The mycobiome in COPD: Descriptive and longitudinal analysis, and participation in research bronchoscopy studies

Einar Marius Hjellestad Martinsen

Thesis for the degree of Philosophiae Doctor (PhD)
University of Bergen, Norway
2021

UNIVERSITY OF BERGEN

The mycobiome in COPD: Descriptive and longitudinal analysis, and participation in research bronchoscopy studies

Einar Marius Hjellestad Martinsen

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

Date of defense: 25.06.2021

Year: 2021

Title: The mycobiome in COPD: Descriptive and longitudinal analysis, and participation in research bronchoscopy studies

Name: Einar Marius Hjellestad Martinsen

Print: Skipnes Kommunikasjon / University of Bergen

Table of Contents

1 Scientific environment

The current thesis had its main site at Haukeland University Hospital, Department of Thoracic Medicine. The thesis is anchored in the Bergen COPD Microbiome study (short name “MicroCOPD”) conducted by The Bergen Respiratory Research Group, led by professor Tomas Mikal Lind Eagan. I have been connected to the Faculty of Medicine, University of Bergen, during the entire thesis work. First as an attendant at The Medical Student Research Programme (MSRP), and later as a PhD candidate. The MSRP is based on one year of full-time research and two years of part-time research. I completed my full-time research as a MSRP attendant in 2013-2014, and was employed as a PhD candidate from 2018 till 2021. My PhD work, including the MSRP, was funded by the University of Bergen.

Main supervisor during the thesis have been:

Rune Nielsen, associate professor, MD. Department of Clinical Science, Faculty of Medicine, University of Bergen.

Co-supervisors during the thesis have been:

Tomas Mikal Lind Eagan, professor, MD. Department of Clinical Science, Faculty of Medicine, University of Bergen.

Harald Gotten Wiker, professor, MD. Department of Clinical Science, Faculty of Medicine, University of Bergen.

2 Acknowledgements

2.1 English acknowledgements

It is the 13th of February in one of the coldest winters I have ever experienced. Three layers of wool, strict COVID-19 restrictions, and home office has become the new everyday life. Some days I have felt as if I was one of the miserable characters in the PhD comics, and honestly, I have frequently questioned myself the last weeks if the PhD work was really worth it. But the answer is simple – yes, it was definitely worth it! The PhD period has given me experiences and friendships for life. I have enjoyed the times with colleagues, travels to courses, stays abroad, and teaching. It has given me several ups and downs, and has contributed to my development as a person and as an academic.

I am grateful to the University of Bergen for the opportunity to do my PhD. I want to give extra thanks to The Medical Student Research Programme at the Faculty of Medicine (MSRP) with Marianne Heldal Stien as front figure. I read about the MSRP for the first time in upper secondary school, and realised it could fulfil my idyllic childhood dream of being a researcher. In my opinion, it is an excellent opportunity. I hope the MSRP will continue to exist and become even more popular. I am also very thankful to all study participants. Without your contribution, none of this could have been done.

Rune, my main supervisor, deserves a huge (I am finally allowed to use those strong adjectives, although it is one of the most corrected things in my paper drafts) thanks! I am forever grateful for your supervision. I have so many times admired your intellect, but your intellect is not what have made you such a great supervisor. Despite being somewhat different as persons, I have felt that you have seen me for who I am and considered me as more than just your student, but also as your friend. You have pushed me forward in tough and ineffective times, and found light and possibilities in the darkest places. You have taught me to stand up for myself. No matter how convincing someone seems to be, I have realised that they can always be argued against. Especially if they

possess a higher academic degree… Thank you for all the help and patience through these ten years. And remember – I chose this project because of the group and the enthusiastic supervisor, not because of an interest in lung medicine.

A special thank you to my co-supervisor Tomas as well. You are one of the most structured and hardworking people I have ever known. Your email replies in the middle of the night and early mornings tells a lot about your capacity and dedication. For a fresh medical student, it was striking to see your calmness and communication skills with the participants. That is really something to live up to. I greatly appreciate your supervision and the genuine interest you have shown in the countless drafts I have bothered you with. To Harald, my second co-supervisor – thank you for some interesting microbiologic discussions on the phone. I appreciate your contribution and sharing of knowledge. I also want to thank Per for your supervision during MSRP.

Dear Elise, my closest colleague and one of my dearest friends. We have known each other since the first year of medical school. I admit being sceptical about sharing supervisor and collaborate on much of the work, but how wrong one can take. I dare to assert that without you, I might not have finished. Always positive and helpful, and more effective than most of us. We have shared nearly everything – highs and lows, lunches, office, strange and personal medical issues, love grief, life crisis, frustrations, hotel rooms, and probably a lot more. I was once asked by the person responsible for the summer job employment at Haugesund hospital if I knew you. I told him you were the brightest student in the class, and that he would not be disappointed if he chose you. The job was yours. Later, you have accused me several times for lying. Well, the fact is, I didn't… Thank you for a great time, Elise!

To Jian Hao, one of my greatest friends. Thank you for fruitful medical discussions, lunches, and social gatherings outside work. You have always been helpful to me, and I

appreciate our friendship a lot. I dream that we end up with internship at the same hospital. You are anyway always welcome in our home, wherever we might end up.
A special thank you to Ane, Christine, Ingvild, and Nazar. Thank you for the lunches and trips during the PhD. I have always looked forward to our social meetings, and value that we could both express frustrations and talk about other things than work and research.

The Bergen Respiratory Research Group – thank you for a safe and including working environment. A big thank you to all who have helped me made this PhD possible. Thank you to Øistein, Marit, Tharmini, Randi, Tuyen, Lise, Hildegunn, Stine, Kristina, and Tove for aid in data collection. And a special thank you to Eli. Our conversations and your kindness have meant a lot to me. Your cordiality has been invaluable. Thanks to Solveig, Gunnar, Sverre, Kristel, Christina, Bahareh, Bernt, Einar, and Marta for valuable discussions, teaching, and social meetings.

Walter, Andreu, and the rest of my Spanish friends at Sequentia – thank you so much for your hospitality and help. My stays in Barcelona are some of my best PhD memories, and I would never have reached my goal without your bioinformatic teaching. Thank you!

To my loving family – thank you for believing in me and for your support. You have always been there for me, and have encouraged me to follow my dreams. My weekend and holiday trips away from Bergen have reminded me that there is definitely more to life than medical school and PhD work, and I have to admit that I have become extremely spoiled during these timeouts. I have always felt an endless love from all of you.

A great thank you to Malena, my wife and best friend. You say you have a poor patience, but the last months with me have proven the opposite. I am without words to describe your flexibility and tolerance during my PhD period. I apologise for the lack of presence the last months, and for my roller-coaster emotional life lately. I am deeply thankful for your endless support, love, and care. The PhD grade is partly yours as well…

2.2 Norwegian acknowledgements (ytringar av takksemd)

Kalenderen viser 13. februar, og vi er midt inni ein av dei kaldaste vintrane eg kan hugse. Tre lag med ull, strenge COVID-19-restriksjonar og heimekontor er blitt den nye kvardagen. Nokre dagar har eg følt meg som ein av dei ulukkelege figurane i «PhD comics», og heilt ærleg har eg spurd meg sjølv gjentatte gonger om heile PhD-arbeidet eigentleg var verdt det. Men svaret er enkelt – ja, det var definitivt verdt det! PhD-perioden har gitt meg erfaringar og venskap for livet. Eg har sett pris på tida med kollegaar, reiser til kurs, utanlandsopphald og undervising. Det har gitt meg fleire opp- og nedturar, og har bidratt til å utvikle meg som person og som akademikar.

Eg er takknemleg overfor Universitetet i Bergen som gav meg moglegheita for å gjennomføre PhD-grada. Eg vil rette ein ekstra takk til Forskarlina ved Det medisinske fakultet med Marianne Heldal Stien i spissen. Eg las om Forskarlina for første gong på vidaregåande, og skjønte at ho kunne oppfylle den idylliske barndomsdraumen min om å bli ein forskar. Etter mi meining er Forskarlina ei glimrande moglegheit. Eg håpar at Forskarlina vil stå ved lag og attpåtil bli endå meir populær. Eg er også svært takknemleg overfor studiedeltakarane. Dette ville ikkje vore mogleg utan dykk.

Rune, hovudrettleiaren min, fortener ein enorm (eg får endeleg lov til å bruke slike sterke adjektiv, sjølv om det er ein av dei mest korrigerte tinga i artikkelutkasta mine) takk! Eg vil alltid vere takknemleg for rettleiinga di. Eg har sett på intellektet ditt med vørdnad mang ein gong, men det er ikkje intellektet som har gjort deg til ein så god rettleiar. Sjølv om vi er noko ulike som personar, har eg følt at du har sett meg for den eg er og for noko meir enn berre studenten din, men også som venen din. Du har skuva meg fram i dei tøffaste og mest ineffektive tider, og har funne ljos og moglegheiter på dei mørkaste stadene. Du har lært meg å stå opp for meg sjølv. Uansett kor overtydande nokon kan verke, har eg skjønt at dei alltid kan argumenterast mot. Særleg om dei held ei høgare

akademisk grad... Takk for all hjelp og tolmod gjennom desse ti åra. Og hugs – eg valde prosjektet grunna gruppa og rettleiar, ikkje grunna interesse for lungemedisin.

Ein særleg takk er på sin plass til min medrettleiar Tomas også. Du er ein av dei mest strukturerte og hardtarbeidande personane eg har kjent nokon gong. Svara dine på e-post kan kome midt på natta og grytidleg på morgonen, og fortel mykje om din kapasitet og dedikasjon. Det var slåande for meg som fersk medisinstudent å sjå korleis du roa og kommuniserte med deltakarane, og det var verkeleg noko å leve opp til. Eg set stor pris på rettleiinga di og di genuine interesse for dei tallause utkasta eg har plaga deg med.
Til Harald, min andre medrettleiar – takk for nokre interessante mikrobiologiske diskusjonar på telefonen. Eg set pris på ditt bidrag og kunnskapsdeling. Eg vil også takke Per for di rettleiing under Forskarlina.

Kjære Elise, min næraste kollega og ein av mine beste vener. Vi har kjent kvarandre sidan det første året på medisinstudiet. Eg må innrømme at eg var skeptisk til å dele rettleiar og samarbeide om mykje av arbeidet, men kor feil kan ein ta. Eg vil våge å påstå at utan deg hadde eg kanskje ikkje fullført. Alltid positiv og hjelpsam, og meir effektiv enn dei fleste av oss. Vi har delt så godt som alt – opp- og nedturar, lunsjar, kontor, merkelege og personlege medisinske problem, kjærleikssorg, livskriser, frustrasjon, hotellrom og truleg mykje meir også. Eg blei ein gong spurd av tilsetjingsansvarleg for sommarjobbane ved Haugesund sjukehus om eg kjente til deg. Eg sa til han at du var den smartaste studenten på kullet, og at han ikkje ville bli skuffa om han valde deg. Jobben var din. Seinare har du fleire gonger skulda meg for å ha loge. Vel, sanninga er at det gjorde eg ikkje... Tusen takk for ei fantastisk tid, Elise!

Til Jian Hao, ein av mine beste vener. Takk for dei fruktbare medisinske diskusjonane, lunsjane og sosiale møta utanom arbeid. Du har alltid vore hjelpsam med meg, og eg set stor pris på venskapet vårt. Eg har ein draum om at vi skal vere LIS1 på det same sjukehuset. Du er uansett alltid velkomen hjå oss, uansett kvar vi skulle ende opp.

Ein spesiell takk til Ane, Christine, Ingvild og Nazar. Takk for alle lunsjane og turane vi har hatt gjennom PhD-grada. Eg har alltid sett fram til å treffe dykk, og verdset at vi både kunne uttrykke frustrasjon, men også snakke om andre ting enn arbeid og forsking.

Lungeforskingsgruppa – takk for eit trygt og inkluderande arbeidsmiljø. Ein stor takk til alle som har hjelpt meg med å gjere denne PhD-grada mogleg. Takk til Øistein, Marit, Tharmini, Randi, Tuyen, Lise, Hildegunn, Stine, Kristina og Tove for hjelp med datainnsamling. Ein spesiell takk til Eli. Samtalane våre og godheita di har betydd mykje for meg. Hjartevarmen din har vore uvurderleg. Takk til Solveig, Gunnar, Sverre, Kristel, Christina, Bahareh, Bernt, Einar og Marta for verdifulle diskusjonar, undervising og sosiale møte.

Walter, Andreu, og resten av mine spanske vener ved Sequentia – takk for gjestmildskapen og hjelpa dykkar. Opphalda mine i Barcelona er somme av dei beste PhD-minna, og eg hadde aldri nådd måla mine utan dykkar bioinformatiske læring. Takk!

Til min kjære familie – takk for at de trur på meg og for støtta dykkar. De har alltid vore der for meg, og har oppmuntra meg til å følgje draumane mine. Helge- og ferieturane mine bort frå Bergen har minna meg på at det definitivt finst meir i livet enn medisinstudiet og PhD-arbeide, og eg må innrømme at eg har blitt ekstremt bortskjemt i løpet av desse pausane. Eg har alltid kjent på ein endelaus kjærleik frå dykk alle.

Ein stor takk til Malena, mi kone og min beste ven. Du seier at du ikkje er så tolmodig, men dei siste månadene med meg har vist det motsette. Eg er utan ord til å skildre den fleksibiliteten og toleransen som du har vist gjennom min PhD-periode. Eg må berre be om orsaking for å ha vore så lite til stades dei siste månadene og for å ha hatt eit kjensleliv som ein berg-og-dal-bane i det siste. Eg er djupt takknemleg for di endelause støtte, kjærleik og omsorg. PhD-grada er delvis di også...

3 Terms and abbreviations

16S rRNA	16S ribosomal RNA.
ALDEx2	The second version of ANOVA-Like Differential Expression.
Alpha diversity	The level of diversity found within a single sample. Includes several metrics, for instance number of observed ASVs (richness) or Shannon index (richness and evenness/distribution).
Amplicon	DNA product of DNA amplification via polymerase chain reaction.
ANCOM v2	The second version of analysis of composition of microbiomes.
ASV	Amplicon sequence variant. The result of grouping of DNA sequences from the marker gene analysis. Represents features or taxa in a microbiome study.
BAL	Bronchoalveolar lavage.
BCCS	The Bergen COPD cohort study.
Beta diversity	The level of diversity or dissimilarity found between samples. Used to examine whether samples within a group are more similar to each other than those in another group.
BLASTN	The Nucleotide Basic Local Alignment Search Tool.
BOLD	The Burden of Obstructive Lung Diseases.

CAT	COPD assessment test.
CF	Cystic fibrosis.
COPD	Chronic obstructive pulmonary disease.
COPD exacerbations	Periods with worsening of respiratory symptoms experienced by COPD patients.
CRP	C-reactive protein.
CT	Computed tomography.
Culture-independent sequencing techniques	Methods for microbial identification without the use of conventional culture-based approaches, i.e., not requiring culturing of microorganisms in the laboratory.
DADA2	The Divisive Amplicon Denoising Algorithm version 2.
Diversity	The richness and/or distribution of taxa in a sample and similarity/dissimilarity of taxonomic composition between samples.
Embase	A biomedical research and literature database provided by the medical publisher Elsevier.
FEV_1	Forced expiratory volume in one second.

FVC	Forced vital capacity.
GOLD	The Global Initiative for Chronic Obstructive Lung Disease.
HMP	The Human Microbiome Project.
IBD	Inflammatory bowel disease.
ICS	Inhaled corticosteroids.
ICU	Intensive care unit.
IL-8	Interleukin 8.
IS	Induced sputum.
ISHAM	The International Society of Human and Animal Mycology.
ITS region	The internal transcribed spacer region of the fungal ribosomal RNA gene cluster. Used for sequencing and fungal identification. Divided into ITS1 and ITS2.
LEfSe	Linear discriminant analysis effect size.
LLN	The lower limit of normal.
M^2	The summed squares of deviations/measure of fit in a Procrustes analysis.

mcL	Microlitre.
MeSH	Medical subject headings.
Metagenomic sequencing	A way of studying all the genetic material including DNA from host and microbes. Could be achieved by shotgun random sequencing of total DNA in a sample.
mg	Milligram.
Microbiome	The genetic information of microorganism within a defined habitat and the properties of the microorganisms' gene products.
MicrobiomeDDA	The Microbiome Differential Distribution Analysis omnibus test.
Microbiota	The collection of microorganisms found within a habitat, such as the human body or the lungs.
MicroCOPD	The Bergen COPD Microbiome study.
mL	Millilitre.
mm	Millimetre.
MMPs	Matrix metalloproteinases.
mMRC	Modified Medical Research Council dyspnoea scale.

MSRP	The Medical Student Research Programme at The Faculty of Medicine, University of Bergen.
Mycobiome	The fungal microbiome, i.e. the collection of fungal organisms in a defined habitat.
NCBI	The National Center for Biotechnology Information.
NCS	Negative control sample.
NGS	Next-generation sequencing.
OLD	Obstructive lung disease.
OTUs	Operational taxonomic units. Used to cluster microorganisms based on similarity of DNA sequences of a specific taxonomic marker gene. Now widely replaced by ASVs.
OW	Oral wash.
Pack years	Number of cigarettes smoked per day divided by 20, and then multiplied with the number of years smoked.
PAMPs	Pathogen-associated molecular patterns.
PBS	Phosphate buffered saline.
PCoA	Principal coordinates analysis.

PCR	Polymerase chain reaction.
PERMANOVA	Permuted analysis of variance.
PERMDISP	Permuted multivariate analysis of beta-dispersion.
PICO	A mnemonic for Patient/problem/population, Intervention, Comparison, and Outcome. Used to define research questions, for instance in a literature search.
PIL	Patient information leaflet.
PMA	Propidium monoazide.
Prebiotic	Administered molecules that promote growth of specific microorganisms.
PRISMA	The Preferred Reporting Items for Systematic reviews and Meta-Analyses.
Probiotic	Microorganisms administered to patients with therapeutic intent.
PRR	Pattern-recognition receptors.
PSB	Protected specimen brushings.
PubMed	An archive of biomedical and life sciences journal literature at the U.S. National Institutes of Health's National Library of Medicine.

QIIME	Quantitative insights into microbial ecology.
Rarefaction	A process which subsamples each sample to a given rarefaction or sampling/sequencing depth without replacement. Samples with a sequence count below this value will be discarded.
rDNA	Ribosomal DNA.
Relative abundance	The proportion of an ASV measured as the absolute count of sequence reads for this ASV in a particular sample divided by the total number of sequence reads in the particular sample.
RLL	The right lower lobe.
RML	The right middle lobe.
SML	Small-volume lavage.
Taxon	A unit of organisms which are related and share common characteristics, for instance a fungus.
Taxonomy	The classification or arrangement of groups of biological organisms based on shared characteristics.

4 Abstract

Background

Chronic obstructive pulmonary disease (COPD) is a high-prevalent lung disease with high mortality rates. The COPD pathogenesis is only partly understood, and we do not know why some risk factor exposed subjects develop the disease. With the introduction of culture-independent sequencing techniques, it has been shown that healthy lung is not sterile, and associations between colonising bacteria, or the lung microbiota, and diseases have been proposed. Early lung microbiome studies, however, focused exclusively on the bacteria, but lungs also contain viruses, fungi, and other eukaryotes. Few studies have examined the fungal community, the mycobiome, in the lungs of COPD patients. Protected bronchoscopic sampling seems to be a well-suited method to collect samples from the COPD lung mycobiome. But the invasive nature of the bronchoscopy procedure might make the recruitment of participants challenging. To facilitate large-scale bronchoscopy studies on the lung mycobiome, we need information on participation in studies involving a bronchoscopy.

Aims

The aim of the thesis was to examine participation in research bronchoscopy studies, and to analyse the lung mycobiome of participants in the Bergen COPD Microbiome study (short name «MicroCOPD»). In paper I, we sought to look for information on motivation to participate, participation barriers, response rates, and recruitment strategies in research studies involving a bronchoscopy by reviewing the literature. In paper II, we wanted to examine participation in the MicroCOPD study by reporting response rates and participation motives and declining reasons. In paper III, our aim was to examine the lung mycobiome in subjects with or without COPD, and to look for effects on the mycobiomes from inhaled corticosteroids (ICS) use. Finally, in paper IV, we aimed to assess the stability of the lung mycobiome of the participants in the MicroCOPD study, and to determine if intercurrent antibiotic use influences the stability.

Material and methods

We performed a literature review on participation in research bronchoscopy studies using the search engines of PubMed and EMBASE. Titles and abstracts of retrieved papers were sifted according to prespecified criteria, and included papers were subject for a more in-depth review.

Individuals with and without COPD or asthma were invited to participate in the MicroCOPD study. Participants underwent at least one bronchoscopy with sample collection by bronchoalveolar lavage (BAL), and a structured interview regarding medical history, medication use, and smoking habits. Additionally, post-bronchodilator spirometry was performed. Subjects that declined participation at the screening interview were asked about their reason not to participate, while subjects accepting participation were asked about their motivation to participate just prior to the bronchoscopy using an open question. All participants also provided an oral wash (OW) sample, and we collected some of the phosphate-buffered saline used in samples to serve as a negative control sample (NCS).

Fungal DNA was extracted from the OW, BAL, and NCS samples, before sequencing of the ITS1 region on an Illumina HiSeq sequencing platform. We compared the taxonomic composition and alpha and beta diversity between participant groups and by ICS use (paper III), and between the first and the second bronchoscopy in paper IV.

Results

Only seven papers were included for in-depth reading in the literature review on research bronchoscopy studies in the literature (paper I), reflecting a paucity of information on this topic. Still, data suggested that most participated for personal benefit and altruistic reasons, while fear and inconvenience hindered participation. Response rates varied from 3 to 73%, and radio advertisement was the most effective recruitment strategy.

The response rate in the MicroCOPD study was just above 50%, and men had a significantly higher response rate than women. Procedural fear was reported as the most common non-response reason. The most frequently stated participation motive was personal health benefit, but after merging participation motives into broader categories, altruism was the most frequent main motive. Men were less likely to state altruism as their main motive.

In paper III, we found that most samples, both OW and BAL, were dominated by *Candida*. *Malassezia* and *Sarocladium* were also frequently found taxa in BAL samples. There was more *Candida* in OW samples compared to BAL samples, and analyses suggested that beta diversity differed significantly between OW and BAL samples. Neither taxonomy nor alpha or beta diversity analyses found consistent differences between participant groups. The mycobiomes did not seem to be affected by use of ICS.

Taxonomy differed more between the repeated bronchoscopies for BAL samples than OW samples, while no apparent effect was seen from participant category and intercurrent antibiotic use (paper IV). Alpha and beta diversities were consistent by time.

Conclusions

We have shown that most participants in research bronchoscopy studies participate for personal benefit or altruism, and that most decliners fear the bronchoscopy (paper II), in line with what is known from the literature (paper I). A response rate just above 50% in the MicroCOPD study means that large-scale bronchoscopy studies are feasible.

Oral and pulmonary samples differed in taxonomic composition and diversity, possibly indicating the existence of a pulmonary mycobiome (paper III). We have also shown that the lung mycobiome is less stable than the oral mycobiome, and neither COPD diagnosis, intercurrent antibiotic use, nor time between bronchoscopies seemed to influence the stability (paper IV).

5 List of publications

Paper I

Martinsen EM, Leiten EO, Bakke PS, Eagan TM, Grønseth R. Participation in research bronchoscopy: a literature review. Eur Clin Respir J. 2016;3:29511.

Paper II

Martinsen EMH, Eagan TML, Leiten EO, Nordeide E, Bakke PS, Lehmann S, Nielsen R. Motivation and response rates in bronchoscopy studies. Multidiscip Respir Med. 2019;14:14.

Paper III

Martinsen EMH, Eagan TML, Leiten EO, Haaland I, Husebo G, Knudsen KS, Drengenes C, Sanseverino W, Paytuví-Gallart A, Nielsen R. The pulmonary mycobiome - a study of subjects with and without chronic obstructive pulmonary disease. Accepted for publication in PLOS ONE.

Paper IV

Martinsen EMH, Eagan TML, Wiker HG, Leiten EO, Husebo G, Knudsen KS, Tangedal S, Sanseverino W, Paytuví-Gallart A, and Nielsen R. A longitudinal study of the pulmonary mycobiome in subjects with and without chronic obstructive pulmonary disease. Submitted.

