, which were then excluded from the study. However, the lack of negative controls remains a weakness, as that could possibly have led to a more precise identification of contaminants. Second, over the three years of the study, two different technicians performed the initial processing of the samples, although with the same protocol. And, although the same study personnel later analyzed all samples with the same protocol, all paired samples were not always analyzed on the same laboratory runs. Analyses of taxonomy and beta diversity did not reveal clear differences in relative abundance between runs or significant differences in beta-diversity, and thus no adjustment for runs were used. However, some laboratory induced inter-pair variation cannot be excluded. Third, since the study compares pairs both where the stable state comes prior to the exacerbation and vice versa, the study is a comparison between disease states, and no chronological sequence of events can be assumed. Fourth, the sample size of the study is too small to make inferences about whether some subgroups like patients with different disease severity have larger variations in their microbiota than other subgroups. Fifth, variability between consecutive samples could not be addressed as participants delivered only one sample at each visit. Sixth, sputum was examined and discarded if the number of cells was < 1 million/mL or number of epithelial cells $> 20\%$. However, even if deemed representative of the lower airways, sputum will invariably contain some microbial contamination from the relatively high-biomass oral cavity. For less contamination prone sampling of the lower airways, bronchoscopy is preferred, however, that is not feasible during COPD exacerbations. Seventh, the true stability of the airways' microbiome is yet unknown, thus some of the change between the stable state and the exacerbations, may just reflect the fluctuating nature of the microbiome. Finally, amplicon sequencing only tells us which bacteria are present and their relative abundance based on the amplicon sequenced (in our case 16S rDNA).

This study not only confirms that there are considerable changes in the respiratory microbiota between disease states in COPD patients, it further shows an important heterogeneity between patient's microbiota. This is indicative of future challenges in development of applicable anti/pro-biotic treatment for groups of COPD patients. At the same time local inflammation is associated with the changes in microbiota, indicating the microbiota has significant implications for respiratory health. Further mechanistic studies are needed to examine the interaction between the microbiota and local inflammation.

Supporting information

S1 Text. More on bioinformatics and statistical methods. Supplementary details on bioinformatics and statistical methods.
(DOCX)

S1 Table. P-values for differential abundances of features between disease states at different taxonomic levels evaluated by use of Aldex2 in R. p*: FDR-corrected p-values, Wilcoxon test. Effect size cut off 0.5.
(XLSX)

Acknowledgments

The authors wish to thank Lene Svendsen, Tina Endresen-Vinsjevik, Eli Nordeide and Inge Zwetzig for help with data collection and sputum processing, and Tuyen Hoang and Christine Drengenes for help with PCR and sequencing.

Author Contributions

Conceptualization: Solveig Tangedal, Per S. Bakke, Tomas M. Eagan.

Data curation: Solveig Tangedal, Tomas M. Eagan.

Formal analysis: Solveig Tangedal.

Funding acquisition: Tomas M. Eagan.

Investigation: Solveig Tangedal, Marianne Aanerud, Per S. Bakke, Tomas M. Eagan.

Methodology: Louise J. Persson, Harald G. Wiker, Pieter S. Hiemstra.

Project administration: Solveig Tangedal, Harald G. Wiker, Per S. Bakke, Tomas M. Eagan.

Resources: Marianne Aanerud, Louise J. Persson, Harald G. Wiker, Per S. Bakke, Pieter S. Hiemstra, Tomas M. Eagan.

Supervision: Rune Nielsen, Harald G. Wiker, Tomas M. Eagan.

Visualization: Solveig Tangedal.

Writing – original draft: Solveig Tangedal, Tomas M. Eagan.

Writing – review & editing: Solveig Tangedal, Rune Nielsen, Marianne Aanerud, Louise J. Persson, Harald G. Wiker, Per S. Bakke, Pieter S. Hiemstra, Tomas M. Eagan.

References

1. Hilty M, Burke C, Pedro H, Cardenas P, Bush A, Bossley C, et al. Disordered microbial communities in asthmatic airways. PLoS One. 2010; 5(1):e8578. https://doi.org/10.1371/journal.pone.0008578 PMID: 20052417; PubMed Central PMCID: PMC2798952.
2. Sze MA, Dimitriu PA, Suzuki M, McDonough JE, Campbell JD, Brothers JF, et al. Host Response to the Lung Microbiome in Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2015; 192 (4):438–45. https://doi.org/10.1164/rccm.201502-0223OC PMID: 25945594; PubMed Central PMCID: PMC4595667.
3. Gronseth R, Drengenes C, Wiker HG, Tangedal S, Xue Y, Husebo GR, et al. Protected sampling is preferable in bronchoscopic studies of the airway microbiome. ERJ Open Res. 2017; 3(3). https://doi.org/10.1183/23120541.00019–2017 PMID: 28875147; PubMed Central PMCID: PMC5576223 openres.ersjournals.com.
4. O'Donnell R, Breen D, Wilson S, Djukanovic R. Inflammatory cells in the airways in COPD. Thorax. 2006; 61(5):448–54. https://doi.org/10.1136/thx.2004.024463 PMID: 16648353; PubMed Central PMCID: PMC2111192.
5. Barnes PJ. Inflammatory mechanisms in patients with chronic obstructive pulmonary disease. J Allergy Clin Immunol. 2016; 138(1):16–27. https://doi.org/10.1016/j.jaci.2016.05.011 PMID: 27373322.
6. Eagan TM, Ueland T, Wagner PD, Hardie JA, Mollnes TE, Damas JK, et al. Systemic inflammatory markers in COPD: results from the Bergen COPD Cohort Study. Eur Respir J. 2010; 35(3):540–8. https://doi.org/10.1183/09031936.00088209 PMID: 19643942.