6 Background

Respiratory diseases are afflicting millions of people worldwide, and one of the most common is chronic obstructive pulmonary disease (COPD). COPD was the third most frequent cause of death in the world in 2016 with approximately 2.9 million registered deaths (1). The global number of deaths globally due to COPD increased from 2006 to 2016 with a percentage change of 5.5 (1), and the COPD burden has been projected to increase in coming years in part due to an ageing population (2). The Nord-Trøndelag Health Study (HUNT) in Norway report that prevalence of pre-bronchodilator COPD was 14.8% in 2006–2008 using fixed-ratio criteria, and 7.3% in 2006–2008 using lower limit of normal criteria. Furthermore, the annual treatment-related societal costs of COPD in Norway are more than 1,000,000,000 NOK, and are driven by those with most severe disease (3).

Despite extensive research, there is yet significant knowledge gaps of the COPD pathogenesis. It is still unclear why some smokers develop COPD while others do not. Additionally, patients with COPD are constantly at risk of experiencing periods with worsening of respiratory symptoms, so-called COPD exacerbations. Some patients become “frequent exacerbators”, whereas other patients nearly never experience these events, for reasons yet to be explained. Bhowmik et al. have found that frequent exacerbators have shown heightened airway inflammation at stable state (4), but potential confounding factors such as inhaled corticosteroids (ICS) treatment, smoking, or bacterial colonisation make results unclear. It has also been suggested that increased inflammation might be due to the presence or load of bacteria in the lower airways (5). The common perception was for a long time that healthy lungs were sterile, and that a failure of innate immune mechanisms in COPD allows bacteria to proliferate and persist in the airways (6). Furthermore, COPD exacerbations were thought to happen as a result of infections (6). However, with the advent of techniques independent on culturing of microorganisms, it has been shown that there exists a diverse low-biomass environment of pulmonary

microorganisms also in healthy subjects (7). The collective microbes were termed "microbiota", and associations between the pulmonary microbiota and diseases were proposed (8).

That a disturbed lung microbiota is related to the development of respiratory disease is an appealing concept, but microbiota studies are often limited by low sample sizes and challenging sampling. Reliable sampling of the lower airways could be done using protected techniques during bronchoscopy to avoid contamination from oral and upper respiratory microbiota (9). The semi-invasive nature of a bronchoscopy, however, can be associated with discomfort and anxiety. Including a bronchoscopy as part of the study might lower participation, and is one of the challenges microbiota researchers face.

Another challenge is that the early airway microbiota studies have focused exclusively on the bacteria present in the lungs. Fungi, on the other side, are an important but hitherto largely ignored part of the airway microbiome. The fungal microorganisms found in the airways are called the airway mycobiota. There exists some evidence that mycobiotas harbour different body sites including the airways, and that these mycobiotas might be associated with various diseases (10). The study of mycobiota and respiratory diseases is still in its infancy, and future lung mycobiota studies should focus on the changes in the fungi present under health and respiratory diseases.

6.1 Chronic obstructive pulmonary disease (COPD)

COPD is a disease of the airways and lung characterised by persistent respiratory symptoms and airflow limitation (11). Shortness of breath, or dyspnoea, is a main symptom of COPD (12). Another important symptom of COPD is cough, which might be accompanied with sputum production (12). COPD is diagnosed in patients with these chronic airway symptoms and chronic airflow obstruction verified by a lung function test (spirometry). Chronic airflow obstruction is defined as a lower-than-normal ratio of forced expiratory volume in one second (FEV_1) to forced vital capacity (FVC) after administration of a bronchodilating drug (post-bronchodilator). Various cut-off values exist, but frequently, a fixed value of 0.7 is used (13). The disease is often staged according to the degree of obstruction, indicated by post-bronchodilator FEV_1 values. The Global Initiative for Chronic Obstructive Lung Disease (GOLD), categorises FEV_1 in percentage of predicted, and patients are often labelled as GOLD 1 (mild, $FEV_1 > 80\%$ of predicted), GOLD 2 (moderate, FEV_1 50-80% of predicted), GOLD 3 (severe, FEV_1 30-50% of predicted), or GOLD 4 (very severe, $FEV_1 < 30\%$ of predicted) (14).

A major risk factor for COPD is cigarette smoking (15). But never smokers also comprise a fair proportion of COPD patients, often with a history of exposure to other harmful factors like indoor open fire with coal or coke, or organic dusts in the workplace (16). It is also believed that air pollution, noxious fumes and vapours, and secondhand smoke are risk factors for COPD (15). The mechanisms explaining why some develop COPD is not fully understood, but the disease is associated with chronic inflammation in the lung caused by risk factor exposure (17). However, exposure per se is not sufficient to develop the disease. Disease development depends on complex interactions between the host and the environment, for instance type and amount of exposure, the host's susceptibility, and host responses (17). The chronic inflammation leads to a destruction of the lung parenchyma and narrowing of the small airways, eventually resulting in a progressive airflow limitation (18). As the disease progresses, inflammation will increase,

contributing to more airway damage, remodelling (structural/histological changes), loss of small airways, and emphysema (overdistended alveoli) (17-21). Peripheral airflow limitation and destruction of alveolar attachment will result in airway closure or collapse, thereby trapping air during expiration and create hyperinflation (18, 22). Hyperinflation is suggested to correlate with exertional dyspnoea (23, 24). Inflammation is the body's normal reaction to an injury, or in this case inhaled irritants, but the inflammation observed in the respiratory tract of COPD patients appears, for reasons not yet understood, to be an amplification of the normal inflammatory response (18, 22). The inflammatory response in COPD involves both the innate and the adaptive immune system, and macrophages, neutrophils, T- and B lymphocytes, and sometimes eosinophils are found in increased numbers in the respiratory tract (17, 18, 20). Interestingly, the most common inflammatory phenotype is a neutrophilic inflammation (17), an inflammatory pattern resembling that observed during a bacterial infection.

6.1.1 COPD and bacteria

As early as 1959, researchers suggested that recurrent bronchial infections were the reason that some smokers developed progressive airways obstruction and others did not (25). This was called "the British hypothesis", which emphasised a clear distinction between COPD and the related disease asthma. At the same time, professor Orie presented what was later called "the Dutch hypothesis", saying that COPD and asthma rather had common origins and clinical expressions, and were determined by host factors (e.g. age, sex) and environmental factors (e.g. allergens, infections, smoking, air pollution) (26). The hypotheses have been subject for debate, with some researchers supporting the British (27), some supporting the Dutch (28), and some arguing for both (29). The British hypothesis was met with some criticism, but regained enthusiasm when Sethi et al. showed that acute exacerbations of COPD were commonly associated with the emergence of new bacterial strains (30). Later studies reported that recurrent respiratory infections were associated with accelerated loss of lung function in COPD patients, and

thereby supporting the British hypothesis further (31-33). It was proposed that the exacerbations were associated with increased airway inflammation (4), which in turn led to the FEV_1 decline. The notion of infections as a driver of lung function decline made bacteria more likely to be the antigen associated with the inflammatory immune response observed in COPD (34, 35).

Exacerbations are a common feature of COPD, and it has been reported that 31% of a COPD cohort in the ECLIPSE study experienced exacerbations requiring hospital admission during a 3-year follow-up (36). It is suggested that 50-70% of exacerbations are caused by airway infections, in which the causative agent might be bacteria, atypical organisms or respiratory viruses (37). The exact proportion of viral infections as the cause of COPD exacerbations is not known, and depends somewhat on the identification techniques. Still, studies have shown that viruses are a common cause of COPD exacerbations (38-40). A study using a polymerase chain reaction (PCR) based method has detected virus in as much as 64% of the exacerbations in a COPD population (41). Presence of a potential pathogenic bacteria have been found during exacerbation in approximately 30 and 50% in sputum cultures and bronchial secretion cultures, respectively (37). The bacterial association to exacerbations was thought not only to be infective, but it seemed reasonable to expect that patients with chronic colonisation would have a greater inflammatory burden, and more frequent exacerbations (5, 37).

6.1.2 Healthy lungs and bacteria

Despite growing evidence of chronic bacterial colonisation in the lungs of adults with COPD, the healthy lungs have long been considered sterile (6). This assumption is somewhat surprising considering the lower airways' direct communication with surrounding air, and the constantly exposure of microorganisms from the upper respiratory tract including microaspiration (42). The idea of lung sterility has its origin in data published more than 100 years ago due to minimal bacterial growth from excised

nasal and tracheal mucosa of rabbits (8). The hypothesis was supported despite observations of bacteria cultures from lung specimens of several different animals as early as 1922 (43), and the possibility of microbial transmissions to the lungs through microaspirations (44). One of the reasons that the sterility hypothesis survived, is related to the methods used by laboratories to detect microorganisms, mainly culture-based methods. Different growth media were optimised to allow culturing of bacteria from the lung specimens. Methods were mainly aimed at known pathogenic organisms (45), and would probably miss important bacteria not unlikely to be present in the lungs. Indeed, it has been shown that only a fraction of bacteria is cultivable using standard medical microbiology media (46), and in addition, there is always the possibility of infection or colonisation with hitherto unknown agents. The introduction of culture-independent techniques for identification of microbes was a turning point in the view of lung sterility.

6.2 The microbiome

Culture-independent molecular methods for identification of microbes do not require that bacteria should be cultivable, i.e. grown in a laboratory, and has the potential to reveal a higher diversity of microbial organisms compared to conventional cultures. This allowed for the identification of the "microbiota", defined as the microbial community membership associated with a defined habitat, for instance different body parts (47). The genetic information and associated physico-chemical properties of the gene products of a microbiota is termed the "microbiome" (47).

As a consequence of the introduction of culture-independent technology discussed above, the interest in the human microbiome raised in the first decade of the twenty-first century (8). In 2012, it was proposed that relationships existed between the microbiome and several human diseases (48), and large-scale studies were initiated to investigate the human microbiome in detail. For instance, the Human Microbiome Project (HMP) by the US National Institutes of Health (49) and the European MetaHIT (50). Studies reported

on substantial variability in microbial composition among human populations, lifespan, disease, and events like antimicrobial exposure (48). Studies also indicated associations between various diseases or conditions and body site specific microbiomes. In particular, association have been found between the gut microbiome and obesity (51, 52), colorectal cancer (53, 54), and inflammatory bowel disease (IBD) (55). It became clear that each body site harboured a microbiota with unique characteristics (48). However, the study of the human microbiome is a young research area, and associated analyses lack standardisation and pose several challenges and pitfalls.

6.2.1 Analysis of the microbiome

Human microbiome studies usually follow a common study design. Studies are initiated with sampling of the microbiome by collection of appropriate samples from the body site of interest, for instance bronchoalveolar lavage (BAL) samples from the lower respiratory tract or faecal samples from the gut. The total DNA content should then be extracted from the cells in each sample, performed by cell lysis using mechanical-, chemical-, or enzymatic methods (56). Extracted DNA is subsequently sequenced in order to determine the order of nucleotides in the DNA. As discussed above, most microbiome studies utilise culture-independent sequencing techniques due to its improved identification of microbes compared to conventional cultures. When BAL samples from lung transplant recipients were analysed using both culture-independent techniques and a routinely used culture-based technique, bacteria were found in significantly more samples using the culture-independent approach (45). Targeted amplicon sequencing is a common culture-independent sequencing technique well-established to identify microbes from complex environments (47), based upon amplification of so-called marker genes. Marker genes are highly conserved and thus shared among the microbes of interest. At the same time, these marker genes exhibit enough variation to allow for taxonomic identification, a term to describe classification of biological organisms based on their characteristics (57). Examples of commonly used marker genes are the 16S ribosomal RNA (16S rRNA) for

bacterial studies (56), and the 18S small subunit ribosomal DNA (rDNA) (58) or either of two fungal rDNA internal transcribed spacer regions (ITS1 or ITS2) for fungal studies (59). Marker genes are PCR amplified to make multiple DNA amplicons, which serve as input to a sequencing platform. Until the early 2000s, mass sequencing of marker gene amplicons was a cumbersome, expensive process. But the so-called next-generation sequencing (NGS) technologies allowed high-throughput sequencing processes from platforms like Illumina® (e.g., MiSeq or HiSeq; Illumina Corporation, San Diego, CA), that could identify millions of DNA reads from collected samples. In short, NGS is achieved by parallel sequencing to facilitate high-throughput (60). Sequencing could be done at both ends of the DNA fragment, so-called paired-end sequencing, and several samples are sequenced simultaneously by appending a unique sample identifier to one or both ends of the amplicons, a process called multiplexing.

Sequencing results in a large number of short sequences, which then are transferred back to the researcher in form of data files. Such data files often include several thousand DNA reads, and consequently, a microbiome project might end up with millions of raw DNA sequences. Sequencing data needs to be processed using bioinformatic tools. Many tools and pipelines have been developed, for instance quantitative insights into microbial ecology (*QIIME*) (61) and *mothur* (62). Additionally, the free software *R Project for Statistical Computing* could be used for further statistical analyses (63). Papers usually report the composition or taxonomic information, the relative abundance of species, and diversity analyses. The latter includes alpha diversity, which describes the number and distribution of organisms within a sample, and beta diversity, which describes diversity between pairs of samples (47). Taxonomic assignments are achieved by comparing sequences to existing databases such as SILVA for 16s rRNA data (64) or UNITE for fungal ITS data (65, 66). Finally, it is common to integrate results with clinical metadata (46), for instance looking at differences among participant groups, or predict outcomes using species data.

Several of the steps in a common microbiome study possess challenges. Contamination is a concern in every microbiome study (47), and could happen through sampling directly (e.g. oral contamination in sputum samples from the lungs) or sample preparation (e.g. contaminants in DNA extraction kits and laboratory reagents) (67). Despite being user-friendly, several parameters of the analytical tools and pipelines could impact the final result (68). Furthermore, different sequencing platforms have been shown to produce different results, and variability can be introduced by differences in DNA extraction, PCR amplification, and patient diversity (68). Databases used for taxonomic assignment contain unresolved information for some sequences, excluding identification on species level for some microorganisms (47). Sampling of adequate material is unproblematic in many microbiome studies, such as faecal sampling in gut studies, but could be an issue in other studies. Sampling the lung microbiome is difficult (69). Numerous studies have sampled the lung microbiome by bronchoscopy, including studies of healthy subjects (42, 70-74), participants with COPD (75-79), and participants with asthma (7), but sputum sampling is also frequently used (80-82). The bronchoscopy method is prone to contamination when passed through the upper respiratory tract. Protected sampling minimises the problem, and should be used (9). However, including bronchoscopy in the study design could lower participation in lung microbiome studies.

A semi-invasive procedure like bronchoscopy can be associated with discomfort and pre-procedural anxiety, which might lead possible participants to decline an invitation, or withdraw their consent. Studies on the lung microbiome are not necessarily therapeutic or beneficial to the participants, perhaps lowering participation even more. Little is known of motivation and predictors of participation in studies involving a research bronchoscopy. A systematic review of participation in colorectal screening trials has reported higher participation rates with general practitioner involvement and face-to-face invitation, and lower probability of accepting an invitation if a long travel distance was involved (83). The response rates from seven Norwegian respiratory healthy surveys from 1965 to 1999 have also been examined, and results show that response rates were

higher in women than in men and higher in the middle-aged/elderly than in young adults (84). A bronchoscopy was however not included in the Norwegian studies. Another Norwegian study on respiratory health showed that non-response was associated with male sex, younger age, and living in a rural area (85). Additionally, they found a weak trend towards more manual occupations among non-responders compared to responders (85). Table 1 summarises the literature on participation in research bronchoscopy studies as of February 2014 (reproduced with editing of a minor correction (see **14 Minor errata**) from paper I published in European Clinical Respiratory Journal).

Table 1. Papers from a literature search and review between December 2013 and February 2014 on participation in research bronchoscopy studies

Author, year of publication, country, reference	Study objective	Main outcomes	N	Population	Inclusion criteria	Design	Purpose of bronchoscopy	Bronchoscopy procedure	Relevant findings
Lipman, 1998, United Kingdom and USA, (86)	Examine acceptability of research bronchoscopy in asymptomatic HIV subjects	Pre- and post-bronchoscopic participation motives, willingness to second procedure. Response rates	Responders: 75 Nonresponders: 23	HIV-infected individuals, 18 - 60 yrs	No acute/chronic respiratory disease, IV drug use or antiretroviral therapy	Prospective. Semistructured, self-completed questionnaires	Harvest cell populations from the lung	Bronchoscopy with BAL	Response rate: 70%. Participation motives: personal benefits, altruism. Non-responders younger. 79 % would accept second research bronchoscopy
Mtunthama, 2008, Malawi and United Kingdom, (87)	Examine adequacy of study information, motives for participation and complication rate in a research bronchoscopy study	Participation motives, perceived sufficiency of study information, self-reported adverse events	100	Malawian volunteers	None given	Prospective. Interviews with both open and closed questions	Determine factors underlying susceptibility to respiratory infection among adults	Bronchoscopy with BAL	Response rate: Not given. Participation motives: personal benefit, study enthusiasm. All agreed to a second bronchoscopy. 94 % considered pre-information adequate and useful
Kerrison, 2008, United Kingdom, (88)	Examine patient's experience of clinical research	Participation motives, importance of referral and consent procedures, sufficiency of study information	18 in bronchoscopy study	6 clinical studies	None given	Retrospective qualitative study. Telephone interviews, questionnaires or focus groups	Identify and sample suspicious lesions	Fluorescence bronchoscopy and optical biopsy	Response rate: 36%. Participation benefits: Extra care, increased surveillance, see an expert and help researchers and others. Most encountered studies through the care giving hospital
Kye, 2009, USA, (89)	Examine various recruitment strategies	Response rates and cost calculations	137	Healthy ex-smokers, ≥30 pky	No contraindications for bronchoscopy or celecoxib, no use of certain medications or no cancer	Prospective study. Telephone interview	Examination and sampling	White light and fluorescence bronchoscopy with BAL and bronchial biopsies	Response rate: 3,1%. Radio advertisement generated most inquiries followed by internet, print media, posted flyers and mass mailing
Schook, 2010, The Netherlands, (90)	Examine whether participation in a chemoprevention study could influence smoking cessation	Smoking cessation rate and influence of participation on cessation	146	Healthy current or former smokers, ≥20 pky	No serious comorbid disease, FEV_1 below 1000 mL or use of systemic or inhaled corticosteroids in previous year	Retrospective study. Questionnaire-based telephone interviews	Examination	Autofluorescence- or white light bronchoscopy with biopsies	Response rate: 73%. Participation motives: Personal benefit
Patel, 2012, United Kingdom, (91)	Examine methods and participation in a lung cancer screening study	Participation motives, declining reasons and views on screening methods	60	Current or ex-smokers, ≥20 pky and or ≥20 yrs, with mild to moderate COPD	No serious comorbid disease, life expectancy ≥5 yrs	Prospective qualitative study. Semi-structured interviews	Examination	Fluorescence bronchoscopy with biopsies	Response rate: Not given. Participation motives: personal benefit, altruism. Participation barriers: fear, bad experiences and travel
Chudleigh, 2013, United Kingdom, (92)	Examine recruitment and retention of CF infants and healthy controls, and parental attitude to participation	Participation motives, benefits and disadvantages. Response rates	Cases: 85 Volunteers: 56	Infants with and without CF, including parents	CF: No contraindicated disorders, no preterms. Healthy controls: No medical and/or social contraindications, ≥2500 g	Prospective longitudinal, observational study. Self-completed questionnaires	Examination	Bronchoscopy with BAL	Response rate: 69% (CF), 21 % (healthy controls). Participation motives: Personal benefit, altruism

BAL, bronchoalveolar lavage; COPD, chronic obstructive pulmonary disease; CF, cystic fibrosis; FEV_1, forced expiratory volume in one second; pky, pack-years.

An important task for researchers is to obtain an overview of the published literature in the research field of interest. Systematic reviews are summaries of the available papers in a given research area, conducted with comprehensive literature searches and explicit, reproducible criteria for inclusion and exclusion (93). Researchers conducting a systematic review need to define a robust research question, and include usable terms in their literature search. The PICO framework, a mnemonic for Patient/problem/population, Intervention, Comparison, and Outcome, is often used (94). A systematic review could be a suitable first step to dig into a novel research area, such as the use of bronchoscopy in research studies. More information on response rates, participation motives, and non-response reasons could lead to better-targeted recruitment for lung microbiome studies.

6.2.2 The lung microbiome

Most of the early microbiome studies, including the leading HMP, did not include samples from the lungs, but data from lung microbiome studies has increased from 2008 (46). The first study of the healthy lung microbiome using culture-independent techniques was performed in 2010 (7). Patients with COPD or asthma were also included, and the study reported on distinct microbiomes from the healthy and diseased lungs (7). The result challenged the long-standing notion of a sterile lung community. Nevertheless, most attention was given to diseased lungs, and especially to patients with cystic fibrosis (95-97). The microbiome of patients with asthma (7, 98) or interstitial lung diseases (99), lung transplant recipients (100), and patients receiving mechanical ventilation (101) was also investigated. Furthermore, an increasing number of studies also reported a rich lung microbiome in clinically stable COPD.

6.2.3 The microbiome and COPD

A summary of published papers on the COPD microbiome is presented in Table 2. Studies have shown that the lung microbiome found in COPD patients seem to differ from that found in healthy controls (7, 76, 78, 102, 103). The COPD microbiome studies

reported that *Proteobacteria*, *Bacteroidetes*, *Actinobacteria* and *Firmicutes* were commonly found phyla, while *Pseudomonas*, *Streptococcus*, *Prevotella*, *Moraxella*, *Acinetobacter*, *Fusobacterium*, *Neisseria*, and *Haemophilus* were common genera (7, 75-78, 104, 105). It has also been suggested that patients with more advanced disease stages showed lower alpha diversity, i.e. lower number and/or more similar bacteria (106). Study participants exposed to inhaled bronchodilators and/or ICS harboured microbiomes that clustered together (78). Associations between the lung microbiome and COPD exacerbations have also been examined (80, 81, 101, 103, 105, 107, 108). One study suggested that rhinovirus infection resulted in increased bacterial burden especially due to outgrowth of *Haemophilus influenzae* with additionally increases in sputum inflammatory cells, neutrophils, and neutrophils elastase levels (80). Other studies have found that some genera changed in relative abundance during exacerbations (81, 103), which, together with other changes in the microbiome, were also associated with inflammation (81).

Table 2. An overview of COPD lung microbiome and mycobiome studies using targeted amplicon sequencing as of 16th of September 2017

Author, year of publication, country, reference	Study objective	Main outcomes	N	Population	Inclusion criteria	Design	Handling of contamination	Sampling method	Relevant findings
					Bacterial microbiome				
Hilty, 2010, England, (7)	Compare airway microbiota in adult patients with asthma, COPD, and controls, plus asthmatic children and controls	Taxonomy, diversity, and cladistic analysis.	Asthma: 11 COPD: 5 Control: 8 Paediatrics: 20	Adult subjects with and without obstructive diseases. Paediatric patients with difficult asthma and non-asthmatic controls	No antibiotics or clinical infections at study time	Observational study	Not mentioned	Swabs from nose and OP and bronchoscopic cytology brushings (LUL). BAL in paediatric participants.	*Prevotella*, *Streptococcus*, *Staphylococcus*, *Neisseria*, *Corynebacterium*, and *Haemophilus* spp. were abundant. Nasal microbiota differed from OP and LUL. Cases: *Proteobacteria*, controls: *Bacteroidetes*
Huang, 2010, USA, (101)	Profile the microbiome in COPD patients with exacerbation	Taxonomy, diversity, abundance	COPD: 8	Intubated COPD patients with exacerbation-related respiratory failure treated with antibiotics	Subjects admitted to the ICU with a primary diagnosis of "COPD exacerbation"	Observational study	Use of endotracheal tubes	Endotracheal aspirates and quantitative clinical laboratory bacterial cultures on m-BAL	A core community of 75 taxa was detected in all patients. They saw interpersonal variation in bacterial richness, and two distinct groups on NMDS influenced by duration of intubation
Erb-Downward, 2011, USA, (77)	Compare the lung microbiome in "healthy" smokers and non-smokers. Analyse multiple sample sites from surgical explants	Taxonomy, abundance, and diversity	COPD: 10 Smokers: 7 Non-smokers: 3	Three never-smokers with normal spirometry, seven smokers with normal spirometry, and four subjects with COPD	No mental incompetence or active psychiatric illness, primary asthma diagnosis, CF, bronchiectasis, infl or fibrotic lung disease, prednisone >20 mg daily	Observational study	Bronchoscopy through mouth	BAL and transplantation tissue	Data suggests the existence of a core pulmonary bacterial microbiome. There were significant micro-anatomic differences in bacterial communities within the same lung of subjects with advanced COPD. Less diversity in moderate and severe COPD
Cabrera-Rubio, 2012, Spain, (75)	Examine the COPD microbiome, and to compare diversity from upper and lower bronchial samples	Taxonomy and diversity	COPD: 8	COPD patients not showing signs or symptoms of infection	Moderate disease. No exacerbations, antibiotics, or hospitalisation prvs year. No severe lung function or regular treatment for COPD or other severe disease	Observational study	Transnasally bronchoscopy without aspiration before vocal cords	Sputum, bronchial aspirate, BAL, and bronchial mucosa	Sputum samples showed significantly lower diversity than the other three sample types. Suggests BAL is most suitable for lower respiratory tract sampling
Sze, 2012, Canada, (102)	Characterise the bacterial community	Abundance, diversity, and taxonomy	Non-smokers: 8 Smokers: 8 COPD: 8 CF: 8	Non-smokers and smokers, very severe COPD (GOLD 4) and CF		Observational study	Negative controls used for OTU subtraction	Tissue	More *Firmicutes* in COPD. Similar diversity between non-smokers, smokers, and COPD, and higher than CF. COPD clustered alone in PCoA
Pragman, 2012, USA, (78)	Compare lung microbiomes in COPD and controls	Taxonomy, diversity, and abundance	COPD: 22 Controls: 10	Controls, moderate COPD, and severe COPD	COPD patients without a recent exacerbation. No smoke or systemic steroid in past 6 months or antibiotic use in the past 2 months	Observational study	States that standard clinical protocols were followed	BAL	Significant increase in microbial diversity with the development of COPD. COPD contained mostly *Actinobacteria*, *Proteobacteria*, and *Firmicutes*. PCoA revealed clustering of control and COPD samples, and on ICS or IBD use. Suggests oral influence on lung microbiome content
Zakharkina, 2013, Germany (79)	Examine the pulmonary microbial communities in COPD patients as compared to healthy individuals	Taxonomy and diversity	COPD: 9 Healthy: 9	Healthy never smokers and participants with clinically diagnosed COPD	Healthy: Never smokers, no chronic disorders, no respiratory illnesses in the prvs year. COPD: Clinically diagnosed COPD	Observational study	Included positive and negative controls in the PCR	BAL	Results of the terminal restriction fragment analysis correlated partly with the data obtained from clone sequencing. The two different methods revealed partly diverse microbial spectra
Molyneaux, 2013, United Kingdom, (80)	To investigate the effect of rhinovirus infection on the airway bacterial microbiome	Abundance and diversity	COPD: 14 Controls: 17	Mild COPD and controls without any obstructive airway disease (10 non-smokers and 7 smokers)	No asthma or atopy, no other systemic or respiratory conditions, no history of respiratory tract infection, exacerbation, or antibiotic use in the prvs 3 months	Observational study	States they avoided invasive sampling with associated contamination	Induced sputum	Increased bacterial burden, especially *Haemophilus influenzae*, after rhinovirus infection in COPD, but not controls

Author, year of publication, country, reference	Study objective	Main outcomes	N	Population	Inclusion criteria	Design	Handling of contamination	Sampling method	Relevant findings
Huang, 2014, USA, (81)	Examine dynamics of the microbiome during COPD exacerbation	Abundance and diversity	COPD: 12	COPD patients longitudinally followed to detect exacerbations.	No other lung disease	Observational study	Mentions saliva contamination, but argues for sputum	Expectorated sputum	Observed compositional changes at exacerbation onset compared to pre-state for several taxa, and exacerbation compared to post-period. Antibiotics and steroid treatment had effect on composition
Millares, 2014, Spain, (103)	Determine characteristics of the microbiome of severe COPD patients with and without *Pseudomonas aeruginosa* colonisation, and its changes during exacerbation	Abundance and diversity	COPD: 16	Severe COPD in stable state with more than 2 exacerbations/year with and without *Pseudomonas aeruginosa* colonisation	Age > 40 yrs, no respiratory hospital admission prvs 6 months, no asthma, CF, neoplasia or bronchiectasis, no oral steroids or immunosupp, no change in smoking or bronchodil in follow-up	Observational study	Mentions risk of oral and oropharynx contamination	Spontaneous sputum	Observed abundance changes in exacerbations. Differences between colonised and non-colonised disappeared during exacerbation. No diversity difference between colonised/non-colonised
Galiana, 2014, Spain, (109)	Compare the microbiota in two groups of patients with COPD of different severity (mild/moderate vs severe)	Abundance and diversity	COPD: 19	9 patients with mild or moderate COPD and 10 patients with severe or very severe COPD in a stable condition	At least 3 months without exacerbation or use of antibiotics for any other reason	Observational study	Not mentioned	Expectorated sputum	Moderate disease had higher diversity. Bacterial load was higher in severe patients. Presence/absence of *Actinomyces* associated with disease severity. Moderate disease seemed to have a core, not seen in severe
Garcia-Nuñez, 2014, Spain, (106)	To identify severity-related modifications of the bronchial microbiome in COPD	Abundance and diversity	COPD: 17	Stable COPD (> 4 weeks) with moderate/severe or advanced disease	Age > 40 yrs, no respiratory hospital admission prvs 6 months, no asthma, CF, neoplasia or bronchiectasis, no oral steroids or immunosupp	Observational study	Mentions oropharyngeal contamination, but argues for sputum	Sputum	Higher alpha diversity in moderate/severe, paralleled by a substitution of the original flora by colonizing bacteria only marginally present in the bronchial microbiome of COPD patients with less severe function
Millares, 2015, Spain, (107)	Analyse the composition and the gene content of the COPD microbiome in stability and exacerbation	Abundance, diversity, and metabolic capabilities	COPD: 8	Severe COPD (FEV_1 < 50% of ref) with ≥3 exacerbations in prvs year	No asthma, CF, neoplasia or bronchiectasis, no immunosupp, no antibiotics between sampling. Included if exacerbation sample and stable 1-6 months before were given	Observational study	PCR controls included to assess contaminants in extraction buffer. Extraction done in cabinet	Sputum	No difference in relative abundance or diversity between stability and exacerbation. Functional differences were identified during exacerbations in certain pathways in MG-RAST, but not significant in PICRUSt
Sze, 2015, Canada and USA, (110)	Examine the relationship of the COPD microbiome to emphysematous tissue destruction, number of terminal bronchioles, infiltrating inflammatory cells, and host gene expression	Diversity, abundance, and host response/relation ships between OTUs and host factors	COPD: 5 Controls: 4	Very severe COPD (GOLD 4). Organ donors were controls if lungs were unsuitable for transplantation		Observational study	Included a negative water control sample	Tissue	Decline in microbial diversity that was associated with structural changes, and the CD4+ T cell infiltration. Specific OTUs were also associated with neutrophils, eosinophils, and B-cell infiltration. The expression profiles of 859 and 235 genes were associated with either more or less of *Firmicutes* and *Proteobacteria*, respectively
Aguirre, 2015, Spain, (104)	Describe bacteria present in expectorated sputum using a pyrosequencing approach and to compare these findings with those obtained with conventional culture	Taxonomy and distribution	COPD: 19	Stable COPD. Nine patients with moderate COPD and ten patients with severe COPD	Min 3 months without exacerbation or use of antibiotics. No current or recurrent symptomatic IHD, congestive heart disease, cerebrovascular disease, dementia, lung cancer, known psychiatric illness, systemic steroids use (oral or parenteral), active TB, IBS or insulin-dependent DM	Observational study	Not mentioned	Expectorated sputum	More genera with pyrosequencing. In pyrosequencing, best represented phyla were *Firmicutes*, *Proteobacteria*, *Actinobacteria*, *Bacteroidetes*, and *Fusobacteria*

Author, year of publication, country, reference	Study objective	Main outcomes	N	Population	Inclusion criteria	Design	Handling of contamination	Sampling method	Relevant findings
Su, 2015, China, (108)	Examine the structure and varieties of lung microbial community in patients with severe AECOPD (multiple sampling times)	Diversity, taxonomy, and abundance	COPD: 6	Patients experiencing an acute exacerbation of severe COPD and receiving medical treatment		Observational study	Not mentioned	Sputum	Bacterial analysis: Longitudinal sampling showed differences in taxonomy over time. No consistent pattern in alpha diversity. Samples clustered on CRP value in PCoA
Wang, 2016, USA and United Kingdom, (105)	Examine the lung microbiome at stable state, exacerbation, 2 weeks post-therapy, and 6 weeks recovery	Taxonomy, abundance, diversity, and human-host interactions	COPD: 87	COPD patients	No asthma, or significant respiratory disease other than COPD. Must be able to produce sputum after sputum induction	Observational study	Negative reagent controls. Compared with previous contamination study	Sputum	Longitudinal sampling showed that changes appeared to be associated with exacerbation events and indicative of specific exacerbation phenotypes. Antibiotic and steroid treatments appear to have differential effects on the lung microbiome
Einarsson, UK, 2016, (76)	Determine airway microbial composition using both culture-dependent and culture-independent methodologies	Taxonomy, diversity, correlation, and abundance	COPD: 18 Smokers: 8 Non-smokers: 11	COPD patients in stable state. Healthy subjects recruited from the general population	COPD: Mild to severe airflow obstruction, no bronchiectasis, interstitial lung disease, or lung carcinoma	Observational study	Negative controls (PBS)	Bronchial wash	Microbial community composition in the lower airways of patients with COPD was significantly different to that found in smokers and non-smokers
Segal, 2017, USA, (111)	Test whether AZM treatment affects the lung microbiome and bacterial metabolites	Diversity, abundance, metabolomics, and inflammation	Smokers with emphysema: 20 (6 met GOLD COPD criteria)	Subjects with CT-defined emphysema with significant smoking history	No treatment with antibiotics or steroids prvs. 3 months, FEV1>70% pred., age>70, no known cardiovascular, liver, or renal disease, no inhaled med. for at least 1 month	Observational study	Supraglottic sample, no suction before BAL, background samples	Acellular BAL	AZM treatment altered both lung microbiota and metabolome, affecting anti-inflammatory bacterial metabolites that may contribute to its therapeutic effects
Engel, 2017, several European countries, (112)	Examine whether COPD sub-types, defined by quantitative CT, are associated with lung microbiome changes	Abundance, diversity, and correlation	COPD: 16 Healthy: 9	COPD patients with CT defined subtypes and controls from 4 European countries	Samples: >2 µg of total DNA in combination with high percentage airway wall area or low lung density in CT for cases	Observational study	Accounted for possible PCR contamination. Used protected sampling	Brushes	CT detectable changes in the COPD lung are associated with alterations in bacterial communities, which may induce further changes in the microbe-host interaction
				Fungal microbiome					
Cui, 2015, USA and China, (113)	Characterise the mycobiome at different respiratory tract levels in persons with and without HIV infection and in HIV-infected individuals with COPD	Diversity, abundance, and association with disease group and clinical risk factors	HIV uninf: 24 HIV COPD+: 22 HIV COPD-: 10	Outpatients. HIV uninfected, and HIV infected either with or without COPD	No acute change in respiratory symptoms in the prior 4 weeks and no systemic antimicrobials or immunosuppressive medications within 6 months	Observational study	Bronchoscopic control samples	OW, IS, and BAL	Mycobiomes of OW, IS, and BAL shared common organisms, but each also had distinct members. *Candida* was dominant in OW and IS, but BAL had 39 fungal species that were disproportionately more abundant than in the OW. Fungal communities in BAL differed significantly by COPD with *Pneumocystis jirovecii* significantly overrepresented in the COPD group

Author, year of publication, country, reference	Study objective	Main outcomes	N	Population	Inclusion criteria	Design	Handling of contamination	Sampling method	Relevant findings
Su, 2015, China, (108)	Examine the structure and varieties of lung microbial community in patients with severe AECOPD (multiple sampling times)	Diversity, taxonomy, and abundance	COPD: 6	Patients experiencing an acute exacerbation of severe COPD and receiving medical treatment		Observational study	Not mentioned	Sputum	Fungal analysis: The fungal population was typically dominated by *Candida*, *Phialosimplex*, *Aspergillus*, *Penicillium*, *Cladosporium*, and *Eutypella*. Each patient had an unstable microbial community during the observation period. No consistent pattern in alpha diversity

COPD, chronic obstructive pulmonary disease; N, number of participants; OP, oropharynx; LUL; left upper lobe; BAL, bronchoalveolar lavage; ICU, intensive care unit; m-BAL, minibronchoalveolar lavage; NMDS, Nonmetric multidimensional scaling; CF, cystic fibrosis; infl., inflammatory; prvs., previous; OTU, operational taxonomic unit; PCoA, principal coordinates analysis; ICS, inhaled corticosteroids; IBD, inhaled bronchodilators; PCR, polymerase chain reaction; yrs, years; immunosupp, immunosuppressive drugs; bronchodil, bronchodilator treatment; FEV_1, forced expiratory volume in one second; ref, reference; MG-RAST, Metagenomics RAST server; PICRUSt, Phylogenetic Investigation of Communities by Reconstruction of Unobserved States; min, minimum; IHD, ischaemic heart disease; TB, tuberculosis; IBS, inflammatory bowel syndrome; DM, diabetes mellitus; AECOPD, Acute exacerbation of COPD; CRP, C-reactive protein; PBS, phosphate buffered saline; AZM, Azithromycin; CT, computed tomography; uninf, uninfected; OW, oral wash; IS, induced sputum.

There is growing evidence that the microbiome plays a role in COPD, but it still remains a problem that most of the studies have mainly focused on the bacteria present in the lungs (47). Because lungs also contain viruses, fungi, and other eukaryotes.

6.3 The mycobiome

The revolution in culture-independent techniques greatly expanded our understanding of the bacterial microbiome, but other important community members like fungi were largely ignored (114). In fact, the word “microbiome” refers to only inhabiting bacteria (115). The term “mycobiome” was introduced by a paper in 2010 as a combination of the words “mycology”, i.e. study of fungi, and “microbiome” (116). The mycobiota could be defined as the fungal microorganisms in an anatomically defined niche, whereas the mycobiome is the collective genetic material of the mycobiota.

Some important differences between a bacterial microbiome study and a mycobiome study are worth mentioning. The total number of fungal cells found on body surfaces are usually lower than bacterial cells (59). However, the fungal cell is typically a hundred-fold larger than bacterial cells (59). Fungal cells are eukaryotic in contrast to the prokaryotic bacterial cells, and mycobiome studies using 18S rDNA for amplification runs the risk of amplifying DNA from other eukaryotes, for instance human DNA (69). Fungi are encased in thick cell walls, and methodology originally developed for isolating bacterial genomic DNA is not necessarily ideal for recovery of fungal DNA (59). ITS is the recommended marker-gene region for fungal studies (117), but the ITS region has a length variation, and ITS fragments from mice and human faeces have been shown to vary in length between 100-550 base pairs (118). The fungal databases are also incomplete, and suffers from a dual naming system depending on the fungus’ reproductive stage (69).

Fungi might be important to maintain microbial community structure, metabolic function, and immunological responses in a host (119). As with the bacterial microbiome, unculturable fungi comprise the largest part of the human mycobiome, and early mycobiome studies were based on a few selected culturable fungi like *Candida albicans*, *Candida glabrata*, *Cryptococcus neoformans*, *Aspergillus fumigatus* and the dimorphic fungi (*Coccidioides*, *Histoplasma* and *Blastomyces*) (58). The introduction of culture-independent techniques showed that the various niches harboured a considerable mycobiome (58). Evidence emerged that specific fungal populations were found at different body sites like the skin, the gastrointestinal tract, and the genitourinary system (58, 59). The most commonly found fungus on the skin is called *Malassezia*, and pathogenic fungi like *C. albicans* has been found as part of the mycobiome in asymptomatic healthy individuals (59). Fungal genera such as *Candida*, *Saccharomyces*, and *Cladosporium* are also found throughout the gastrointestinal tract and in the genitourinary system (59). New evidence suggests that interactions between mycobiomes found at different body sites are possible, especially a gastrointestinal-respiratory interaction. For instance, an animal model has shown that a disrupted gut microbiome due to yeast overgrowth subsequently predispose the host to allergic airway disease after allergen challenges (120). Another study has suggested that fungal cell wall components might translocate into the bloodstream of HIV-infected individuals without known active fungal infection, possibly stimulating a systemic immune response and inflammation (121). Furthermore, fungal-bacterial interactions seem to promote or inhibit growth, eventually leading to altered immune responses or antimicrobial therapy (58). Another important interaction is the one between the mycobiome and the host. It is suggested that host factors like genotype, lifestyle choices, and physiologic and immunologic factors all play a role in shaping the mycobiome (58). On the other hand, fungi are capable of elicit immune responses in humans. Fungi contain pathogen-associated molecular patterns (PAMPs) such as glucans, chitin, and mannans present in the fungal cell wall (69). Pattern-recognition receptors (PRR) on phagocytes recognise these PAMPs and initiate an immune response, illustrating how the interaction between fungi and the immune

system is bidirectional. Some fungi are known to have pathogenic traits (122), and, taking the beforementioned links between the microbiome and several diseases into account (see **6.2 The microbiome**), associations between the human mycobiome and various pathologies do not seem unrealistic.

There has been a growing literature that the mycobiome indeed is associated with a number of diseases. *Malassezia spp.* have been shown to play a role in several skin diseases. The skin is naturally inhabited by *Malassezia*, but the fungus has pathogenic properties. Pityriasis versicolor is the only skin disease known to be caused by *Malassezia*, but *Malassezia* is suggested to contribute, albeit not directly, in seborrheic dermatitis, atopic dermatitis, and psoriasis as well (123). Fungi are considered to affect healing of chronic skin wounds. A study suggested that the baseline mycobiome from wounds could be predictive of healing time (124). Stool samples are a feasible and easy way to investigate the intestinal mycobiome, and have been utilised in studies on IBD. Disease-specific alterations in fungal diversity, and reciprocal changes in phyla in IBD patients especially during IBD flare has been reported in a large study, suggesting considerable shifts in the mycobiome (125). IBD patients also had reduced levels of *Saccharomyces cerevisiae*, a pattern also shown during active flare (125). Furthermore, the intestinal mycobiota is suggested to play a part in alcohol-induced liver disease in humans and mice (126). The researchers observed that alcohol-dependent patients displayed reduced intestinal fungal diversity and *Candida* overgrowth compared with healthy individuals. Additionally, an increased systemic exposure and immune response to fungal products were seen in patients with alcohol abuse. Growing focus has also been given on the lung mycobiome, and papers on various lung diseases have been published (95, 100, 108, 113, 127-134).

6.3.1 The lung mycobiome

A thorough scientific focus should undoubtedly be put on the lung mycobiome for several reasons. First, fungi are ubiquitous, and the human respiratory tract are exposed for numerous airborne fungi. Indeed, fungal spores represent more than 50,000 spores per cubic meter of air during the fungal season (135). Healthy airways possess effective removal of such spores through mucociliary clearance and phagocytosis. Second, fungal infections have increased in previous years, especially in immunocompromised individuals (58). New fungal pathogens are being discovered, and an extensive use of antibiotics may have promoted resistant respiratory fungal infections (135). Third, ICS are frequently used by individuals with chronic lung diseases. Use of ICS will have anti-inflammatory and immunosuppressive effects, possibly resulting in outgrowth of commensal fungi such as *Candida* (136). It would thus be interesting to examine whether the lung mycobiomes is affected by the use of ICS. Fourth, information on the lung mycobiome might reveal an advantageous composition of fungi, enabling us to design pro- or prebiotic treatment, or detect specific species or strains with detrimental effect on lung health, enabling effective antimycotic treatment. Finally, besides the ability to cause life-threatening infections, fungi can interfere with the immune system. As discussed earlier, PAMPs present in the fungal cell wall are capable of initiating an immune response if they are recognised by PRR. The respiratory epithelium plays a key role in the response to fungi, and both innate and adaptive immune responses are activated (69). Knowing that part of major respiratory diseases like COPD and asthma are thought to include an inflammatory ingredient, revealing the potential role of fungi in these diseases seems important. Also, other chronic respiratory diseases like CF and other bronchiectatic conditions have been linked to fungi (135), and a rich lung mycobiome exists in both disease and health.

6.3.1.1 The healthy lung mycobiome

The healthy lung mycobiome has been investigated using culture-independent techniques in a few studies. The first published paper on the healthy lung mycobiome reported a low fungal DNA burden from environmental agents such as *Davidiellaceae* and *Cladosporium*, and some low abundances of *Aspergillus* (100), while a later study found that *Eremothecium sinecaudum*, *Systenostrema alba*, *Cladosporium cladosporioides* and *Vanderwaltozyma polyspora* were particularly prevalent in the sputum of control subjects (128). *E. sinecaudum*, *V. polyspora*, and *S. alba* are fungi and microsporidia isolated from soil and plants with no known clinical pathogenicity, but *C. cladosporioides* has been described in infections in immunocompromised patients (135). Bittinger et al. included 12 healthy controls in a mycobiome study, and found that BAL samples contained *Cladosporium* and *Debaryomyces*, while oral washes (OW) included amounts of *Candida* (130). Another study reported on the mycobiota from 24 controls in a HIV study (113). The investigators performed both 18S and ITS sequencing, and included OW, induced sputum (IS), and BAL samples. They found an overrepresentation of *Pneumocystis jirovecii* in HIV patients compared to the healthy controls, and the healthy BAL mycobiome was characterised by *S. cerevisiae*, and *Penicillium brevicompactum*. They also showed that BAL samples clustered together with OW samples in principal coordinates analysis (PCoA), i.e. a method to explore and to visualise similarities or dissimilarities of data, while IS samples overlapped in part with OW samples, but not with BAL. Together with observed differences in species prevalence between IS and BAL, and BAL compared with OW, they concluded that shared and unique fungi were found in each respiratory tract level (113). Among intensive care unit (ICU) patients without pneumonia serving as controls in a pneumonia mycobiome study, the majority of the mycobiome was made up of *Saccharomycetes,* including *Candida* (129). However, it is doubtful whether these should be considered as healthy subjects. For instance, as much as 28% of the control subjects had acute respiratory distress syndrome, and 16% were on immunosuppression. A somewhat similar study was performed in another ICU setting to examine the presence of *Candida* in the lungs, and also included healthy subjects (132).

Healthy subjects were adult patients with healthy respiratory tract undergoing elective plastic surgery. *Candida* sequences were not present in mycobiota of healthy controls. All in all, the healthy lung mycobiome is understudied, but some studies are starting to fill the void. A growing literature is also found on the diseased lung mycobiome (95, 100, 108, 113, 127-134).

6.3.1.2 The diseased lung mycobiome

The diseased lung mycobiome was investigated with cultured-independent techniques for the first time in 2012. Delhaes et al. included four patients in a CF study, and collected two temporal sputum samples from each participant (95). They found that reduced fungal diversity and richness of fungal and bacterial communities were associated with poor clinical status, which means that more fungal species were seen in patients with better clinical measures. Another CF study found that *Candida* represented the dominant genus, and that *Malassezia* occurred in all samples (127). *Malassezia* appeared in one of the two samples from each participant in the CF study by Delhaes et al. as well (95). *Malassezia* is known to be associated with atopic dermatitis, and it was suggested to pay particular attention to *Malassezia* species in further lung mycobiome research (135). CF patients were also studied by Kramer et al (131). They found that the airway mycobiome was dominated by *Candida*, possibly acting as a coloniser of the CF airways.

Several factors involved with a lung transplantation including antibiotics, immunosuppression, and structural changes all predispose for fungal growth. Charlson et al. have investigated BAL and OW samples of lung transplant patients and compared to healthy controls (100). OW samples were dominated by *Candida*, and BAL samples contained mostly *Candida*, but also some *Aspergillus*. Control samples had a higher fungal richness than samples from lung transplants. This data set was expanded with new samples by Bittinger et al. (130). The added data revealed that increasingly severe pulmonary and immunologic deficits resulted in a higher colonisation in BAL by fungi with known pathogenic potential (130).