7. Agusti AG, Noguera A, Sauleda J, Sala E, Pons J, Busquets X. Systemic effects of chronic obstructive pulmonary disease. Eur Respir J. 2003; 21(2):347–60. https://doi.org/10.1183/09031936.03.00405703 PMID: 12608452.

8. Caramori G, Ruggeri P, Di Stefano A, Mumby S, Girbino G, Adcock IM, et al. Autoimmunity and COPD: Clinical Implications. Chest. 2018; 153(6):1424–31. https://doi.org/10.1016/j.chest.2017.10.033 PMID: 29126842.

9. Hurst JR, Vestbo J, Anzueto A, Locantore N, Mullerova H, Tal-Singer R, et al. Susceptibility to exacerbation in chronic obstructive pulmonary disease. N Engl J Med. 2010; 363(12):1128–38. https://doi.org/10.1056/NEJMoa0909883 PMID: 20843247.

10. Sapey E, Stockley RA. COPD exacerbations. 2: aetiology. Thorax. 2006; 61(3):250–8. https://doi.org/10.1136/thx.2005.041822 PMID: 16517585; PubMed Central PMCID: PMC2080749.

11. Wilkinson TM, Hurst JR, Perera WR, Wilks M, Donaldson GC, Wedzicha JA. Effect of interactions between lower airway bacterial and rhinoviral infection in exacerbations of COPD. Chest. 2006; 129 (2):317–24. https://doi.org/10.1378/chest.129.2.317 PMID: 16478847.

12. Sethi S, Evans N, Grant BJ, Murphy TF. New strains of bacteria and exacerbations of chronic obstructive pulmonary disease. N Engl J Med. 2002; 347(7):465–71. https://doi.org/10.1056/NEJMoa012561 PMID: 12181400.

13. Wang Z, Bafadhel M, Haldar K, Spivak A, Mayhew D, Miller BE, et al. Lung microbiome dynamics in COPD exacerbations. Eur Respir J. 2016; 47(4):1082–92. https://doi.org/10.1183/13993003.01406-2015 PMID: 26917613.

14. Tangedal S, Aanerud M, Persson LJ, Brokstad KA, Bakke PS, Eagan TM. Comparison of inflammatory markers in induced and spontaneous sputum in a cohort of COPD patients. Respir Res. 2014; 15(1):138. https://doi.org/10.1186/s12931-014-0138-6 PMID: 25398249; PubMed Central PMCID: PMC4237726.

15. Global Strategy for the Diagnosis, Management and Prevention of Chronic Obstructive Pulmonary Disease 2019 report, Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2013. Available from: https://goldcopd.org/wp-content/uploads/2018/11/GOLD-2019-v1.7-FINAL-14Nov2018-WMS.pdf. Date last updated: January 2019. Date last accessed: February 18 2019.

16. Husebo GR, Bakke PS, Aanerud M, Hardie JA, Ueland T, Gronseth R, et al. Predictors of exacerbations in chronic obstructive pulmonary disease—results from the Bergen COPD cohort study. PLoS One. 2014; 9(10):e109721. https://doi.org/10.1371/journal.pone.0109721 PMID: 25279458; PubMed Central PMCID: PMC4184893.

17. Tangedal S, Aanerud M, Gronseth R, Drengenes C, Wiker HG, Bakke PS, et al. Comparing microbiota profiles in induced and spontaneous sputum samples in COPD patients. Respir Res. 2017; 18(1):164. https://doi.org/10.1186/s12931-017-0645-3 PMID: 28851370; PubMed Central PMCID: PMC5576328.

18. Persson LJ, Aanerud M, Hardie JA, Miodini Nilsen R, Bakke PS, Eagan TM, et al. Antimicrobial peptide levels are linked to airway inflammation, bacterial colonisation and exacerbations in chronic obstructive pulmonary disease. Eur Respir J. 2017; 49(3). https://doi.org/10.1183/13993003.01328–2016 PMID: 28298400.

19. Tjabringa GS, Vos JB, Olthuis D, Ninaber DK, Rabe KF, Schalkwijk J, et al. Host defense effector molecules in mucosal secretions. FEMS Immunol Med Microbiol. 2005; 45(2):151–8. https://doi.org/10.1016/j.femsim.2005.03.004 PMID: 16051067.