The lung mycobiota of ICU patients have been investigated in two studies. *Candida* species were the most abundant in ICU patients with pneumonia, and *Candida utilis* was more abundant in controls compared to pneumonia patients (129). Krause et al. showed that *Candida* was part of the lung mycobiome of various intubated and mechanically ventilated ICU patients with and without antibiotic therapy and with and without pneumonia. They also showed that admission to the ICU altered the lung mycobiome to be dominated by *Candida*, while antibiotics to treat pneumonia did not (132).

Only a few culture-independent studies have been performed on the lung mycobiota of asthma patients. One study included 30 asthma patients and 13 control subjects (128). Around 50% of the reads from asthma patients were classified as *Psathyrella* and *Malassezia* spp.. *Psathyrella candolleana*, *Malassezia pachydermatis*, and *Termitomyces clypeatus* were overrepresented in asthma patients, while *E. sinecaudum*, *S. alba*, *Cladosporium clasdosporioides*, and *V. polyspora* showed a higher percentage of reads in the controls (128). Another closely related disease, COPD, is also understudied.

6.3.1.3 The COPD mycobiome

A summary of published papers on the COPD mycobiome is presented as part of Table 2. Little interest has been shown to date regarding the role of the mycobiome and development and disease progression of COPD. This is somewhat surprising, knowing that COPD possesses several factors that facilitate a rich fungal community. For instance, a disturbed airways architecture, bronchiectasis, use of ICS, and frequent use of antibiotics predispose for increased fungal growth, and all factors are quite frequent in COPD patients. HIV animal studies have related the fungus *Pneumocystis* to airway remodelling and inflammation (137, 138), and a human study has shown that HIV patients that are colonized with *P. jirovecii* have more airway obstruction (139). The *P. jirovecii* colonised patients in the latter study also had significantly higher levels of matrix metalloproteinases (MMPs). MMPs have been suggested to be involved in the

COPD pathogenesis. Thus, colonisation with *Pneumocystis* could result in the release of MMPs both from the fungi and the host lung, which eventually deteriorate the airway obstruction. Other fungi are largely unexplored. In fact, only two studies investigating the lung mycobiome by culture-independent techniques have included COPD patients. The first study was performed in 2015 by Cui et al. (113). OW, BAL, IS, and control samples were collected from 10 HIV-infected individuals with COPD and compared to HIV-infected individuals with normal lung function. The primary fungus enriched in the lungs of individuals with HIV and COPD was *Pneumocystis*. But the fungal communities as a whole were also altered, and other fungi not yet reported to be associated with human disease were also overrepresented in the BAL of HIV-infected individuals with or without abnormal lung function. The second COPD mycobiome paper collected repeated sputum samples in patients with severe COPD hospitalised for acute exacerbations (108). None of the participants had a stable mycobiome during their hospital stay. Samples were dominated by *Candida*, *Phialosimplex*, *Aspergillus*, *Penicillium*, *Cladosporium,* and *Eutypella*. No comparisons were done to a healthy or stable COPD cohort. There is clearly a need for more studies on the COPD mycobiome.

Some limitations apply for the few existing COPD mycobiome studies. Due to the risk of oropharyngeal contamination, it is suggested that sampling of the airways should be done using bronchoscopy with protected sampling (9). It is difficult to reach a sufficient number of participants which has been underlined by the two COPD studies including 10 (113) and 6 (108) COPD patients. Furthermore, COPD patients in the paper by Cui et al. were also infected with HIV, while the participants in Su et al.'s paper were sampled during COPD exacerbations. No one has ever sampled non-immunocompromised patients and compared to a large healthy control population. The lower airways are also exposed to oral fungi through aspiration, and the sampling of the lung mycobiome is prone to contamination from the oral mycobiome. The oral mycobiome should be investigated in all lung mycobiome studies to account for contamination and explore associations and (dis)similarities between the two mycobiomes. The healthy oral

mycobiome was investigated with culture-independent techniques for the first time in 2010 by Ghannoum et al. (116). They found *Candida* in 75% of individuals (28), but *Cladosporium*, *Aureobasidium*, *Saccharomycetales*, *Aspergillus*, *Fusarium*, and *Cryptococcus* were also frequently observed. Later research has been in good agreement with this result (29), only adding *Malassezia* as a commensal (30). Contamination from air and study equipment are also likely to occur in lung mycobiome studies, and studies should include contamination controls from every sample collection in their study design. Only one of the studies including COPD patients collected repeated samples (108). The sampling period only extended from 7 to 16 days, and patients experienced an exacerbation of severe COPD. Four other lung mycobiome studies have included sequential sampling (95, 127, 131, 132). However, studies differed in elapsed time between sampling, included different study groups, and had low sample sizes, which complicate a general conclusion. Longitudinal data is essential to reveal fungi's implication in the progress of chronic respiratory diseases, and also to examine therapy outcome (135).

Summarising data on the COPD mycobiome up to 16th of September 2017 (Table 2), there is clearly a need for larger studies with a well-characterised disease population, ample healthy controls, and longitudinal data. To facilitate bronchoscopy studies of this scale, it is also of interest to examine participation in research bronchoscopy studies both in the literature and by original data.

7 Objectives

The objectives of this thesis were to:

Investigate participation in research bronchoscopy studies by performing a literature review on participation motives and barriers to participation, response rates, and recruitment strategies in studies involving bronchoscopy with an emphasis on studies including COPD patients.

Investigate response rates and reasons for non-response in a study with bronchoscopic sampling. Additionally, we aimed to report participation motives and examine demographic predictors for these motives.

Characterise the oral and lung mycobiomes in health and disease by comparing subjects with COPD to subjects without lung or airway diseases. Furthermore, we sought to relate the taxonomy and alpha and beta diversity of the mycobiome to the use of ICS.

Assess the stability of the lung mycobiome over time by comparing the lung mycobiome at the first and second bronchoscopy in 51 subjects with and without COPD.

8 Material and methods

The current thesis is based on data from a systematic literature review on participation in research bronchoscopy studies, and research data from MicroCOPD. The study design of MicroCOPD was described in a previously published paper (140). MicroCOPD was performed to compare the lung microbiome in subjects with and without (controls) COPD. The initial goal was to include 300 subjects with COPD, and 200 subjects without COPD, aiming to be the world's largest single-site study of the lung microbiome. As the project moved forwards, some subjects with asthma were also included. Sampling of the lungs was performed with bronchoscopy. Data collection was between April 11th, 2013, and June 5th, 2015, and the sequencing of the bacterial microbiome was completed in June 2017. The results from the latter have been published (9). A pilot study including eight COPD patients was conducted in 2012 for protocol improvement. Data from the pilot study was included in the paper examining participation, but not in the lung mycobiome papers. Eventually, 103 controls, 130 subjects with COPD, and 16 subjects with asthma participated in the MicroCOPD study.

8.1 Systematic literature review

Two systematic literature searches were performed between December 2013 and February 2014 in the PubMed search engine of the US National Library of Medicine (141) and the Excerpta Medica Database (Embase) provided by the medical publisher Elsevier (142). Relevant search terms were found by completing a modified PICO scheme (94) with colleagues, and by identifying search terms from papers in initial searches. The PICO scheme is shown in Table 3 (reproduced from paper I published in European Clinical Respiratory Journal).

Table 3. Modified PICO scheme used for a literature review on participation in bronchoscopy studies

P1	P2	I	O
COPD (MeSH)	Patients (MeSH)	Bronchoscopy (MeSH)	Patient participation (MeSH)
COPD (tw)	Patients (tw)	Bronchoscopy (tw)	Participation (tw)
Chronic obstructive pulmonary disease (tw)	Participants (tw)		Response (tw)
	Human volunteers (MeSH)		Non-response (tw)
	Volunteers (tw)		Attitude (MeSH)
	Study (MeSH major topic)		Attitude (tw)
	Trial (MeSH major topic)		Motivation (MeSH)
	Research subjects (MeSH)		Motivation (tw)
	Research subjects/psychology (MeSH)		Refusal to participate (MeSH)
	Clinical research (MeSH)		Refusal to participate (tw)
			Informed consent (MeSH)
			Participation rate, patient (MeSH)
			Participation rate (tw)
			Patient selection (MeSH)
			Patient selection (tw)
			Advertising as topic (MeSH)
			Advertising (tw)
			Risk assessment (MeSH)
			Risk assessment (tw)
			Altruism (MeSH)
			Altruism (tw)
			Perception (tw)

P, population; I, intervention; C, comparison; O, outcomes; MeSH, medical subject headings; tw, text words. The P1 column was excluded in the final search due to a paucity of results.

COPD was intended to be part of the PICO scheme, but the term was excluded due to a low number of retrieved papers. PubMed uses so-called “medical subject headings” (MeSH) terms to index their papers (143). Relevant search terms were included as both MeSH terms and text words, meaning that we included two search strategies. In a text word search, PubMed scans the whole record of the article: title, abstract, list of applied MeSH terms, list of authors, and journal name, and papers containing the queried text word are retrieved (144). A more focused search can be done by using MeSH headings and subheadings (144). Because of the hierarchical organisation of MeSH terms, generalised MeSH terms include papers classified by specific MeSH term. A similar subject heading system called Emtree exists in Embase. By including subject headings with the explode function in Embase, the search will include the selected subject heading

and all of the narrower terms below the heading. The multipurpose function in Embase is a keyword search in several parts of a paper, including title, abstract, subject headings, and more. MeSH terms were thus replaced by explosion search, and text words were replaced by multipurpose terms in our Embase search. All titles and abstracts from the PubMed and Embase searches were reviewed and classified according to exclusion and inclusion criteria. Number of retrieved papers, and how they distributed according to the given classification criteria is presented in Table 4 (reproduced from paper I published in European Clinical Respiratory Journal). Papers from both searches were included for in-depth review if they examined recruitment in bronchoscopy studies. Specifically, papers in English or a Scandinavian language including participation motives or perceived benefit of participation, non-response reasons, recruitment sources, and/or response rates in studies involving respiratory invasive procedures were included. Papers were not relevant if no humans were included, if they were case studies or secondary publications (literature reviews, reports, comments, letters, guidelines, newspaper articles, books, or book chapters), or if they had other main outcomes than participation.

Table 4: Number of retrieved papers for a literature search in the databases PubMed and Embase on participation in research bronchoscopy studies according to classification criteria

Classification criteria	PubMed	Embase
Total number retrieved	**989**	**987**
Non-English/non-Scandinavian language	102	None
Case studies/series	82	427
Secondary publications	116	108
Non-human studies	7	None
Participation not a major topic	674	443
Papers included in review	**8**	**9**
Papers common to PubMed and Embase searches	7	

Secondary publications included reviews, expert panels, letters, guidelines, and so on. The Embase search excluded studies in languages other than English and Scandinavian languages, as well as non-human studies.

8.2 MicroCOPD

8.2.1 Study design

MicroCOPD was a single-centre observational study carried out in Bergen, Western Norway, at the outpatient clinic at the Department of Thoracic Medicine, Haukeland University Hospital. It was, to our knowledge, the largest single-centre lung microbiome study performed when the study enrolment ended. The MicroCOPD study included 249 participants, with 323 performed bronchoscopies. That means, 62 participants underwent two bronchoscopies (re-bronchoscopy), and 12 participants had three bronchoscopies. The study was conducted in accordance with the declaration of Helsinki and guidelines for good clinical practice. The regional committee of medical ethics Norway north division (REK-NORD) approved the project (project number 2011/1307). Participants received no immediate benefit other than a thorough examination including a consultation with a physician, free transportation to the hospital, and reimbursement of parking expenses. All participants had to provide a written consent. Response rates and motives and non-response reasons for participation in research bronchoscopy studies, and investigation of the lung mycobiome was also explored as part of the MicroCOPD study.

8.2.2 Study population

Participants were mainly recruited from two previous large COPD studies performed by our research groups, namely The Bergen COPD cohort study (BCCS) (145), and the GeneCOPD study (146). Some COPD patients attending the outpatient clinic at Haukeland University Hospital for clinical purposes were asked to participate if deemed eligible. Furthermore, some participants were recruited among hospital staff at Haukeland University Hospital. A few individuals contacted our research staff and asked to be included through attention from a local media coverage. In total, 14 participants were recruited from the outpatient clinic and by own initiative, accounting for a small part of the total number of participants.

Individuals with and without COPD or asthma were invited to participate, but asthma patients were not included in the mycobiome analysis. Controls could either be current-, ex-, or never-smokers. Current smokers smoked at the time of participation, while ex-smokers and never-smokers did not smoke at the time of participation. Ex-smokers had a history of smoking in the past, while never-smokers had never tried smoking. Potential participants were not eligible for participation if they had increased bleeding risk or poor cardiac status judged by the study physician, hypercapnia, hypoxemia despite oxygen supply, or were allergic against lidocaine. To investigate the stable state of the microbiome, we postponed participation for subjects that were currently on, or had recently had, a course of systemic antimicrobial treatment or systemic steroid treatment. Specifically, no participant should have used antibiotics the 2 weeks immediately preceding participation, and COPD patients should not have been admitted to hospital due to COPD last 2 weeks. Furthermore, participants with symptoms of an ongoing systemic or respiratory infection could not attend, but had to postpone participation. A modification of Anthonisen's symptom criteria (147) were used to rule out or confirm a likely respiratory infection or acute exacerbation of COPD, where presence of two major symptoms (i.e. increased dyspnoea, colour change or increased sputum) or one major and one minor symptom (increased cough, sore throat, runny/stuffed nose, asthenia, wheezing from the chest) excluded attendance. A summary of exclusion and postponement criteria are given in Table 5.

Table 5: Exclusion criteria for participants with COPD and controls in the MicroCOPD study

Exclusion criteria	Detailed description
Increased bleeding risk	Double platelet inhibition, oral anticoagulant therapy, treatment with clopidogrel, ticagrelor, or low molecular weight heparin. Total platelet count < $75*10^9$, INR > 2.0, or presence of a coagulopathy.
Poor cardiac status	Cardiac valve prosthesis, known severe pulmonary hypertension, or acute coronary syndrome during the preceding 6 weeks.
Hypercapnia	Arterial CO_2 tension > 6.65 kPa
Hypoxemia	Arterial O_2 tension < 8.0 kPa or SpO_2 < 90% despite 3 litres/minute oxygen supply
Comorbidities	Control: Treatment for lung or airway disease
Allergy	Allergic against lidocaine or alfentanil
Healthcare utilisation	Antibiotic or systemic steroid use last two weeks. COPD: Admission due to COPD last two weeks.
Infection	Ongoing respiratory symptom exacerbation defined as either two major symptoms (increased dyspnoea, colour change, or increased sputum production) or one major and one minor symptom (increased cough, sore throat, runny/stuffed nose, asthenia, or wheezing from the chest).

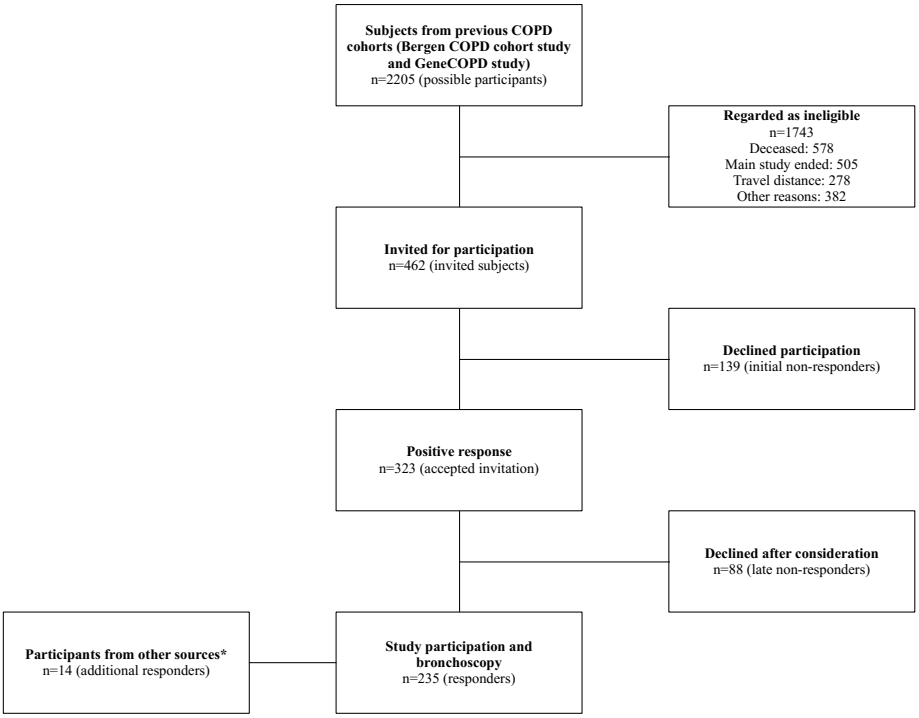
INR, International Normalized Ratio; CO_2, carbon dioxide; kPa, kilopascal, O_2, oxygen; SpO_2, oxygen saturation as measured by pulse oximetry. Absolute exclusion criteria are given in bold, while participants with recent healthcare utilisation or infection had to postpone participation.

Included COPD patients had a post-bronchodilator FEV_1/FVC ratio below 0.7 according to the GOLD report (14). Subjects without COPD or other lung diseases and without significant airflow obstruction were defined as controls. In some instances, the differentiation between COPD and asthma can be challenging, and also differentiation between COPD stage I and healthy smoking controls. Diagnoses where the study physician was in doubt were re-evaluated by three experienced pulmonologists from our research group. These decisions were based on available spirometry, radiologic imaging, respiratory symptoms, and disease history. A small number of subjects were then re-categorised: 6 “unclear asthma or COPD” were defined as COPD in 5 of the cases, and one control was re-categorised as COPD. A total of 22 controls providing samples for the fungal analyses had a ratio of FEV_1/FVC lower than 0.7, but did not have symptoms of COPD and remained classified as controls.

Theoretically, all participants included in the BCCS and the GeneCOPD study were eligible for participation. However, the majority were classified as ineligible, mainly due

to death or that the MicroCOPD inclusion period had ended. A flow chart of the inclusion process is presented in Figure 1 (reproduced from paper II published in Multidisciplinary Respiratory Medicine).

Figure 1: Flow chart of the MicroCOPD study



* Local media, hospital staff, and outpatient clinics were regarded as other sources.

In total, 235 potential participants accepted an invitation to participate, while an addition of 14 individuals from recruitment sources other than the BCCS and the GeneCOPD studies also participated.

8.2.3 Data collection

All data collection was performed at the outpatient clinic of the Department of Thoracic Medicine at Haukeland University Hospital over one or two days. Both biologic materials in the form of samples, and non-biologic information were collected. Collected information was used to create the so-called metadata, i.e. data about the data. This term is used in bioinformatics where the metadata describes co-variates that describe the samples, circumstances surrounding the sampling, which subject that provided the sample, demographics related to this subject, and more. The metadata is stored in a large tab-delimited file which is used for analysis.

8.2.3.1 Metadata

Most of the metadata was collected through a structured interview conducted by study personnel at the day of the bronchoscopy. However, invited individuals that declined to take part were asked about their reason not to participate in order to examine non-response reasons. All study interviews were conducted in the same way as specified by a study questionnaire. The questionnaire was developed by our research group, mostly consisting of previously validated instruments, written in Norwegian (see **16.1 Appendix A. Questionnaire, MicroCOPD study**). The questionnaire gathered information on:

- Potential contraindications, including questions on COPD exacerbation/lower respiratory tract infection
- Current medications and vaccination status, including antibiotic and steroids use
- Comorbidities
- Demographics (age, sex, marital status, children, education, menopause, pets/livestock)
- Tobacco smoking and alcohol habits

- Dyspnoea assessments (COPD assessment test (CAT) (148), modified Medical Research Council dyspnoea scale (mMRC) (149) except that 100 yards was replaced with 100 meters, and a Borg CR10 Scale® with 19 points as found in the 2010 folder (150))
- Motivation for participation, and pre-procedural anxiety
- Participants' perception of the bronchoscopy procedure

Perceptions of side effects and discomfort associated with the bronchoscopy was evaluated immediately after the procedure, two hours after the procedure, and after one week. The interviewer was instructed to ask these questions on perception exactly as they were written. An open question on motivation was first included in the study questionnaire from the fifth pilot patient and was asked immediately prior to the procedure.

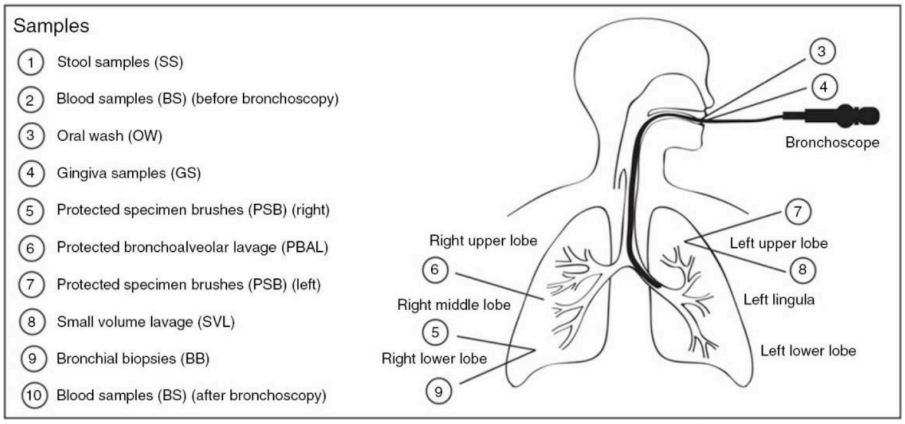
All subjects underwent a spirometry at least 15 minutes after bronchodilation with 400 microgram salbutamol administered through a large-volume spacer prior to the bronchoscopy. Spirometry was performed with a Viasys Vmax ENCORE (VIASYS Healthcare, Inc. Conshohocken, PA, USA). Absolute numbers of FEV_1 and FVC in litres were obtained, and values in percentage of predicted were calculated using Norwegian reference values from Johannessen et al. (151). Peripheral venous blood samples were drawn both before, and one hour after the bronchoscopy. An arterial blood gas was taken and analysed on a Radiometer ABL 800 flex (Radiometer Medical ApS, Brønshøj, Denmark). Anthropometric measures including weight, height, waist circumference, and bioelectrical impedance using Bodystat 1500 (Bodystat Ltd, Douglas, Isle of Man), and a blood pressure measurement were also collected with Omron HEM-757 (Omron Corporation, Kyoto 600-8530, Japan). Spirometry, blood samples, arterial blood gases, anthropometric measures, and blood pressure measurements were performed by trained study technicians. Participants were offered participation in a concurrent study which included a Siemens Somatom definition flash (Siemens Healthcare GmbH, Erlangen,

Germany) computed tomography (CT) scan of the heart and lungs. For the current studies, CT results were only used when we sought to clarify the participant status (COPD or asthma patients/controls).

8.2.3.2 Sample collection

An overview of the biological sample collection performed before, during, and after the bronchoscopy is given in Figure 2.

Figure 2: Sample collection in the MicroCOPD study



The figure is a modified version of Fig. 1 in the MicroCOPD study design paper (140).

Samples from the lungs were obtained from given locations and in a given order, as specified in Figure 2. Note, however, that the left-right order was converted a while into the study implementation to look for microbiota differences related to sampling order. Biopsies were always taken last.

8.2.3.2.1 Pre-bronchoscopic sample collection in MicroCOPD

Stool samples, blood samples, OW, gingiva samples, and negative control samples (NCS) were collected before the bronchoscopy. Participants were given a faecal sample kit at

their first visit at the outpatient clinic. Samples were obtained at home, and brought to the next outpatient clinic visit. The gingival samples, blood samples, and stool samples were not utilised in the current papers. The OW sample was collected for each participant by gargling 10 millilitre (mL) of phosphate-buffered saline (PBS) water for 1 minute, and then delivered in a sterile Eppendorf tube. The PBS was taken from a sterile bottle which was opened just prior to the procedure, or within 24 hours of a previous procedure. A PBS sample was taken directly from the bottle to serve as a NCS to account for contamination. That means, that a NCS was obtained for each participant using the same PBS as for OW and the samples obtained during bronchoscopy. The PBS had been sterilised by sterile filtration (0.22 μm) and autoclaving at 121 °C for 15 minutes by the manufacturer. Gingival samples were obtained by inserting sterile paper points to the gingiva in 5 interdental spaces in both the upper and the lower jaw.

8.2.3.2.2 Bronchoscopic sample collection in MicroCOPD

All bronchoscopies were performed by one of six experienced study physicians and one of two trained study nurses. Participants had to be fasting for at least 4 hours. Prior to procedure initiation, participants were under surveillance using three-lead ECG-monitors, pulse oximetry, and non-invasive blood-pressure measurement. Our outpatient clinic uses Olympus bronchoscopes, and actual scope used per procedure was determined by availability and wanted diameter. The largest bronchoscope used (Olympus BF-XT160) had an outer diameter of 6.3 millimetre (mm), and working channel diameter of 3.2 mm. The smallest used (Olympus BF-P180 or similar) had an outer diameter of 4.9 mm, and inner working channel of 2.0 mm. Bronchoscopes were cleaned after each procedure. We performed at maximum three bronchoscopies per day, explaining the use of different bronchoscopes. Figure 3 includes pictures from the cleaning machine and the bronchoscopy storage cabinets at the outpatient clinic where all the bronchoscopies were performed.

Figure 3. Facilities at the outpatient clinic where all bronchoscopies in the MicroCOPD study were performed

A) Cleaning machine **B) Storage of bronchoscopes**

The bronchoscopy was performed with the participant in supine position using oral access. Topical anaesthesia was applied using 10 milligram (mg)/dose lidocaine oral spray onto the base of the tongue and down the oropharynx pre-procedurally. 20 mg/mL lidocaine was delivered per-operatively through a catheter within the bronchoscope's working channel to the vocal cords, trachea, and bronchi. All participants were offered light sedation with alfentanil, and midazolam could be given intravenously in case of great anxiety as evaluated by the study physician. No suction was applied prior to having entered the trachea. The bronchoscopic sampling was completed using three different sampling methods, namely protected specimen brushings (PSB), lavage (fractioned BAL and small-volume lavage (SML)), and endobronchial biopsies. Three wax-plug PSBs (Conmed, Utica, NY, USA) were obtained from the right lower lobe (RLL) and the left upper lobe. We installed two fractions of 50 mL PBS in the right middle lobe (RML) through a wax-tip protected catheter (Plastimed Combicath, prod number 58229.19) inserted in the bronchoscope working channel. PBS used for BAL was taken from the same procedure-specific PBS bottle used for NCS and OW (see 8.2.3.2.1 Pre-bronchoscopic sample collection in MicroCOPD). Yield of the installed PBS was immediately collected by manual aspiration to the same 50 mL sterile syringe used for installation. For BAL to be performed, we required a measured FEV_1 of at least 30% of predicted and above 1.0 litres. SVL was obtained by installing 20 mL PBS, and then suctioned into two serially connected lavage traps. A maximum of six biopsies from the RLL were taken. Three of the biopsies were placed in fixative for immunohistochemistry, and three were flash frozen in liquid Nitrogen.

For mycobiome analysis, only OW, the second fraction of BAL, and a NCS was used for each participant. That means that for each participant we had three samples. However, due to minimum requirements in FEV_1 measures for BAL collection, BAL samples were not obtained for all participants. In total, 206 participants provided OW, BAL, and a NCS, where 13 of the participants had asthma and were not included in fungal analysis. Thus, 579 samples were included in the mycobiome analyses.

8.2.4 Laboratory processing

8.2.4.1 Fungal DNA extraction

Fungal DNA was extracted together with bacterial DNA using a protocol developed in-house by Tuyen Thi Van Hoang (University of Bergen) and professor Harald Gotten Wiker (University of Bergen). The protocol for DNA extraction, bacterial PCR, and bacterial sequencing of samples from the MicroCOPD study has been published (152, 153). Only the DNA extraction step was used for the fungal samples.

In summary, fungal DNA was extracted using a combination of enzymatic and mechanical lysis. We used 1800 microlitre (mcL) of the OW samples and the BAL samples, and 450 mcL of the NCS as input to the extraction process. We then treated each sample with an equal amount of Sputasol (Oxoid Limited, England) as the input sample volume, before samples were incubated at 37 °C for 15 minutes on a thermomixer (1000 rpm). Samples were centrifugated at 15700 x g at room temperature for 8 minutes. The resulting cell pellet was resuspended in 250 mcL PBS. Each sample were then treated with an enzyme cocktail solution consisting of 25 mcL lysozyme (10 mg/ml, Sigma-Aldrich, USA), 3 mcL mutanolysin (25 KU/mL, Sigma-Aldrich), 1.5 mcL lysostaphin (4000 U/mL, Sigma-Aldrich), and 20.5 mcL TE5 buffer (10 mM Tris-HCl, 5 mM EDTA, pH 8). Samples were then incubated on a thermomixer at 37°C with shaking at 350 rpm for 1 hour. An additional centrifugation was performed to pellet any cells not lysed. The pellet was resuspended in 800 mcL CLS-TC buffer (FastDNA SPIN kit, MP Biomedicals, LLC, Solon, OH, USA). The resulting suspension was transferred to a Lysing Matrix A tube (FastDNA SPIN kit), and homogenised in the FastPrep-24 instrument at a speed setting of 6.0 m/s for 40 seconds. Samples were centrifuged at 14000 x g at room temperature for 10 minutes before the 650 mcL supernatant was combined with 300 mcL of the supernatant from the enzymatic lysis. The supernatant from the enzymatic lysis had not been treated with mechanical lysis to avoid any potential DNA shearing. Further processing included addition of Binding Matrix (FastDNA SPIN kit), incubation, spinning through a SPIN filter (FastDNA SPIN kit), and

addition of prepared SEWS-M (FastDNA SPIN kit) and DES (FastDNA SPIN kit) before incubation and centrifugation.

DNA concentration, or yield, were measured using the Qubit quantification platform (Thermo Fisher Scientific Inc, Waltham, MA, USA).

8.2.4.2 Fungal PCR and high-throughput sequencing

The Illumina 16S Metagenomic Sequencing Library Preparation guide (Part no. 15044223 Rev. B) served as a template for the sequencing library preparation, but with some minor modifications as noted in the following. First, an amplicon PCR was carried out targeting the fungal ITS1 region using primer set ITS1-30F/ITS1-217R, which sequences are GTCCCTGCCCTTTGTACACA and TTTCGCTGCGTTCTTCATCG (154). The number of cycles was increased from 25 to 28 in the amplicon PCR. An index PCR was then performed to add Illumina sequencing adapters and dual-index barcodes. The number of cycles was increased from 8 to 9 in the index PCR. Libraries were then quantified again. Samples were loaded on an Illumina HiSeq sequencing platform (Illumina Inc., San Diego, CA, USA) and underwent 2x250 cycles of paired-end sequencing in three separate sequencing runs. The PCR and sequencing were performed by Sequentia Biotech SL (Barcelona, Spain, https://www.sequentiabiotech.com/).

8.2.5 Bioinformatic analysis

The Illumina sequencing generates a large number of data files with biological data and quality scores in a specific format called the FASTQ format. The generated data need to be processed with methods and software tools to make any sense. We chose to use selected tools (plugins) provided within the second version of Quantitative Insights into Microbial Ecology (*QIIME 2*) bioinformatic package for the upstream analysis (155), and suitable packages in *R* (63) for downstream analysis.

8.2.5.1 Upstream bioinformatic analysis

FASTQ files created by the Illumina sequencing include up to thousands of lines with DNA sequences or reads. Due to the paired-end sequencing design in the current study, both ends of the fragment were sequenced in opposite directions. One FASTQ file is generated for the forward direction, and one FASTQ file is generated for the reverse direction. The forward and reverse read can be merged into a single read if they were sufficiently overlapping. Overlap in paired reads in our study were evaluated using *BBMerge* (156), and samples with a lack of overlapping forward and reverse reads were removed. Most of the remaining upstream bioinformatic analysis were then performed using *QIIME 2* (versions 2019.1 and 2019.10).

Fungal ITS primers are often located in the more conserved 18S/5.8S genes or the 5.8S/28S genes flanking the ITS regions and keeping these more conserved regions has been shown to create mis-assignments in downstream analyses (157). We therefore trimmed our reads using the *q2-itsxpress* plugin specified for the fungal ITS1 region and clustered at 100% identity (157). We then performed a quality control of the trimmed reads using the Divisive Amplicon Denoising Algorithm version 2 (*DADA2*) *q2-dada2* plugin (158). *DADA2* is used to perform a so-called denoising by filter out low-quality sequences and remove chimeric sequences and singletons. Chimeric sequences are spurious sequences made by combining fragments from different sequences during PCR. PCR amplification spontaneously abort from time to time. The resulting, incomplete amplicons could then serve as primers in the next PCR cycle, anneal to another template and generate an artificial sequence called a chimera. *DADA2* utilise an algorithm that models the errors introduced during sequencing to infer the sample composition. The resulting inferred sample sequences are called exact amplicon sequence variants (ASVs). ASVs are replacing the formerly used operational taxonomic units (OTUs). OTUs are sequences that are clustered based on a similarity threshold chosen by each individual researcher for their projects, usually 97%, and then used as a proxy for taxa (47). The ASVs are suggested to detect more fine-scale variations than the OTUs (158). Finally,

DADA2 merge the paired-end reads. *DADA2* was performed on a per-sequencing run basis with no truncation or further trimming in accordance with the recommendations in the *DADA2* ITS pipeline workflow (159).

Samples originating from different sequencing runs were merged after *DADA2*. Taxonomy was assigned to ASVs generated in *DADA2* using the *q2-feature-classifier* (160) *classify-sklearn* (161) with a UNITE database for fungi with clustering at 99% threshold level (162). As mentioned in the background section (see **6.3 The mycobiome**), fungal databases are often incomplete, and a portion of the fungal ASVs are often insufficiently assigned. We defined ASVs assigned only as *Fungi* at kingdom level, or *Fungi* at kingdom level with unidentified phylum as unclassified. Unclassified ASVs were subject for a manual examination. First, unassigned ASVs were checked using the Nucleotide Basic Local Alignment Search Tool (BLASTN) program maintained by the National Center for Biotechnology Information (NCBI) (163) in the Nucleotide database (164). ASVs without hits were discarded, while accession identifiers from the remaining unassigned ASVs were further investigated using NCBI's Nucleotide database tool Batch Entrez in order to filter on fungi (165, 166). Only fungal records from the Batch Entrez results were included for further assessments. The corresponding ASVs to the fungal Batch Entrez results were manually checked in detail using BLASTN. A max score is generated in each BLASTN search to show how well the query sequence aligns with sequences in the nucleotide database in terms of matched or mismatched nucleotides. Based on the max score in BLASTN, a decision was made whether it was likely that the ASV represented a fungus, and if so, what fungus it was. ASVs with ambiguous or non-fungal BLASTN results, for instance plants, were discarded. ASVs with an unambiguous BLASTN result with a high max score were repeatedly assigned to new taxonomic assignments using UNITE databases with fungi or all eukaryotes with different threshold levels (162, 167-169) (via *q2-feature-classifier* (160) *classify-sklearn* (161) and *classify-consensus-blast* (170)). If one of the repeated taxonomic assignments matched the decision made on the BLASTN result (i.e. a given fungus), the particular ASV was

assigned to this particular taxonomy. A thorough examination of the unclassified are important to reduce inclusion of likely non-fungal ASV initially assigned as fungi.

Despite removal of low-quality reads and chimeras by *DADA2*, some PCR and sequencing errors are likely to still exist in the data, potentially creating an unreasonable diversity. The rarest OTUs or ASVs are often discarded at an arbitrary level. *LULU* was created aiming to exclude artefactual OTUs without discarding rare, but real OTUs (171). It does so by investigating whether a rare OTU could be explained by co-occurrence with a more abundant OTU, and if so, subsequently merged. We curated our ASVs with *LULU* in R with 80% coverage and 84% identity in the database blast search, and the actual *LULU* curation with default settings (*minimum_ratio_type = "min"*, *minimum_ratio = 1*, *minimum_match = 84*, *minimum_relative_cooccurence = 0.95*).

ASVs present in only one sample, and ASVs with a total number of sequence reads less than 10 obtained across all samples, were filtered out.

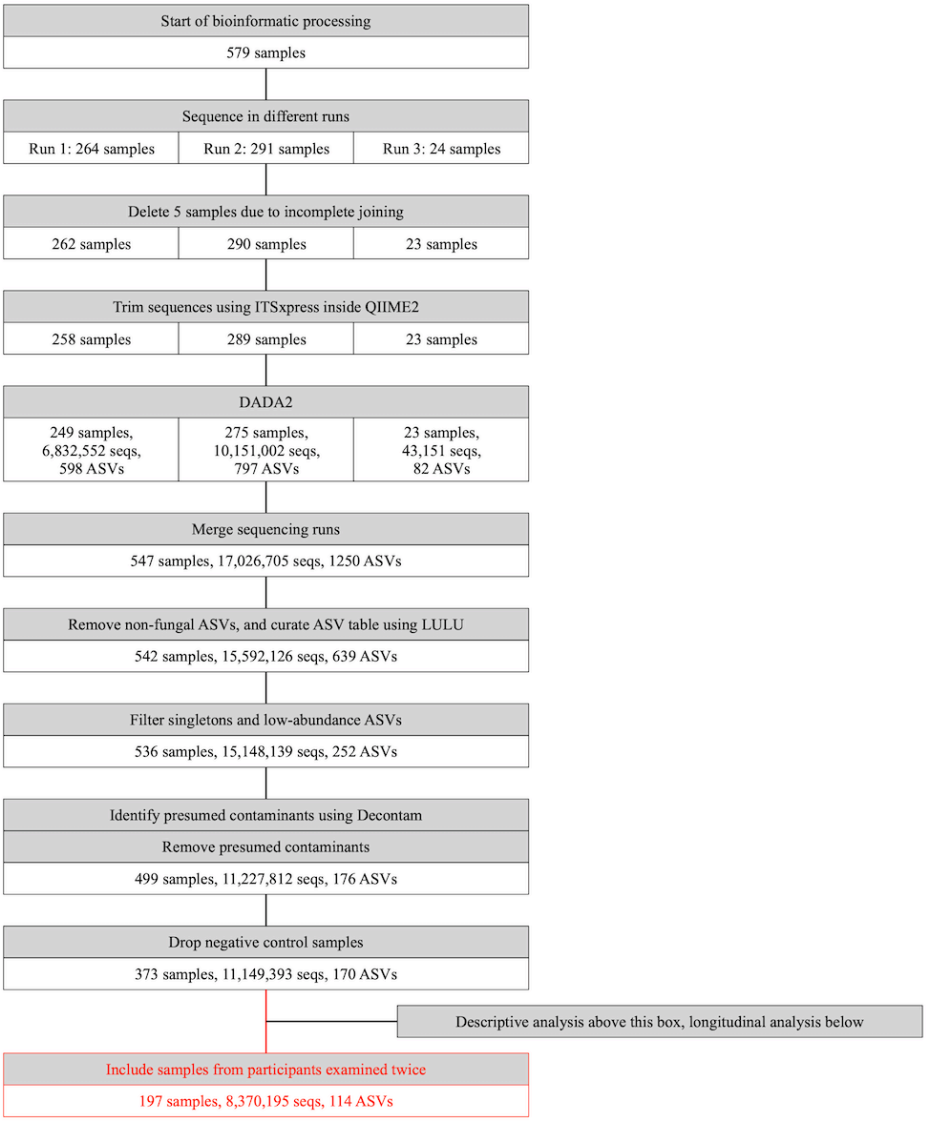
Mycobiome studies are impacted by contamination. Principally, this contamination can stem from the sampling procedure or the laboratory handling/analyses. Contaminants can be found in PCR reagents, and because of their ubiquity as airborne particles, fungi can also contaminate equipment and samples directly (172). The lower respiratory tract is a low-biomass environment, and mycobiome studies thus observe modest numbers of organisms in their samples (130). This makes contamination a particular concern. There seems to be no consensus how to handle contamination in mycobiome studies (172). We included NCS to evaluate contamination. However, different strategies exist on how to best utilise the information found in the NCS. Usage of bioinformatic tools is one solution. One of them, *Decontam* (173), has proven effective for contamination removal in lung microbiome sequencing data (174), and has recently been used by a COPD mycobiome study (175). We thus chose to identify likely contaminants with *Decontam*. *Decontam* was run with the prevalence-based approach, and a user defined

threshold of 0.5. A batched classification was used, meaning identification of contaminants was done on a per-sequencing run basis. Identified contaminants were removed from the fungal ASVs.

Samples from re-bronchoscopies were only included in the longitudinal study (paper IV), and thus removed prior to the descriptive analyses (paper III).

Flowchart Designer version 3 (http://flowchart.lofter.com) was used to create a flow chart of the bioinformatic analysis with remaining samples, sequences and ASVs as shown in Figure 4. Only samples from participants with two bronchoscopies were included in the longitudinal mycobiome analysis (paper IV), shown by the red boxes at the end of the flow chart (Figure 4).

Figure 4: Flow chart of the fungal upstream bioinformatic analysis in the MicroCOPD study



DADA2: Divisive Amplicon Denoising Algorithm version 2, seqs: sequences, ASV: amplicon sequence variant. Samples were sequenced in three different runs before trimming and denoising. Data from different sequencing runs were merged, and then further processed to exclude presumed contaminants and spurious ASVs prior to analyses. Boxes with a red line is exclusively for the longitudinal mycobiome paper (paper IV).

8.2.5.1 Downstream and statistical analysis in MicroCOPD

8.2.5.1.1 Participation analysis (paper II)

All analyses were performed using *Stata* version 14 (176). We defined all participants that underwent a bronchoscopy as responders. Non-responders were divided in two – *initial non-responders* declined participation at first approach, while *late non-responders* reconsidered an initial acceptance to participate. Bivariate comparisons between responders and non-responders, and between initial non-responders and late non-responders were performed using parametric (t-test) and non-parametric tests (chi-square test or Fisher's exact test), when judged appropriate. Demographic continuous variables were reported as means with standard deviations, and demographic categorical variables as absolute numbers with percentages.

Participants could provide as many unique participation motives as wanted, and all unique participation motives were merged into broader principal motives afterwards. These principal motives were further combined into three main groups:

1. *Altruism*. A wish to help others or continue previous study participation, or a willingness to contribute to science.
2. *Personal benefit*. A perceived hope to improve own health by participation.
3. *Obligation*. A subjective feeling of being bound to participate.

Unspecified reasons were labelled missing. We fitted bivariate logistic regression models on participation motives. We used "exclusive altruism" and "exclusive personal benefit" as binary variables. Participants that *only* had altruism as main motive or *only* personal benefit as main motive were labelled 1 in the respective variables. If participants had more than two main motives, or had obligation as main motive, they were coded "0" on the binary variables. Variables were included in the model in a step-wise manner, and covariates with a p less than 0.20 were included in multivariate models. Age and FEV_1 in percentage of predicted values were treated as continuous variables, but divided by 10 to

provide ratios for an increase of 10 units. Smoking habits were included in the logistic regression model as a binary variable with current smoking and never-/ex-smoking.

For other analyses, never-smokers and ex-smokers were handled as separate groups. Pack years were calculated by dividing number of cigarettes per day by 20, and then multiplied by years smoked. Percentage of predicted values of FEV_1 and FVC, as well as the FEV_1/FVC-ratio, were used as measures of lung function. Dyspnoea was assessed by the mMRC (149). A response rate was defined as the number of bronchoscopies performed divided by the number of invited subjects. COPD and asthma were merged as obstructive lung diseases in the response rate calculations, and calculations were stratified by sex and participant category (control/obstructive lung disease). Frequencies of given non-response reasons were compared using Chi-square test or Fisher's exact test stratified by time of non-response decision and participant category.

8.2.5.1.2 Descriptive mycobiome analysis (paper III)

Demographics were analysed using *Stata* version 15 (177). Smokers were divided into daily smokers, ex-smokers, and never-smokers in demographic analysis, but ex-smokers and never-smokers were merged into one category for taxonomic composition analysis and diversity analysis. Alpha and beta diversity analyses including participants with COPD were stratified by GOLD stage (14). Age and FEV_1 in percentage of predicted were treated as binary variables (below or above and equal to 70 years for age, and below or above and equal to 80% for FEV_1) in diversity analysis and differential abundance/distribution testing. Fungal ASVs from the upstream bioinformatic processes were collapsed at genus and phylum level before taxonomic bar plot creation and differential abundance/distribution testing either by the *q2-taxa barplot* or the *q2-taxa collapse* commands. No collapsing was done prior to the diversity analysis.

Taxonomic bar plots, rank abundance plots, and *Basidiomycota*/*Ascomycota*-ratios were made using the packages *readr* (178), *janitor* (179), *tibble* (180), *SOfun* (181), *dplyr*

(182), *tidyr* (183), *glmnet* (184), and *ggplot2* (185) in *R* (63). Taxonomic bar plots were displayed at subject level, and divided by sample type, participant group, and use of ICS. Medians of the *Basidiomycota*/*Ascomycota*-ratios were calculated in *R* (63).

The *q2-diversity core-metrics* plugin was used to generate several diversity metrics, including Shannon index for alpha diversity analysis. Samples were rarefied (subsampled without replacement) to 1000 sequences per sample before making the diversity metrics. Samples with a sequence count below this value were discarded in the diversity analysis. The rarefaction depth (i.e. 1000 sequences) was chosen based on the plateau in alpha rarefaction plots, and multiple testing with different values. We aimed to find a value as high as possible while excluding a minimum of samples. Statistical differences in alpha diversity measured with Shannon index was tested in *R* using Kruskal-Wallis for unpaired variables (age, sex, FEV_1, smoking habits, ICS use, and participant group with GOLD stages), and Wilcoxon signed-rank test for paired analyses (sample type, i.e. OW vs BAL). Shannon index comparisons on participant group and sample type were plotted using the *R* packages *ggpubr* (186) and *gginnards* (187). Both the Bray-Curtis dissimilarity and the Jaccard similarity coefficient were calculated from the rarefied ASV table using the *vegan* package in *R* (188). The *R* package *ape* was used to create PCoA plot values corrected for negative eigenvalues with Cailliez method (189). Differences in beta diversity between age, sex, FEV_1, smoking habits, and ICS use were tested with permuted analysis of variance (PERMANOVA) on OW and BAL separately, and for controls, COPD patients, and both participants groups combined. Comparisons on age, sex, FEV_1, smoking habits, and ICS use were performed unadjusted. We also tested for beta diversity differences between study groups in BAL samples both unadjusted and adjusted for age, sex, and FEV_1. Differences in spread was tested with permuted multivariate analysis of beta-dispersion (PERMDISP) for all variables. PERMANOVA tests were run with 10000 permutations, except for pairwise comparisons in which 999 were used. PERMDISP tests had 1000 permutations. Bray-Curtis and Jaccard distances were shown in PCoA plots divided by participant group and ICS use. Furthermore, PCoA

plots divided by sample type were created with and without Procrustes transformation with 999 permutations restricted to the three first axes from the PCoA result. The summed squares of deviations (M^2) and p-values from the Procrustes analysis were given in the plots. PERMANOVA, PERMDISP, and Procrustes were analysed using the *vegan* package in R (188), while *ggpubr* (186) was used to create the PCoA plots.

Differences in distributions and relative abundances on group level were evaluated by the Microbiome Differential Distribution Analysis (*MicrobiomeDDA*) omnibus test (190), the second version of analysis of composition of microbiomes (*ANCOM v2*, https://github.com/FrederickHuangLin/ANCOM) (191), and the second version of ANOVA-Like Differential Expression (*ALDEx2*) (192-194) at genus level with a processing step using *tibble* (180). All three tests were used to compare participants groups, while only *ANCOM v2* and *ALDEx2* were used to compare sample types due to the possibility to define a paired design. Differences in relative abundances on subject level, i.e. between samples provided by the same participant, were evaluated by calculating the Yue-Clayton measure of dissimilarity (1-θYC) (195). Relative abundances were plotted with the calculated Yue-Clayton measures between OW and negative control samples, BAL and negative control samples, and OW and BAL. The *R* packages *tibble* (180), *tidyverse* (196), *reshape2* (197), *randomcoloR* (198), and *ggpubr* (199) were used to create the plots.