20. Bolyen E RJ, Dillon MR, Bokulich NA, Abnet C, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope E, Da Silva R, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley G, Janssen S, Jarmusch AK, Jiang L, Kaehler B, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MG, Lee J, Ley R, Liu Y, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton J, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson II MS, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CH, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG. QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science. PeerJ Preprints. 2018;6:e27295v2. doi: 10.7287.

21. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016; 13(7):581–3. https://doi.org/10.1038/nmeth.3869 PMID: 27214047; PubMed Central PMCID: PMC4927377.

22. Rognes T, Flouri T, Nichols B, Quince C, Mahe F. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 2016; 4:e2584. https://doi.org/10.7717/peerj.2584 PMID: 27781170; PubMed Central PMCID: PMC5075697.

23. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome. 2018; 6 (1):226. https://doi.org/10.1186/s40168-018-0605-2 PMID: 30558668; PubMed Central PMCID: PMC6298009.

24. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013; 41(Database issue):D590–6. https://doi.org/10.1093/nar/gks1219 PMID: 23193283; PubMed Central PMCID: PMC3531112.

25. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol. 2009; 26(7):1641–50. https://doi.org/10.1093/molbev/msp077 PMID: 19377059; PubMed Central PMCID: PMC2693737.

26. Yue JC, Clayton MK. A similarity measure based on species proportions. Commun Stat Theory Methods. 2005; 34(11):2123–31. https://doi.org/10.1080/sta-200066418

27. Gloor GB, Reid G. Compositional analysis: a valid approach to analyze microbiome high-throughput sequencing data. Can J Microbiol. 2016; 62(8):692–703. https://doi.org/10.1139/cjm-2015-0821 PMID: 27314511.

28. Jari Oksanen FGB, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs, Helene Wagner. vegan: Community Ecology Package. R package version 2.5–2. 2018.

29. Seemungal TA, Donaldson GC, Bhowmik A, Jeffries DJ, Wedzicha JA. Time course and recovery of exacerbations in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2000; 161(5):1608–13. https://doi.org/10.1164/ajrccm.161.5.9908022 PMID: 10806163.

30. Fenwick PS, Macedo P, Kilty IC, Barnes PJ, Donnelly LE. Effect of JAK Inhibitors on Release of CXCL9, CXCL10 and CXCL11 from Human Airway Epithelial Cells. PLoS One. 2015; 10(6):e0128757. https://doi.org/10.1371/journal.pone.0128757 PMID: 26090665; PubMed Central PMCID: PMC4474874.

31. Liu M, Guo S, Hibbert JM, Jain V, Singh N, Wilson NO, et al. CXCL10/IP-10 in infectious diseases pathogenesis and potential therapeutic implications. Cytokine Growth Factor Rev. 2011; 22(3):121–30. https://doi.org/10.1016/j.cytogfr.2011.06.001 PMID: 21802343; PubMed Central PMCID: PMC3203691.

32. Egesten A, Eliasson M, Johansson HM, Olin AI, Morgelin M, Mueller A, et al. The CXC chemokine MIG/CXCL9 is important in innate immunity against Streptococcus pyogenes. J Infect Dis. 2007; 195 (5):684–93. https://doi.org/10.1086/510857 PMID: 17262710.

33. Waters JP, Pober JS, Bradley JR. Tumour necrosis factor in infectious disease. J Pathol. 2013; 230 (2):132–47. https://doi.org/10.1002/path.4187 PMID: 23460469.

34. Mayhew D, Devos N, Lambert C, Brown JR, Clarke SC, Kim VL, et al. Longitudinal profiling of the lung microbiome in the AERIS study demonstrates repeatability of bacterial and eosinophilic COPD exacerbations. Thorax. 2018; 73(5):422–30. https://doi.org/10.1136/thoraxjnl-2017-210408 PMID: 29386298; PubMed Central PMCID: PMC5909767.

35. Huang YJ, Sethi S, Murphy T, Nariya S, Boushey HA, Lynch SV. Airway microbiome dynamics in exacerbations of chronic obstructive pulmonary disease. J Clin Microbiol. 2014; 52(8):2813–23. https://doi.org/10.1128/JCM.00035-14 PMID: 24850358; PubMed Central PMCID: PMC4136157.

36. Su YC, Jalalvand F, Thegerstrom J, Riesbeck K. The Interplay Between Immune Response and Bacterial Infection in COPD: Focus Upon Non-typeable Haemophilus influenzae. Front Immunol. 2018; 9:2530. https://doi.org/10.3389/fimmu.2018.02530 PMID: 30455693; PubMed Central PMCID: PMC6230626.

37. Tufvesson E, Bjermer L, Ekberg M. Patients with chronic obstructive pulmonary disease and chronically colonized with Haemophilus influenzae during stable disease phase have increased airway inflammation. Int J Chron Obstruct Pulmon Dis. 2015; 10:881–9. https://doi.org/10.2147/COPD.S78748 PMID: 26005341; PubMed Central PMCID: PMC4427610.

Graphic design: Communication Division, UiB / Print: Skipnes Kommunikasjon AS

uib.no

ISBN: 9788230859032 (print)
9788230857311 (PDF)