8.2.5.1.3 Longitudinal mycobiome analysis (paper IV)

The demographic analyses were similar in the longitudinal mycobiome paper (paper IV) as in the descriptive mycobiome paper (paper III), except that intercurrent antibiotic use and time between procedures were also reported. Variables were analysed with data from the first bronchoscopy, while demographics from the second time point were not shown. Demographics were analysed with *Stata* 16 (200). Taxonomic bar plots were generated using the *R* packages *dplyr* (182), *tibble* (180), *tidyr* (183), *readr* (178), *glmnet* (184), and *ggplot2* (185) in *R* (63), and displayed at genus level after collapsing with the *q2-taxa*

barplot command. Taxonomic bar plots were stratified by participant group and use of antibiotics in the time between the bronchoscopies. Only the nine most abundant taxa were visualised individually, while taxa with low abundances were merged and visualised as *Others*. Additionally, genus level taxonomy from each bronchoscopy was plotted for each participant separately, labelled with the corresponding Yue-Clayton measure. The Yue-Clayton measures were calculated as in the descriptive paper (paper III), and plots were generated with the *R* packages *tibble* (180), *tidyverse* (196), *reshape2* (197), *randomcoloR* (198), and *ggpubr* (199).

Shannon indexes were estimated using *q2-diversity core-metrics* after samples were rarefied (subsampled without replacement) to 2000 sequences per sample. To visualise changes in alpha diversity, Shannon indexes were plotted with lines between Shannon indexes from the same participant, i.e between Shannon indexes from the first and the second bronchoscopy. The plot was made with *ggpubr* (186), and *gginnards* (187) in *R* (63). We compared Shannon indexes from the first and the second bronchoscopy using Wilcoxon signed-rank test for paired analysis stratified by sample type. The *q2-diversity core-metrics* command provided a rarefied table that were used to make beta diversity metrics (Bray-Curtis dissimilarity and Jaccard similarity coefficient) using the *vegan* package in *R* (188). Plot values used to create PCoA plots were made using the *ape R* package (189). Negative eigenvalues were corrected with Cailliez method. PCoA plots using Bray-Curtis or Jaccard were made using *ggpubr* (199). We made plots coloured by bronchoscopy number and sample type, intercurrent antibiotic use and sample type, and bronchoscopy number and intercurrent antibiotic use for both OW and BAL. The PCoA plots coloured by bronchoscopy number and intercurrent antibiotic use for OW and BAL were also plotted with a Procrustes transformation with 999 permutations using the *vegan* package in *R* (188). The Procrustes analysis was restricted to the three first axes in the PCoA result, and M^2 and p-values were given in the plots with Procrustes transformation.

Pairwise distances between a participant's samples from the first and the second bronchoscopy were calculated from both Bray-Curtis and Jaccard distance matrixes using *q2-longitudinal* (201). Calculated pairwise distances were grouped on sample type in one plot with colours indicating intercurrent antibiotic use, and comparisons were done between sample types using Kruskal-Wallis test. Pairwise distances were grouped on time between bronchoscopies measured in days in another plot. The plot grouped on time between bronchoscopies were divided in OW and BAL, and we added a different colour for each distance metric type. Pairwise distance plots were generated with *ggpubr* (199).

9 Results

9.1 Paper I – Literature review on participation

Studies involving a semi-invasive procedure like bronchoscopy is inevitable associated with some discomfort. During recruitment, possible participants are informed about the procedure and consequently, recruitment to bronchoscopy studies could be challenging. Furthermore, one might wonder why at all people choose to participate in studies involving a bronchoscopy.

As organizers of a large research bronchoscopy study, we wanted to investigate participation in bronchoscopy studies in detail. As a preparation for this, we performed a systematic literature review on this subject.

We did a systematic literature search on participation in research bronchoscopy studies in February 2014 using the search engines of PubMed and EMBASE. There was clearly a lack of information on participation in research bronchoscopy studies in the literature, and after a classification of 1,976 resulting papers, only seven relevant papers were included. Still, some main findings were worthy of attention:

1) Participants were driven by four main motives, namely personal benefit, altruism, perceived importance of research, and obedience to the authority of the researchers.
2) Response rates varied from 3 to 73%.
3) The invasive nature of the procedure influenced potential participants' view on participation negatively.
4) Radio advertisement was the most effective recruitment strategy.

The literature review also showed difficulties for bronchoscopy studies in recruiting control subjects and younger individuals. No studies included participants with COPD.

9.2 Paper II – Participation in the MicroCOPD study

The MicroCOPD study was initiated to compare the lung microbiome in subjects with COPD and controls. As one of the largest single-centre bronchoscopy studies on the lung microbiome, the study gave the opportunity to investigate participation in detail.

Participation was thus examined as a sub-study, and included examination of participants' motivation for participation, reasons to decline participation, and calculations of response rates. Participation motives and non-response reasons were collected using open questions. We compared responders and non-responders, and non-responders declining participation immediately with those declining after some waiting time. Motives were merged into broader categories called *altruism*, *personal benefit*, and *obligation*, and we fitted a bivariate logistic regression model using altruism and personal benefit as outcome.

The MicroCOPD study had an overall response rate of 50.9%. The response rate in men was significantly higher than the response rate in women (56.5% vs 44.8%, p-value 0.01), and there was no difference in the response rates for subjects without obstructive lung disease and subjects with COPD or asthma. There was a higher percentage of participants with obstructive lung diseases among the late non-responders compared with initial non-responders, although non-significant (p-value 0.06).

Altruism as participation motive was mentioned by 67.3% of the participants, while 52.2% stated personal benefit as important. Men were less likely to state altruism as their main participation motive. The most given non-response reasons were fears and worries related to study participation (40% of initial non-responders). More late than initial non-responders stated worries and fears as their non-response reason (34.1% vs 16.5%, p-value < 0.01). Health issues was a frequently given non-response reason, and especially so for subjects with obstructive lung diseases (27.3% vs 9.5%, p-value < 0.01).

9.3 Paper III – Descriptive mycobiome analysis

The fungal part of the microbiome is understudied, and we set out to characterise and compare the oral and pulmonary mycobiomes in a large cohort of participants without lung disease and participants with COPD. Additionally, we examined if OW samples differed from BAL samples, and if participants using ICS had a different lung mycobiome compared to participants not using ICS.

We created taxonomic bar plots grouped by sample type, participant group, and ICS use. Alpha diversity was calculated using Shannon index, and Bray-Curtis dissimilarity and Jaccard similarity coefficient were estimated to assess beta diversity using PCoA plots with and without a Procrustes transformation. We calculated the Yue-Clayton measure of dissimilarity between pairs of samples, i.e. OW and BAL samples. Three different differential abundance/distribution tests were used to look for differences in sample type, participant group, sex, age, smoking habits, and ICS use.

Taxonomies from OW differed compared to taxonomies from BAL with a mean Yue-Clayton measure of 0.63. OW and BAL taxonomies were dominated by *Candida*, but there was significantly more *Candida* in OW compared to BAL for both participant categories. *Malassezia* and *Sarocladium* were other common fungi found in both OW and BAL samples. No difference was seen in alpha diversity between OW and BAL samples, but PCoA plots before and after symmetric Procrustes transformation indicated that there were differences in the composition between OW and BAL samples from the same individual. We observed no difference in neither alpha nor beta diversity between participant groups, but there seemed to be a tendency towards higher proportions of *Basidiomycota* in participants with COPD compared to controls both in OW and BAL. Differential abundance/distribution tests provided varying results when comparing taxa between the participant categories, and thus, no consistent differences were found. The results indicated that ICS use did not seem to significantly affect the lung mycobiome.

9.4 Paper IV – Longitudinal mycobiome analysis

To our knowledge, no study has published data on the stability of the lung mycobiome in subjects without lung disease, while few studies have performed repeated sampling of the COPD lung mycobiome. Thus, we aimed to report on the stability of the oral and lung mycobiome in participants with and without COPD. We also aimed to examine the potential effects on the mycobiomes from intercurrent antibiotic use.

We performed repeated sampling with bronchoscopy in 30 participants with and 21 without COPD. Taxonomic bar plots were generated grouped by bronchoscopy number, study group, and intercurrent antibiotic use. We calculated Shannon index to compare alpha diversity between repeated bronchoscopies, and the Yue-Clayton measure of dissimilarity between each sample pair from the first and the second bronchoscopy. Bray-Curtis dissimilarity and Jaccard similarity coefficient were estimated, and differences in beta diversity between repeated bronchoscopies were presented as PCoA plots with and without a Procrustes transformation. Pairwise distances between the first and the second bronchoscopy were calculated and grouped on sample type and time between procedures.

We observed that the oral mycobiome showed a higher stability compared to the lung mycobiome based on three observations:

1) A visual impression of higher instability in BAL compared to OW samples.
2) BAL had a higher average Yue-Clayton measure than OW samples (0.69 vs 0.22).
3) Significantly higher pairwise distances in BAL samples compared to OW sample using the Bray-Curtis distance metric (p-value < 0.01).

We found no evidence of differences in Shannon indexes or obvious clusters in PCoA plots between repeated bronchoscopies. Intercurrent antibiotic use and time between bronchoscopies did not seem to influence the mycobiomes.

10 Discussion of methodology

A plethora of choices related to study designs, generation of variables, and statistical tests are made by the researchers. These different approaches might yield different results. Consistency in data analysis was examined in a study by Silberzahn et al. (202). They recruited 29 research teams to analyse the same data set to address whether soccer referees are more likely to give red cards to dark-skin-toned players than to light-skin-toned players. Twenty teams (69%) found a statistically significant positive effect, and 9 teams (31%) did not observe a significant relationship. With Silberzahn et al.'s study in mind, this section focuses on methodological issues for the current thesis.

10.1 Study design

One of the first choices that researchers are faced with, is the choice of study design, i.e. how the study should be conducted. To conclude on a suitable study design is of particular importance because a poorly designed study can never be recovered, whereas a poorly analysed study can be reanalysed to reach a meaningful conclusion (203). Research designs can simplified be classified into descriptive or analytical, where analytical studies attempt to analyse and draw inferences while descriptive are more summarising (204). Analytical studies can further be divided into observational if no exposure is determined or experimental if exposure is determined (204). Ultimately, studies can be cross-sectional, based on cases, or cohorts (204). Choosing an appropriate study design is a necessity to generate reliable and valid data. In contrast, inappropriate study designs generate poor data and are prone to biases.

10.1.1 Reliability, validity, and bias

Measurements of reliability and validity are important factors to assess the quality of a research study. Measurement reliability relates to the consistency of the measure (205). If a researcher repeats the study measurements under consistent conditions and receives similar results, we say that the measurements have high reliability. But measurements are

useless if they consequently evaluate inaccurately. Measurement validity refers to the extent to which a concept is accurately measured (205). A high measurement validity means we measure what we intend to measure. The combination of high measurement validity and reliability creates data that we trust. Similarly, a valid study design is needed to draw valid conclusions. The validity of study designs is commonly divided into internal and external study validity. Internal study validity assesses how well the study can draw conclusion regarding the study population, while external study validity covers the generalisation of the study results to a larger population (206). Skewness in data, or biases, are known to influence study validity. Biases are systematic errors that arise from the study methods (206), and can be divided into different types. Selection bias comes from any error in selecting the study participants and/or from factors affecting the study participation (206), and becomes a problem if the selected study participants are significantly different compared to subjects not included in the study. Information bias occurs during data collection and handling and creates systematic differences from the truth (207). Information biases could occur as a result of misclassification, observer variation, or bias in recall, and is a probable bias within observational studies (207).

10.1.2 Study design of the MicroCOPD study (paper II, III, and IV)

MicroCOPD was a single-centre observational study with repeated sampling for a sub-cohort of the participants. The aim was to examine the microbiome in participants with and without COPD. Strictly speaking, the aim made participant category an exposure while characteristics associated with the mycobiome became the outcome. Paper III utilised data from the first sampling time point. Because the analyses in paper III relied on samples at a single point in time, we could say that data in paper III mimics data taken from a study with a cross-sectional design. A limitation with cross-sectional studies is that they cannot determine any causality, because it cannot be demonstrated that the exposure preceded the outcome (208). Although we defined participant category as exposure and mycobiome characteristics as outcome above, we are unable to determine

any direction of potential associations in paper III. For instance, we cannot say whether characteristics with the lung mycobiome are involved in the COPD pathogenesis or if COPD disease features alters the lung mycobiome. However, cross-sectional studies are suitable to reveal potential associations, they are relatively easy to conduct, inexpensive, and can be carried out in a short time frame (203). As discussed above, observational studies are prone to information biases. The use of a questionnaire creates several participant-reported variables, and the risk of recall biases is present. Recall biases will be further discussed later (see **10.2.2.1 Collection of metadata**).

The MicroCOPD study is, to our knowledge, the largest single-centre study on the COPD lung mycobiome using targeted amplicon sequencing. Another COPD mycobiome study has included 337 participants with stable COPD and 47 non-diseased controls (175). In this latter COPD study, participants were recruited across five hospital sites which could introduce experimental bias. However, a strength of the multi-centre study is that all experiments were done at a single site, and that all samples were quality-controlled on arrival at the site where experiments were performed. We did not perform any sample size calculation for the MicroCOPD study, but we had a larger number of participants compared to the other published single-centre COPD lung mycobiome studies (108, 113). Participants in the MicroCOPD study were recruited from residents in the same county, and were represented by a quite heterogeneous population. We included a thorough examination of diversity and differential abundance/distribution testing to look for potential confounding effects from sex, age, smoking habits, FEV_1 in percentage of predicted, and ICS use, but no obvious effects were seen. Although the study design facilitated high internal validity, we acknowledge that generalisation to a larger population (external validity) is somewhat limited due to the confined recruitment area.

Motivation for participation, response rates, and non-response reasons were examined in the MicroCOPD population as a sub-study (paper II). While originally being designed as a microbiome study, the study design has some limitations in regards to analyses of

participation. The two main recruitment sources for the MicroCOPD study were the earlier conducted BCCS (145) and the GeneCOPD study (146). For paper II, inviting individuals with a known history of willingness to participate in previous research, we cannot exclude an introduction of selection bias with an increase in response rates as consequence. On the other hand, previous participation could frighten or tire out individuals, making future participation less likely.

Paper IV examined the stability of the lung mycobiome in participants with and without COPD, utilising that 62 participants underwent bronchoscopy twice. The longitudinal study design was well-suited to look for intra-individual differences in the mycobiome. A limitation of the study design is that participants underwent repeated bronchoscopies with different time intervals, which could introduce biases. However, the interval was random, and did not differ systematically between the groups of interest. Furthermore, we did not observe any trend in pairwise distances drawn from Bray-Curtis and Jaccard distance matrices, suggesting that beta diversity was little affected by elapsed intercurrent time.

10.1.3 Study design of paper I

For paper I, a literature search was conducted to summarise knowledge on participation in research bronchoscopy studies using both the PubMed search engine and Embase. We used a modified PICO scheme to detect relevant search terms. Specifically, keywords were found by looking at MeSH terms from relevant papers and by discussion with a colleague. More information on the literature search process is found in the Material and methods section (see **8.1 Systematic literature review**). Retrieved papers were sifted based on predetermined classification criteria. For instance, we decided that only papers written in English or a Scandinavian language should be included. In total, 102 papers were omitted due to language. The importance of choosing appropriate databases has been emphasised in a paper on successful literature searching, as each database will contain unique material (209). We did only include PubMed and Embase in our literature

search, but acknowledge that several other databases common in medical sciences exist, such as Medline (Ovid), Scopus, and Web of Science (209). Thus, we cannot exclude that we have missed relevant literature in another language or in another database.

The Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement has been developed to improve the reporting of systematic reviews and meta-analyses (210). Using PRISMA to review paper I, we admit that the search could have been described in more detail in order to be repeated by others. More focus on biases in the included papers would also have been beneficial, including publication bias, a bias occurring when the hypothesis and significance of the results have influenced the publication process. A study has described the steps involved to undertake a new or update an existing Cochrane review (211). Critical steps, including study selection, data extraction, and data analysis are suggested to be repeated by a second author. This was not performed in paper I, although close supervision was maintained during the whole review process. It could be that relevant papers were omitted due to a lack of understanding, considering that I only had finished two years at medical school when study selection was done. Analysis and writing were done in the following year. Nevertheless, the study topic was not too clinically difficult to understand, and included papers were written without use of a detailed medical language. Additionally, only seven papers were included for in-depth analysis, enabling that considerable time could be spent on each of them. We did not publish any protocol for the literature review, which runs the risk of making decision during the review process that had not been discussed before the process started. This could have introduced biases (211). However, exclusion- and inclusion criteria were pre-specified before the classification began.

One benefit from our literature review is that we included keywords both as subject headings and as text words in PubMed or explosion searches and multipurpose terms in Embase, thereby broadening our search. It may take a few months before subject

headings are added to an article, which confirms the importance of including text words and multipurpose terms.

10.2 Data collection

The data collection for this thesis is anchored in the MicroCOPD study. In the MicroCOPD study, a self-developed questionnaire was used to interview participants on numerous important exposures (see **8.2.3.1 Metadata** and **16.1 Appendix A. Questionnaire, MicroCOPD study**). Data from this questionnaire was used in both participation and mycobiome analyses. In addition, biological material from the bronchoscopic sampling was processed and sequenced to be used in the mycobiome papers (paper III and IV).

10.2.1 Data collection for participation analyses (paper II)

We examined motivation to participate in the MicroCOPD study by an open question asked in the following way:

"Why did you wish to take part in this project?"

The question was asked some minutes before the procedure. The question is by no means exhaustive, especially considering that we coded the answers into categories at time of analysis. It would have been interesting to perform a profound psychological analysis of participation motives, for instance as a qualitative study. Such detailed analyses were outside the scope of the MicroCOPD study. Previous studies published at the time of data collection for the MicroCOPD study utilised different ways of examining participation motives in research bronchoscopy studies. Questionnaires with open and/or closed questions were used by three studies (86, 88, 92), while four studies used interviews either face-to-face or by telephone (87, 88, 90, 91). Only one study collected information by use of focus groups (88). The benefit with closed questions, i.e. pre-defined answer

options, is the ease in classification afterwards, but one might miss important nuances in answers. We made an open question to reveal such nuances. More information and follow-up questions could perhaps be gained from focus groups. However, as experienced by Kerrison et al., recruitment for such groups can be difficult (88). Addition of follow-up questions in our study could have strengthened results, especially since 8.2% of answers were missing. Some of these missing values were a result of too few details in the answer, such as "wanted to participate" or "was asked". It should be mentioned that different study personnel were involved in data collection, including participation motives. But all had been instructed to ask the question as written in the questionnaire. Participants in the MicroCOPD study were asked for participation motives in the operating room just prior to the procedure. Chudleigh et al. asked parents for motives for including their children in the study at two time points, first early in the study and then again late in the study (92). They observed that parents got less confident that participation was safe, and that they were less motivated by achieving a benefit for their child later in the study (92). This suggests that the timing of the question had an impact on the outcome. The measurement validity of the participation motive variable in the MicroCOPD study could thus been increased if we had asked the participants both at recruitment and at the procedure day. The bronchoscopic procedure was not mentioned in the question text, and one could argue that answers are not specific for bronchoscopy procedures. But the question was asked as part of a 13 pages long questionnaire (see **16.1 Appendix A. Questionnaire, MicroCOPD study**) focusing on the consecutive bronchoscopy, and it was asked immediately prior to the procedure. Thus, we expected that participants considered the bronchoscopy study for this particular question.

Data on non-response was collected from approached subjects that declined participation. Most of the non-responders declined at first approach, while some declined participation after some waiting time. Non-responders were asked directly by one of the study physicians what reason that prevented them from participation using an open question. Non-response reasons were categorised later. A total of 21.6% of initial non-responders

missed non-response reason in paper II, and there were significantly more missing in non-response reasons from the initial non-responders compared to late non-responders. Unfortunately, not all of the recruiting study physicians had been informed that we collected non-response reasons early in the recruitment. Most participants were scheduled for bronchoscopy at that time and few had been bronchoscoped. As it is likely that most late non-responders decline participation close up to their scheduled bronchoscopy, missing information is likely to impact initial non-responders most. It would have been interesting to know more about the non-responders. Unfortunately, due to privacy regulations, this was not possible. Non-responders were previous participants of BSCC and GeneCOPD, and some information was still available. Data were however limited to age, sex, and diagnosis.

10.2.2 Data collection for mycobiome analyses (paper III and IV)

The mycobiome papers are based on analyses of the OW and BAL samples, and NCS. In addition to biological samples, so-called sample metadata is collected in mycobiome studies. Metadata are "data about data" and are essential to gain biological insight from our data. Examples of common metadata are sex, age, diagnosis, and medication use. Metadata is thus specific to a given mycobiome study, and it was collected by use of a questionnaire for paper III and IV.

10.2.2.1 Collection of metadata

Participant category, or diagnosis, was an important variable in the metadata. Participants with COPD in the MicroCOPD study had a chronic airway obstruction (low FEV_1/FVC) in presence of respiratory symptoms. Additionally, 22 controls had a ratio of FEV_1/FVC lower than 0.7, i.e. airflow obstruction. Airflow obstruction detected by spirometry is required to establish a COPD diagnosis, but the individual in whom a COPD diagnosis is considered should also have relevant symptoms and a history of risk factor exposure according to the GOLD report (14). Consequently, it is justified that more than an

obstructive spirometry is needed to establish a COPD diagnosis. Airflow limitation could in fact be a result of advancing age (212-214). The 22 controls with airflow obstruction did not experience sufficient symptoms to be put in the COPD group. Still, these 22 obstructive controls were of particular interest, and were subject for some additional analyses (Table 6).

Table 6. Comparison of obstructive controls and participants with COPD

Variable	Obstructive controls, n=22	COPD, n=93	Comparison
Age, mean years (SD)	68.9 (7.1)	67.4 (7.6)	p-value = 0.4047
FEV_1, mean % of predicted (SD)	97.4 (13.1)	61.1 (17.3)	*p-value < 0.0001*
Pack years, mean (SD)	28.5 (25.8)	30.0 (17.9)	p-value = 0.7421
Smoking status (%)			
Daily	13 (59.1)	22 (23.7)	
Ex-smokers	6 (27.3)	70 (75.3)	*p-value < 0.01*
Never	3 (13.6)	1 (1.1)	

COPD, chronic obstructive pulmonary disease, SD, standard deviation, FEV_1, forced expiratory volume in 1 second. Age, FEV_1 in percentage of predicted, and pack years were tested with a t-test. Smoking status was tested with Fisher's exact test.

Obstructive controls had significantly higher FEV_1 values than participants with COPD. The obstructive controls had a pretty high smoking burden, but less than participants with COPD if we merge daily and ex-smokers. Albeit non-significant, obstructive controls were also older.

The lower limit of normal (LLN) is an alternative to the fixed FEV_1/FVC ratio cut-off. The LLN is taken to be equal to the 5th percentile of the frequency distribution in a reference population (215). The fixed ratio can increase false positive results in males aged above 40 years and females above 50 years, in addition to overdiagnosis of COPD in asymptomatic elderly never-smokers (215). Our definition of COPD relied on the fixed ratio, which thus could have led to overdiagnosis of COPD in our elderly population.

However, we also required that participants with a $FEV_1/FVC < 0.7$ had symptoms that corresponded to COPD to support a clinical diagnosis of COPD. Our research team included pulmonologists with long experience in COPD treatment, and all diagnoses were verified based on spirometry, radiologic imaging, respiratory symptoms, and disease history. We believe that a thorough individual examination in the MicroCOPD study thus prevented overdiagnosis of COPD. We did not perform BAL in participants with a measured FEV_1 below 30% of predicted and below 1.0 litres due to safety concerns. Participants without a BAL sample were excluded in the mycobiome papers (paper III and IV), thus participants with a very low FEV_1 are not included in these analyses. This limited comparison based on severity. We also excluded subjects that were currently on, or had recently had, a course of systemic antimicrobial treatment or systemic steroid treatment, and subjects with symptoms of an infection. By doing so, we were only able to investigate the stable state of the mycobiome.

Inclusion of a questionnaire might introduce biases, especially recall bias (see **10.1.1 Reliability, validity, and bias**). Participants were asked about current and previous smoking habits and previous antibiotics treatment. Precise smoking exposure and dates are difficult to remember, and variables were not objectively verified. If tobacco exposure was more vividly remembered in subjects with COPD, this could lead to an incorrect calculation of pack years. Information on previous antibiotic treatment was used to create the intercurrent antibiotic variable (i.e. antibiotic treatment between two sampling time points). It is possible that participants that frequently require antibiotic treatment, for instance participants with COPD with frequent exacerbations, might be more inclined to forget a treatment, and are thus subject to recall bias.

10.2.2.2 Collection of procedural samples

An ideal lung sample should be possible to collect safely. Essentially, this means without the use of an invasive procedure. Collecting lung samples without an invasive procedure is difficult of natural reasons, considering the lungs' anatomic location. Furthermore, the

researcher needs to be certain that the sample truly represents the lung environment and is not too much blurred by contamination. Finally, the sampling method should be achievable, both in terms of costs and feasibility. That means, it should not require too much of the participants, as we do not want to introduce sample differences based on participants' effort. Some common sampling methods are discussed below.

Sputum samples is a common way of sampling the lungs and can either be spontaneous or induced. IS samples are obtained by having subjects cough after inhalation of a hypertonic saline solution and then expel the sputum from the mouth into a container (69). There are several advantages of choosing sputum samples. They are easy to perform, non-invasive and thus possible to perform during exacerbations, and inexpensive. However, sputum samples are the most vulnerable to upper airway contamination due to its direct contact with microbes at the passage through the mouth. Furthermore, sputum samples proximal and distal airways in all parts of both lungs, in contrast to BAL and biopsies. It has indeed been shown that IS mycobiomes are distinct from the oral and BAL mycobiome (113). Different environmental conditions are proposed to exist in different portions of the lungs (42), and such regional differences will be indistinguishable in a sputum sample (69).

Bronchoscopic sampling is an alternative to sputum samples, and is usually performed by lavage, biopsies, or brushes. BAL samples from a subset of the alveoli (69), thereby providing researchers with better control of where obtained samples are collected. It also means that, in contrast to sputum samples, BAL might miss regional differences entirely due to its limited sampling area (69). Dickson et al. have proposed that BAL from a single segment is sufficient to characterise the healthy lung microbiome (42). But it has also been shown that different microbiomes have been found in different lung lobes of COPD patients (77), questioning the validity of including BAL samples from only one lobe in analyses of diseased lungs. The invasiveness and costs associated with a bronchoscopy are motives for not including BAL in the study design of a microbiome

study. Still, our research group has shown that bronchoscopy generally is a safe procedure with few serious complications (216, 217). It is associated with some discomfort, though. There are also no agreed standards for bronchoscopic sampling in microbiome studies. Whether the amount of returned BAL volume impact analyses remain to be examined. BAL is often performed fractionated. What total volume one should obtain, and whether analyses should be performed for fractions separately or pooled, are considerations that remain to agree on. A benefit of BAL however is the possibility to perform analyses of the local immune response such as cell counts, measures of inflammation markers, and more.

Mycobiome analyses in the MicroCOPD study (paper III and IV) were based on the second fraction of BAL from the RML. The entire MicroCOPD study included four different sample types to sample the lungs, namely PSB, BAL, SVL, and biopsies. We expected that sampling would have an impact on the surroundings. Thus, we chose different sampling sites for the different sample types to avoid any potential impact from earlier sampling, which explains why PSB was taken from the RLL and BAL was taken from RML. Dickson et al. have proposed that microaspiration is more likely to influence the right lung due to the vertically oriented right main bronchus (42). Additionally, if different mycobiomes are present in different parts of the lungs, similarly to what is proposed for bacteria in healthy state (42) or in disease (77), we might have decreased the measurement validity in the mycobiome analyses by use of RML only. We included BAL rather than sputum samples in the MicroCOPD study as we considered contamination from upper airways on BAL to be less. However, the risk of contamination in BAL samples is not negligible. In fact, contamination in mycobiome studies is an issue which requires special attention.

Contamination in a microbiome setting has been defined as “the observation of sequence reads in a sample coming from microbes that were not originally part of that sample” (218). Contamination could be introduced in different parts of a mycobiome study, for

instance during the procedure. Except for sampling of lung explants, sampling modalities of the lung are impacted from the upper airways and the oral cavity, either through direct contact with the sample (sputum samples) or the sampling device (bronchoscopic samples). Considering the low levels of fungi in the lungs (172), only minor disturbances from the high-biomass oral cavity could significantly impact lung mycobiome samples. In fact, it is assumed that bacteria vastly outnumber fungi on must human mucosal surfaces (59), making contamination particularly problematic for mycobiome studies. The extent of upper airway contamination on bronchoscopic samples is somewhat unresolved. Some claim that lung samples are confounded by bronchoscopic carryover (71), while others claim that there is only minor impact from pharyngeal contamination (70, 219). While no conclusion exists on the impact of upper airway contamination, our research group has shown that protected BAL, as used in the MicroCOPD study, minimises such contamination (9). Only the second fraction of the BAL was used in the mycobiome analyses in the MicroCOPD study (paper III and IV). The paper by our research group has also shown high similarity between the first and the second fraction of BAL, suggesting little impact from choice of fraction. Although, in theory, use of a second fraction could be a method to reduce impact from potential bronchoscopic carryover (71). All other COPD mycobiome studies have included sputum samples in their analyses (108, 113, 175), rendering them more vulnerable to oral contamination, although Cui et al. additionally collected BAL (113). The largest study collected samples as a combination of spontaneous and induced sputum (175). A bacterial microbiome study found different microbiota profiles using both spontaneous and induced sputum (220), meaning that unwanted variation might have been introduced in the mycobiome study by Tiew et al. (175).

As discussed above, lungs are exposed to microorganisms through microaspiration. Consequently, lung samples will have a similarity to oral samples (70, 71), and it remains a challenge to separate contamination from true exposures. We thus included an oral wash in the MicroCOPD study for representation of the upper airway microbiota.

Inclusion of proximal and distal sampling sites in the lung would probably have increased our possibility to examine the role of microaspiration in shaping the lung mycobiome, but only BAL was included in the mycobiome analyses. Nevertheless, we are not aware of other mycobiome studies comparing OW and protected BAL.

Except for contamination introduced by the procedure, mycobiome studies are also prone to laboratory contamination. The laboratory work and the associated contamination will be discussed in the following.

10.3 Laboratory work

10.3.1 Laboratory contamination

Studies have now emphasised that microbiome samples are being contaminated during laboratory processing steps (for instance from DNA extraction kits, PCR reagents, technician, or microbes in air) (67, 174). This especially applies to the low-biomass mycobiome studies, as greatest impact is expected to be on samples with low microbial loads (67). Furthermore, it has been suggested that that fungal DNA can often contaminate PCR reagents (221). Fungi is also found in both indoor dust samples and outdoor air (222) and might be a contamination source during preparation. Laboratory contamination runs the risk of introducing false positive in mycobiome studies and should not be overseen.

In silico methods are developed to handle laboratory contamination. One option is to remove sequences below a user-specified relative abundance threshold, employed by for instance Bittinger et al. (130). The rationale behind removing low-abundance ASVs is that these probably represent spurious sequences and not true biological signals (223). Still, rare ASVs could in fact appear in low abundances, and the method would remove these ASVs rather than abundant contaminants that are the most likely to interfere with subsequent analysis (173). Another option is to remove sequences found in NCS. NCS

are subject to index cross-talk from abundant true sequences found in other samples discussed more in detailed later (224). Removing true sequences, possibly with a high abundance, makes this decontamination option an unwanted pitfall. A third decontamination option would be to remove taxa commonly known to often appear as contamination. Contamination specific for a given study is however not removed by the common contaminant method (173).

In 2018, Davis et al. published a decontamination tool especially developed for marker-gene and metagenomic sequencing called *Decontam* (173). *Decontam* is based on two assumptions: 1) That sequences from contaminating taxa are likely to have frequencies that inversely correlate with sample DNA concentration, and 2) sequences from contaminating taxa are likely to have higher prevalence in control samples than in true samples (173). Both assumptions can be used to identify presumed contaminants, and we employed the second assumption to do a so-called prevalence-based contaminant identification in our mycobiome analyses (paper III and IV). The prevalence-based decontamination method is based on the prevalence of ASVs in NCS compared to actual samples, for instance BAL samples. The relationship between prevalence in NCS and actual samples creates a score for each taxon, which is later used in comparison to a user-defined threshold. A taxon is classified as a contaminant if the score is less than the threshold. The default threshold is 0.1 but should be set individually for each study. A higher threshold results in a more stringent contamination identification, meaning more taxa are classified as contaminants. The interpretation of a threshold of 0.5 is intuitively easy to interpret, because taxa would be classified as contaminants if present in a higher fraction of NCS than actual samples. The prevalence-based decontamination is feasible even in low-biomass environments (173), which suits our mycobiome lung samples well. The *Decontam* algorithm also includes an approach consisting of identifying non-contaminants in very low-biomass samples where samples are expected to constitute of a larger fraction of contamination than actual sequences. We chose the approach to identify contaminants, but we acknowledge that the non-contaminant approach perhaps would

have been more feasible. However, we observed several differences between our BAL and OW samples compared to the NCS including lower DNA yields, lower sequence reads, and differences in taxonomy, suggesting contaminant identification was appropriate.

As suggested by the *Decontam* authors, we did multiple comparisons before deciding on a threshold. Scores from *Decontam* were plotted for each comparison, and presumed contaminants were compared. Specifically, we tested with and without prior filtering of low abundance ASVs, with all sequencing runs together or separately, with OW and BAL together or OW and BAL separately, curation with *LULU* prior or after *Decontam*, and with threshold 0.1 or 0.5. Minor differences were seen from all parameters except for different thresholds. A threshold of 0.5 naturally identified more contaminants than 0.1. Our final model included a pre-filtering step of low abundance ASVs after *LULU* curation, and *Decontam* performed on each sample type separately with a 0.5 threshold. We did not run *Decontam* on each sequencing run independently. However, inherent bias is expected across sequencing runs, and thus include different contaminants. *Decontam* allows users to do a batched classification in which case scores are generated independently within each batch. We specified sequencing runs as batches in our mycobiome analyses. The *Decontam* algorithm has been confirmed to work well on the 16S rRNA data from the MicroCOPD study (174), but it has not been frequently employed by mycobiome studies. However, one recent study on the COPD mycobiome did also use *Decontam* with a prevalence threshold of 0.5 for contamination identification (175), adding some indication that *Decontam* could work well in mycobiome studies as well. Although it has been suggested that five to six NCS are sufficient to identify most contaminants in *Decontam* (173), sensitivity increases with more samples. A NCS was collected for each participant in the MicroCOPD study, creating a solid foundation for contaminant identification. In comparison, the other mycobiome study using *Decontam* only included 4 DNA extraction blanks and 10 sequencing blanks from a total of 437 participants with COPD and 47 healthy controls (175).

NCS were not made specifically for each step in the study, i.e. we did not include a NCS from the PBS, the DNA extraction kit, and the sequencing separately. Nevertheless, our PBS NCS were subject to the same laboratory protocol and sequencing as the procedural samples, and we therefore expect contamination from each step of the study to be included. In addition to be used in *Decontam*, our NCS were subject to detailed analyses. Taxonomy was assigned to all samples, and a Yue-Clayton measurement was calculated between the NCS and OW, and NCS and BAL from each participant. In addition, diversity and differential abundance analysis were performed on the different sample types. We observed lower read counts in NCS compared to OW and BAL in every part of the bioinformatic workflow, and a significantly higher Shannon index in NCS compared to both OW and BAL. Sample types clustered differently in Procrustes plots with M^2 above 0.85 for both comparisons, although not significant. The M^2 statistic increases with less concordance, and a M^2 above 0.3 is often interpreted as unsimilar. The mean value of Yue-Clayton measurements was 0.65 for OW and NCS sample pairs, and 0.73 for BAL and NCS. The Yue-Clayton measure is 0 with perfect similarity and 1 with perfect dissimilarity. There is however no settled threshold to define similarity/dissimilarity using Yue-Clayton measures, but previous studies have used 0.2 as cut-off for taxonomic differences (220, 225). Using this cut-off value, the results indicate dissimilarities between our OW and BAL samples compared to NCS. The observed differences in read counts, DNA yields, and diversity give additional evidence that the NCS mycobiome profiles differed from those made from OW and BAL.

During library preparation, DNA fragments are labelled with indexes so that sequences can be traced back to the sample they belonged to. Problems arise if sequences are assigned to an incorrect sample, commonly referred to as index misassignment or index cross-talk (224). Index cross-talk can be introduced by several mechanisms, including adapter errors from the manufacturer, experimental/sample handling issues, sequencing or bioinformatic errors, or carryover from previous runs on the same instrument (224). In order to reduce the impact from cross-talk, it is recommended to use dual index adapters

(226), and apply quality filtering of the index sequences (227). We did not apply any specific strategy to handle index cross-talk in our data except dual indexes, but we removed ASVs present in only one sample, and ASVs with less than 10 sequence reads across all samples, also for NCS. It is likely to assume that misassigned sequences constitute a minor fraction of a sample's total sequences. By including the most abundant taxa only, we thus expect impact from index cross-talk to be reduced, but recognise that it will not resolve the issue entirely. Other COPD mycobiome studies (108, 113, 175), however, have not commented on cross-talk contamination, suggesting index cross-talk might be an overlooked, albeit important, limitation of mycobiome studies. Index cross-talk could also contribute to contamination of NCS, potentially limiting their use in decontamination. Index cross-talk does not seem to limit contaminant identification using the prevalence method in *Decontam*, but *Decontam* is not designed to remove such cross-talk contamination (173).

Despite the inherent risk of contamination, laboratory work cannot be avoided, and the following sections will focus on the different parts.

10.3.2 Fungal DNA extraction and viability

The fungal cell wall is unique to the fungi, and is primarily composed of chitin, glucans, mannans, and glycoproteins (228). The components of the fungal cell wall create mechanical strength, and different combinations yields varying strengths (228). In contrast to fungi, bacterial cell walls consist of peptidoglycan (229). Due to the differences between bacterial and fungal cell walls, methodologies to extract bacterial DNA might not be suitable for fungal DNA extraction. Fungal cell walls are more difficult to open compared to the bacterial cell wall and might require chemical or mechanical lysis as a pre-extraction step (135). For instance, by employing a harsh bead breaking step, Dupuy et al. showed that *Malassezia* was a prominent community member of the oral mycobiome that had previously been overlooked (230). Use of mechanical

disintegration, however, runs the risk of shearing the DNA, in addition to unwanted release of DNA from other sources than fungi (69). A recent study comparing five protocols for assessing bacterial and fungal DNA recovery found that inclusion of a bead beating step increased the recovery of certain fungal taxa, including *Malasseziaceae*, *Aspergillaceae*, *Cladosporiaceae*, and *Dipodascaceae* (231). Bead-beating combined with pre-treatment consisting of magnetic stirring in PBS, and inclusion of Phenol:Chloroform:Isoamyl alcohol, was suggested to reduce the proportion of dead microorganisms. Nevertheless, whether observed taxa from amplicon marker studies are living and reproducing in the lungs remains a controversy. Sample pre-treatment with propidium monoazide (PMA) has been suggested to facilitate viable microorganism detection. No significant differences were observed between PMA-treated and untreated samples in a study by Nguyen et al. (133), but samples were collected from 5 CF patients with acute exacerbations, thus limiting generalisability.

The fungal DNA extraction in the MicroCOPD study was based on a protocol developed primarily for bacterial DNA extraction. However, our protocol included lysis both by enzymes and bead beating. We observed that *Malassezia*, which is well-known for its thick cell wall (230), was one of the most abundant fungi both in OW and BAL. Observations of *Malassezia* thus suggests that our included DNA extraction protocol was suited to open robust cell walls. We did not include any other quality control of the DNA extraction except quantifying DNA yield by Qubit. The DNA extraction protocols employed in the three other COPD mycobiome studies were not consistent across studies (108, 113, 175), nor to our study. Using different DNA extractions is just one of several factors limiting comparison across different mycobiome studies.

10.3.3 PCR

The PCR procedure can introduce biases in a mycobiome study, especially with increasing PCR cycles. In a paper by Salter et al., it was shown that by increasing the

number of PCR cycles from 20 to 40, a weaker signal from the sample microbiota and an increased signal from contamination was seen (67). In a study by our research group, however, it was shown that the impact from contamination did not vary much between a 30 PCR-cycle and a 45-PCR cycle setup (174). Our research group has recently examined the impact of a one vs two-step PCR protocol using bacterial data (232). The results from the study indicated an increased number of small ASVs when following a two-step PCR protocol.

We used a two-step PCR approach in our studies as recommended by the Illumina sequencing platform. Taking the one vs two-step study above into consideration, we realise the possibility of introducing a PCR related bias using this setup. The study also emphasised that *Decontam* did not resolve differences in results between sequencing setups (232). We do believe, though, that further investigations should be done to fully reveal the impact from different PCR protocols, in particular with fungal data included. We modified the Illumina 16S Metagenomic Sequencing Library Preparation guide (Part no. 15044223 Rev. B) by increasing the first PCR step from 25 to 28, and the second index PCR from 8 to 9 cycles. By increasing the first PCR step, i.e. the amplicon PCR, higher levels of DNA are provided for sequencing, which might be important in low biomass mycobiome studies. As discussed above, the literature is conflicting regarding the impact from increasing PCR cycles (67, 174). The percentage of increase in PCR cycles for our mycobiome data was nonetheless low compared to the setup in study reporting increased signal from contamination with more cycles (67).

10.3.4 Sequencing

The mycobiome papers in the current thesis relied on PCR-amplified gene fragments, or amplicons, with the help of primers with broad fungal specificity. In mycobiome studies, these primers are usually targeted against the gene encoding the 18S region, or the ITS regions located between the 18S and 26S rRNA genes (69). The 18S rRNA gene is more

conserved across eukaryotes, and consequently, any human DNA in samples will also be amplified (69). Use of the 18S approach has somewhat diminished in recent years, and the ITS has been suggested as the recommended universal barcoding target for fungi (117). No consensus seems to prevail whether ITS1 or ITS2 should be used, though (233-235). A drawback of ITS amplification is that different primer pairs are biased towards particular taxa (236). Furthermore, ITS regions of different fungi vary in length (59) and are present in variable copy numbers in fungal genomes (172). Length variation could introduce biases during PCR amplification and sequencing approaches (59), while the copy number variation hampers accurate quantitative analyses (172). Additionally, because it is a non-coding region, ITS sequences cannot be used to determine phylogenetic relationships between unidentified fungi (69). Lastly, species resolution is poor using the ITS1 marker gene (117, 172), and analyses are more reliable at genus level.

We chose to amplify the ITS1 region in our analyses. An earlier comparison between 18S, ITS1, ITS2, and large subunit rRNA found an overrepresentation of *C. albicans* for all targets using a mock community (233). For this comparison, primer pair ITS1F – ITS2 was used to amplify the ITS1 region. Our primer set, the ITS1-30F/ITS1-217R, has shown a further increase in the coverage of *Candida* compared to the ITS1F – ITS2 primer pair (154). Overestimation of *Candida* is of particular interest due to the *Candida* dominance seen in our studies. We did a pilot project prior to the sequencing to compare three different primer pairs for ITS1 amplification, namely ITS1-ITS2, ITS1-ITS4, and ITS1-30F/ITS1-217R. We included four of our real samples, and compared percentage of unknown reads, taxonomy, and diversity. The ITS1-30F/ITS1-217R primer pair came out as most precise and sensitive. This argues for our choice of the ITS1-30F/ITS1-217R primer pair. However, it does not justify why we chose ITS1, and not ITS2, as has been argued for in some papers published after our sequencing was done (233, 235, 237). In hindsight, amplifying both the ITS1 and ITS2 regions, and perhaps even doing shotgun sequencing, could have added additional information to our analyses. In reality, many

decisions in research, especially on study design, are the results of compromise between desired methods and available resources. Sequencing is expensive, which is one of the reasons mycobiome studies frequently sequence ITS1 or ITS2 alone. Amplification of ITS1 using ITS1F-ITS2 has failed to generate sequences for particular taxa, for instance *Yarrowia lipolytica* and *Malassezia* (233). Based on unambiguous BLASTN results in our results, we concluded that *Y. lipolytica* was represented in our unclassified sequences. The sequences account for a small part of the total count though but could have impacted presence/absence metrics like Jaccard. Additionally, several *Malassezia* ASVs were found.

We did not adjust for copy number variation, which could impact analyses based on abundance. It is suggested that researchers should not solely rely on read numbers to determine relative abundances (172), because high copy numbers result in overestimations. Furthermore, we did not include any phylogenetic diversity metrics, and we did all analyses except diversity analyses on genus level. Diversity analyses were done on ASV level because the diversity metrics used (Shannon, Bray-Curtis, and Jaccard) are independent of taxonomy.

10.3.5 Batch effects due to DNA extraction and sequencing

Batch effects can be defined as unwanted variation introduced by confounding factors that are not related to any factors of interest, and where the term batch itself refers to one of the levels of a particular confounding factor (238). Age and sex could for instance be a confounding variable in a microbiome study. In the current thesis, we were concerned about batch effects due to inherent biases expected from the use of different sequencing runs and different DNA extraction kit lots. We noticed that the number of ASVs after merging of the three sequencing runs was close to the sum of ASVs found in the separate runs, suggesting new, potentially spurious ASVs, to be introduced with each sequencing run. We expected a sum closer to the number of ASVs found in the sequencing run with

the highest number of ASVs. Thus, analyses of batch effects were done on our processed data. No differences in Shannon indexes were observed between sequencing runs. There seemed to be some clustering on sequencing run in PCoA plots, especially for BAL samples, but variance explained was low for the first and second axis. Pairwise PERMANOVA tests on differences in beta diversity between sequencing runs were hard to interpret due to low R-squared, contradictory results for Jaccard and Bray-Curtis metrics, and significant differences in dispersion. *ANCOM v2*, *MicrobiomeDDA*, and *ALDEx2* reported *Sarocladium* to be significantly different in abundance/distribution between sequencing run 1 and 2 for OW and BAL separately, but ideally no taxa should be different. Based on PCoA plots and PERMANOVA we do not believe batch effects were introduced from the different DNA extraction kit lots. However, these analyses were somewhat limited due to many missing values. We missed information on extraction kit used for 114 of 230 samples included for diversity analysis after rarefaction. To summarise, no apparent batch effect was seen due to the inclusion of multiple sequencing runs and different DNA extraction kit lots, but could not be entirely ruled out, particularly for sequencing run 1 and 2 for BAL samples. Groups of interest for the analyses in the current thesis, i.e. sample types and participant groups, were present in each sequencing run, and samples from a given participant were always sequenced in the same run. Due to no obvious batch effect, and a robust sequencing design, no further correction was done to adjust for batch effects from sequencing runs or DNA extraction kit lots.

We included a thorough examination of diversity and differential abundance/distribution testing to look for potential confounding effects from several important clinical parameters. No apparent effects were seen on potential confounding effects from several important clinical parameters like sex, age, FEV_1 in percentage of predicted, smoking, or ICS use.

10.4 Bioinformatics

A mycobiome study generates a large amount of data that needs to be processed bioinformatically prior to analyses. Different software packages have been developed, but most are directed towards bacterial data. That means, applying default settings of common software packages on fungal data could lead to errors, especially regarding classification at species level (172). Lack of a standard processing method leaves researchers to freely choose between an excess of bioinformatical tools with multiple parameter settings. A bioinformatic workflow of our mycobiome analyses is included in the Material and methods section (see Figure 4 in **8.2.5.1 Upstream bioinformatic analysis**), summarising which tools that were included.

ITSxpress is a bioinformatical tool developed for marker gene studies using ITS (157). The two ITS regions are flanked by conserved regions and separated by one intercalary conserved region (154). It is shown that removal of the conserved regions improves taxonomic classification (239). *ITSxpress* trims the FASTQ reads to filter out conserved parts. We chose to cluster at 100% identity in the *ITSxpress* software command, as this parameter is chosen by the author of *ITSxpress* in the *QIIME 2* tutorial on *ITSxpress* (240). Because of the length variation found in the ITS regions, another issue called read-through might happen if the read length is longer than the ITS region. The read will then extend into the opposite primer, leaving the reverse complement of the primers in the sequencing read. As the priming regions are often found in the conserved regions, we anticipated read-through to be handled by *ITSxpress*.

Errors are frequent from amplicon sequencing, and it is difficult to distinguish such errors from real biological differences. Sequencing errors generally increase towards the read end, especially in reverse reads (241), and are often evaluated by quality scores (242). PCR errors, such as chimeras (see **8.2.5.1 Upstream bioinformatic analysis**), are also common. Quality control methods are available to correct errors. *DADA2* is one of the

amplicon-specific quality control methods (173). It was previously agreement between microbiome researchers that sequences should be clustered together based on their similarity, usually on a 97% similarity threshold. These clusters were called OTUs. OTUs reduced the rate at which errors were misinterpreted as biological variation but ignored fine-scale variation (173). In contrast, *DADA2* identifies such fine-scale variations to create ASVs, resolving differences of as little as one nucleotide. *DADA2* calculates a probability that the abundance of reads of a given sequence could be explained by errors in the sequencing, or if there are more than could be explained by errors. We applied some specific parameter options to fit fungal data. Specifically, we skipped trimming of the start of the read (primers) and truncation. ITS length variability is biological, and truncation could remove real ITS variants that were shorter than the chosen truncation length (159). However, the *QIIME 2* ITS fungal analysis tutorial applies truncation of the sequencing read ends due to the decreased sequencing quality mentioned above (243). It was not obvious to us which solution we should choose. However, our quality scores were high, with median values above 30 at base position 250 for all forward reads and two of three reverse reads. The quality score of the reverse read in sequencing run 3 was somewhat lower, with a median quality score of 22 at the 250 base position, meaning a base call accuracy of 99% (242). We thus chose to abstain from truncation of our reads as recommended by the *DADA2* ITS tutorial. It has been suggested that *DADA2* inflates alpha diversity, both by outputting rare organisms, but also false positives (244). Filtering of low abundance ASVs, as done in the current thesis, can reduce the number of unmatched ASVs generated by *DADA2* (244). An inflation would impact metrics not taking abundance of ASVs into account, for instance the Jaccard similarity coefficient, which could explain observed differences between the Bray-Curtis and Jaccard metrics in our data.

The result of the *DADA2* process is an ASV table, a table of the number of times each ASV appears in each sample. A mycobiome study should compare abundances of ASVs between groups of interest to look for differential abundant ASV, i.e. ASVs that differ

significantly in abundance. Abundance data in an ASV table is, however, not directly comparable. Each sample has its own sequencing depth or sum of abundances. Imagine that a theoretical sample A has a *Candida* abundance of 200 reads and a sequencing depth of 250, while another sample B has 14,000 *Candida* reads and a sequencing depth of 70,000. *Candida* read count is higher in sample B. But the *proportion* of *Candida* reads is lower (relative abundances are 80% in sample A and 20% in sample B). Abundance tests are thus tests of proportions. Since proportions always add to one, a change in a taxon's relative abundance will change the relative abundances of the remaining taxa. In other words, microbiome datasets are compositional (245). This complicates analysis because we are unable to deduce absolute changes of taxa in a given sample without quantitative information about total microbial load (246). Several techniques have been proposed to deal with this compositionality, including rarefaction (subsampling without replacement to a given sampling depth), median, and quantile normalization, but methods lack sufficient control of false discovery rates (246). Methods specifically developed for differential abundance testing has shown high false positive rates (245). We included three differential abundance tests with distinct foundations. *ANCOM v2* uses Aitchison's log-ratio based methodology (191), *ALDEx2* applies the centred log-ratio values from a modelled probability distribution of the dataset (192-194), while *MicrobiomeDDA* is built on a zero-inflated negative binomial regression model (190). We expected that inclusion of multiple tests would strengthen the reliability of the results, especially since we only considered taxa found by all three tests as a positive result. But we do admit that it can have the opposite effect, creating confusion only. The three methods test for differential abundance of taxa on group level. We calculated the Yue-Clayton measurement between sample types in paper III, and between the bronchoscopy procedures in paper IV. The Yue-Clayton measurement is a calculation based on differences in proportions for all taxa between two samples of interest. We could therefore look for differences for each participant specifically. Another problem in abundance testing is the before mentioned copy number variation (see **10.3.4**

Sequencing) found in the ITS regions. Differences in copy number variation could have impacted our differential abundance analyses, as we did not account for this.

Comparison of diversity is also an important part of a mycobiome study. Creation of diversity metrics is not straightforward due to the differences in sequencing depth across different samples. As a normalisation method, we chose to rarefy our ASV table to a specified depth. That means, all sample counts will be randomly subsampled to the given depth. Samples with a total read count lower than this depth will be excluded. We examined our ASV table to find a suitable depth, which should be as high as possible. A high value keeps most of the information, but will also result in a bigger loss of samples. In contrast, many of the sample may have no shared ASVs at a low depth, which would result in a distance of 1.0 with Jaccard and Bray-Curtis. The rarefaction process has been criticised due to loss of information and reproducibility (247), and other options like additive-, centre-, or isometric log-ratio transform have been suggested (245). Our decision was motivated by one of *QIIME 2*'s learning tutorials (248). *QIIME 2* (155) is one of the main pipelines used to perform microbiome analyses, and their diversity metric plugin or command has a required rarefaction step. Only the non-phylogenetic beta diversity metrics Bray-Curtis dissimilarity and Jaccard similarity coefficient were included in our analyses. We did not include any phylogenetic diversity metrics, because ITS sequences cannot be used to determine phylogenetic relationships between unidentified fungi (69). Although not useful, ITS sequences have been used in analyses including phylogenetical information in a previous COPD mycobiome study (108). We admit that important information is lost without phylogeny, but believe it is a strength that both a qualitative (Jaccard) and a quantitative (Bray-Curtis) distance metric was included or beta diversity analyses. That means, Jaccard values rare taxa equally as the high-abundant taxa, while Bray-Curtis values the high-abundant most. Including two such different metrics made it possible to see how the *Candida* dominance manifested in different methods.

10.5 Mycobiome specific issues

10.5.1 Mock communities and positive controls

It is generally recommended to include positive controls and a mock community in any mycobiome study (47, 235). Positive controls are samples with known species unlikely to be found in the samples, and aid in the assessment of contamination, error accumulation, chimera formation, and cross-contamination (235). A fungal mock community is a community with a predefined content of fungi. By using a mock community, researchers are able to assess their methodologies by comparing their results from the mock community with the known content. Use of mock communities is important to compare results between studies performed in different laboratories (47). Our studies are somewhat limited by the absence of positive controls and mock communities. In hindsight, this is a shortcoming of our study, that also affects a number of other studies, and one can only hope that future investigators will take this into account.

10.5.2 Taxonomy assignment

Fungal databases are used in mycobiome studies to assign taxonomy, but the databases have some limitations that affect the assignments. In contrast to bacterial databases, fungal databases include few sequences (58, 69), and sequences included in the databases are often incomplete and/or incorrect (59, 118, 172, 249). Additionally, databases are curated from gastrointestinal or environmental sources perhaps making them less relevant to the lung mycobiome (250). Fungal taxonomic assignment also suffers from a dual naming system, in which sexual and asexual forms of a fungal species can be classified as different taxa (58, 59, 69, 118, 135). Several fungal databases exist, including UNITE (65, 66), SILVA (64), and the International Society of Human and Animal Mycology (ISHAM)-ITS reference DNA barcoding database (251). The UNITE database was created to overcome the issue of misidentification of fungi. Occasionally, new, curated versions of the UNITE database is released. We chose the UNITE database for taxonomic assignment in the current thesis, as has been recently recommended (235). The sequences

within the UNITE database are heavily biased toward the *Dikarya* sub-kingdom, which includes the *Ascomycota* and *Basidiomycota* phyla (69, 172). This could explain why nearly all of our sequences were assigned to *Ascomycota* or *Basidiomycota*, but that does not mean that the assignment is incorrect. Other authors have suggested to do a further manual investigation after the taxonomic assignment, especially of the most abundant ASVs (59, 235). We did a thorough investigation of unclassified ASVs, i.e. ASVs only assigned to *Fungi* at kingdom level (see **8.2.5.1 Upstream bioinformatic analysis**). A total of 463 ASVs were excluded as they were likely to be non-fungal, while 90 ASVs were included for further analyses, often with an improved taxonomy. Considering that 463 ASVs could have been included as unclassified fungi when they most likely did not represent fungi, our investigation of unclassified ASVs should be viewed as a strength. However, due to time constraints, no ASVs with a more detailed taxonomy than *Fungi* at kingdom level were further investigated. The Nucleotide database is a collection of sequences from several sources, including GenBank, RefSeq, TPA and PDB (164). Even though some of these sources are quality-controlled, researchers can still freely submit their sequences to the database, possibly leading to incorrect or poorly-defined species names (59). It has previously been shown that as many as 20% of the publicly-available fungal ITS sequences are annotated incorrectly at the species level (252). We acknowledge the limitation of the Nucleotide database. The investigation was also limited by subjectivity. There exists, however, data in the Nucleotide database output that could have been used to reduce subjectivity. We could, for instance, have required that a hit needed a certain percent identity and coverage in order to being considered further. Furthermore, it is known that for some taxa, for instance *Aspergillus* spp., use of one target region is not sufficient for species identification (253). Adding another barcode, for instance betatubulin, could provide improved resolution, but was not done in the current thesis. To compensate for issues related to taxonomic assignment, analyses involving taxonomy were performed on genus level. This conforms with recommendations saying species level is not a suitable discrimination level for ITS1 data (117). However, important species-to-species differences are then lost.

10.5.3 Statistics

A sufficient sample size is needed for any statistical test to be valid. The MicroCOPD study included more participants than previous similar studies. Still, some issues with sample size were encountered during time of analysis. As discussed earlier, data was rarefied, i.e. subsampled without replacement, to a given sequencing depth in order to assess diversity independent of sampling depth. Samples with a sequencing depth below the chosen rarefaction value were discarded. That means, rarefaction normalise data at the expense of power loss, a sacrifice that has been criticised in the literature (247). Diversity analyses on smoking habits in BAL samples from controls were omitted due to a lack of current smokers in the descriptive mycobiome paper included in this thesis (paper III). Rarefaction also omitted most BAL samples in the longitudinal paper (paper IV), leaving only 12 left for analyses and decreasing power in the diversity analyses including BAL samples. The low number of samples precludes stratification on participant group in alpha and beta diversity analyses in the longitudinal mycobiome paper. However, we did not see any difference between controls and COPD in the descriptive study, meaning that stratification might not could have added additional information. The statistical analysis of Shannon index between those receiving intercurrent antibiotics and those who did not, was not stratified on sample type, even though different colouring was applied in the plot. We know that OW and BAL are two quite different sampling methods, but again, the descriptive paper did not show any difference in Shannon between the two methods. Furthermore, one of the included differential abundance tests, *ALDEx2*, works poorly if there are only a small number of taxa (less than about 50). Analyses to look for differential abundant features in OW compared to BAL and vice versa using *ALDEx2* were not performed in smoking controls and smoking participants with COPD, and participants with COPD not using ICS regularly. No comparisons were done between sequencing run 1 and 3 in OW and BAL due to few taxa.

The rDNA is dynamic and can exhibit substantial interspecific and intraspecific variation in copy number (254). In fungi, the locus is typically duplicated 100-200 times (118). The variation in copy numbers complicates any conclusions on quantitative comparisons and will influence taxonomic bar plots and comparisons based on quantities, for instance the Bray-Curtis metric and differential abundance tests. Suitable solutions to adjust for copy number variations are yet to be developed, and we thus need to interpret our data in light of possible interspecific and intraspecific responses (254). However, this issue is not specific to our study. Differential abundance testing was discussed in general earlier (see **10.4 Bioinformatics**). Hypothesis testing of microbiome compositional data is an ongoing research area without standardisation, exemplified well by the three COPD mycobiome studies performed (108, 113, 175). Cui et al. used the neutral model and ubiquity-ubiquity plots to look for disproportionately abundant fungi in different sample types (113). The neutral model predicts if the composition in the lung is a result of dispersal from the oral cavity or organisms adapted for growth in the lungs (72). Further, they used two machine-learning classifiers and an R-based open-source software called *Metastats* to identify any species associated with COPD (113). Su et al. did not include any statistical test to look for differences, but relative abundances could be examined visually (108). The multi-centre study by Tiew et al. utilised Linear discriminant analysis effect size (*LEfSe*) to look for dominating fungi in different geographic locations and between participants with different COPD exacerbation status (175). It is worth noting that *LEfSe* does not perform multiple testing correction. Multiple testing correction is offered by *MicrobiomeDDA*, which was included in our study. It is a demanding work to navigate in the differential abundance testing field. With the wealth of available options with different foundations, generalisability across studies is also limited. Furthermore, differential abundance tests are usually tested on bacterial data, and further testing should be done to evaluate their performances on fungal data. Mycobiome studies would certainly have benefitted from a standardisation in abundance testing.

Samples that were taken from the same participant are inevitably dependent and should be analysed with paired tests. Thus, analyses between OW, BAL, and NCS, and analyses between samples from first and second bronchoscopies should be paired. The *MicrobiomeDDA* test did not have any option to specify a paired design and was therefore not used in analyses on sample type. We did not include differential abundance tests in the longitudinal study because we suspected such tests to look for differences on group level rather than subject level. Yue-Clayton measures for each sample pair was used instead. We did not find any suitable paired test for comparison of beta diversity differences between OW and BAL nor between two consecutive bronchoscopies, so we presented the data in PCoA plots both with and without a Procrustes transformation. A Procrustes transformation gave us a measure of fit (M^2) and a p-value as a measure of concordance between the PCoA plots. Tiew et al. analysed temporal changes in beta diversity using PERMANOVA, while statistical method to analyse changes in alpha diversity by time was not given (175). We are not aware of any way to define a paired test using PERMANOVA, suggesting alternative methods should be used by Tiew in the temporal beta diversity analyses. To our knowledge, there is not agreement on what ordination method fits microbiome data best. Several methods exist, each with its unique features (255). We chose to visualise our multidimensional data using PCoA, mainly influenced by *QIIME 2*'s learning tutorial (248). We did, however, tried nonmetric multidimensional scaling, but did not reach any solution using the *metaMDS* function in the *vegan R* package (188). PCoA plots do seem widely used in the COPD mycobiome literature as well (108, 113, 175). Ultimately, plots are visual presentations, and choosing the appropriate statistical method seems at least quite as important.

10.5.4 Reproducibility in mycobiome research

Bittinger et al. has reported that the reproducibility in mycobiome studies is poor by performing repeated extractions from samples (130). A low reproducibility would have an impact on our longitudinal analysis in particular, because we need acceptable

reproducibility to assess the stability between two time points. It should be mentioned that the reproducibility analyses in Bittinger et al.'s paper was based on 18 samples only, but did include up to four repeated extractions. Furthermore, the reproducibility seemed to improve after conversion to PicoGreen-corrected abundance. Fungal OTUs with proportions of 1% to 50% appeared sporadically between replicates (130). Removal of low-abundance ASVs in the current thesis might have reduced noise from spurious ASVs, and together with contamination removal, this have probably increased the reproducibility. Still, examinations of repeated extractions of selected samples, preferably from each sample type and study group, would have strengthened our study results.

To build on the ideas of reproducibility, generalisation and comparison with other studies is worth mentioning. Technical variation is introduced in mycobiome research by different sampling techniques, storage, PCR amplification, and DNA sequencing (256). Without a standardised workflow, results can only be interpreted in each specific study setting, preventing generalisation. There is thus a need for standardisation and open research. An essential requirement in this regard, is the sharing of data. However, it has been shown that differences in bioinformatics and statistics could affect reproducibility, even with open access to the original raw data (256). Work has now been initiated to develop standards in the microbiome field by creation of reference reagents (257). Generalisation of the current thesis is also limited by the heterogeneous nature of the COPD disease (212), and more research is needed to examine the mycobiome in different COPD phenotypes. Moreover, data from different geographical areas should be collected, for instance in Africa where development of COPD due to indoor open fires is more common. But, as emphasised, comparisons are of limited value unless we can agree on a common workflow.

11 Discussion of main results

11.1 Participation in research bronchoscopy studies

Paper I was a literature review of participation in studies involving a bronchoscopy. Only seven papers were included in the final review, of which none included detailed analyses of patients with COPD. The studies were heterogeneous with respect to geography, aims, and disease in focus, which precludes generalisation of the results to a COPD population.

Still, some inferences from the literature are discussed in the following sub-sections and compared with our analyses on participation in the MicroCOPD study.

11.1.1 Response rates in research bronchoscopy studies

Response rates were given or derived in five of the included paper from the literature review. The two studies without response rates included Malawian adults undergoing research bronchoscopy (87), and smoking patients with COPD approached for participation in a lung cancer screening trial (91). The remaining studies included quite heterogeneous study populations. One studied HIV-infected individuals (86), while two studies included individuals who participated in, or were asked to participate in lung cancer chemoprevention trials (89, 90). The fourth examined infants with and without CF, and the last bronchoscopy study with a known response rate included patients with suspected lung cancer (88). Despite differences in study populations, response rates did not vary considerably, ranging from 64% to 73% in four of the five papers (86, 88, 90, 92). In contrast, the response rate in one of the lung cancer chemoprevention trials was markedly lower, with only 3% of pre-screened individuals being enrolled (89). This lung cancer chemoprevention trial had numerous specific exclusion criteria such as a minimal requirement of 30 pack years and no pre-existing medical condition, and subjects had to pass both a pre-screening and a screening in order to participate. Only 7.2% of the approached subjects passed the pre-screening, and received the consent form and were invited for a second in-person screening including both a CT and a bronchoscopy.

Subjects without exclusionary findings on the in-person screening were enrolled in the study, which constituted only 3.1% of the originally pre-screened subjects. Unfortunately, we could not find any information on the number of subjects that declined participation of their own will. Thus, the 3.1% response rate probably mostly reflects exclusion by the research team rather than denial to a specific invitation. Furthermore, several of the recruitment methods relied on the participants to contact the research team for a pre-screening by telephone. Strictly speaking, all subjects who read or heard about the study could thus be defined as invited, and this number is not possible to obtain. The discussed lung cancer chemoprevention trial was the only study examining the optimal recruitment strategy. They found that radio advertisement was the most effective strategy in terms of generated inquiries (89). The costs of each recruitment strategy utilised in the study were also evaluated, but not accounted for in the effectiveness calculation. The paper was published in 2009, and it would be interesting to repeat the study to examine the potential that lies in social media.

We chose to exclude participants that were recruited to the MicroCOPD study from our outpatient clinic and participants that contacted us by their own initiative. Final response rate for the MicroCOPD study was 50.9% (paper II). The response rate is in line with the results from the literature review, despite being somewhat lower. By achieving a satisfactory response rate, we minimise the risk of type II errors, i.e. an erroneous acceptance of the null hypothesis, and we can generalise our results to a greater extent. The highest response rate observed in the literature review (paper I) was 73% in a study where they investigated the effect on smoking cessation by participation in a chemoprevention trial for premalignant lesions (90). No response rate was explicitly given in the study, but we derived the response rate from a flow chart in the paper. The response rate was calculated in line with the definition in our literature review as number of enrolled divided by approached or pre-screened individuals. A total of 201 subjects were pre-screened, and 146 were analysed to examine smoking cessation, giving the response rate of 73%. However, pre-screened individuals had already accepted

participation in the chemoprevention trial, possibly introducing a selection bias. Furthermore, a total of 47 subjects were excluded from the smoking cessation trial due to lack of participation in the chemoprevention trial, being non-smokers, or not answering, leaving only 8 subjects as "true" decliners. Moreover, one could argue that this included response rate is not displaying a bronchoscopy study since the calculated response rate points to the smoking cessation trial and not the chemoprevention trial, in which a bronchoscopy was included. Nevertheless, the calculation example from the smoking cessation trial illustrates the difficulties in comparing response rates in studies not designed with that purpose in mind.

In paper II we reported a significantly higher response rate in men compared to women (44.8% and 56.5%, $p = 0.01$). No difference was observed with regard to age. Most literature on participation in clinical trials stem from cancer research, which has shown that women and elderly are underrepresented in cancer trials (258, 259), and women are also under-enrolled in heart failure research in the United States (260). The situation was somewhat different in seven Norwegian respiratory healthy surveys from 1965 to 1999, in which response rates were higher in women than in men and higher in the middle-aged/elderly than in young adults (84). None of the Norwegian studies included a procedure as invasive as bronchoscopy, and one could speculate that men viewed bronchoscopy as more beneficial compared to perceived benefit from non-invasive studies, thereby explaining the differences in response rates between women and men in the MicroCOPD study.

Another study from the literature review (paper I) showed that recruitment of infants with CF was more feasible than recruitment of healthy controls (92), suggesting higher response rates among subjects affected by the index disease. No differences were seen in response rates for the different participant categories in the MicroCOPD study (paper II), although 70.4% of subjects that declined participation after some consideration time had COPD or asthma. The Norwegian health system is highly ranked by the World Health

Organization (261). That subjects with COPD or asthma were as likely to participate as controls could perhaps be due to close monitoring of patients with obstructive lung disease (OLD) in Norway. In addition, health services are almost without charge in Norway, and together with the close monitoring, participation in a clinical study could thus be seen as no additional benefit. However, both responders with and without disease were driven by personal health benefit in the MicroCOPD study.

11.1.2 Participation motives in research bronchoscopy studies

Both the literature review (paper I) and the original analyses from the MicroCOPD study (paper II) concluded that personal benefit, altruism, and obligation were the most frequent participation motives in research bronchoscopy studies. Personal benefit was listed as a participation motive in all of the papers examining participation motives in the literature review (86-88, 90-92) and a later study (262), and it was most often related to participant's own health. In the MicroCOPD study, the personal benefit motive was a combined label from several unique participation motives merged together, especially personal health benefit. The consent form stated that no direct benefit was to be gained from participation, but still almost half of the participants gave personal health benefit as a participation motive. One could speculate that the consent form was not read properly by all participants, or that the consent form was too advanced to understand for any without a medical background, although it was written in an easy language for optimal understanding.

Similar to bronchoscopy study participants, patients undergoing bronchoscopy for clinical purposes also require information, often given in the form of a patient information leaflet (PIL). One study examined the readability and content of PILs provided to patients prior to gastrointestinal and respiratory endoscopic procedures in Irish public hospitals (263). The readability was evaluated using the Flesch Reading Ease and the Flesch–Kincaid Grade Level scores. The Flesch Reading Ease and the Flesch–

Kincaid Grade Level scores are validated tools for measuring readability (264). The Flesch Reading Ease creates a score which should be above 60 if readability should be acceptable, while the Flesch–Kincaid Grade Level score corresponds to the number of years of education usually required to understand the material, and patient information should be aimed at 6th grade (10–11 years old) level (263). The authors found that no department produced PILs that all met the recommended standard (Reading Ease scores of 60 or more and Grade Level of 6 or less) (263). Although a research setting is difficult to compare to a clinical situation, the Irish study demonstrates that medical information can be hard to deliver in a feasible manner. Generally speaking, regardless of how well-written the consent form is, it is the researcher's responsibility that participation is informed and that the participants have understood the consequences of participation. Researchers should put aside their own desires and not romanticise participation when approaching potential participants. As discussed above, the MicroCOPD consent form stated that no direct benefit was to be gained from participation. Nevertheless, the same consent form stated that all participants would get a proper examination of their lungs including a CT scan, which, understandably, might be perceived as a benefit. All subjects examined by CT scans or bronchoscopy, are potentially subject to unexpected findings. A specialist in the field of cardiology or pulmonology examined all results from a clinical perspective, and appropriate further diagnostics or treatment were given; for example, in the instance of finding significant arteriosclerosis in the cardiac vessels or a suspected malignancy on pulmonary CT scan or during bronchoscopy. Additionally, incidentally found deviations in blood sampling could be dealt with. Our research group has shown that research bronchoscopies have few serious complications, but are associated with discomfort such as post-procedural sore throat and fever (216, 217). Summarised, with few complications and potential advantages from examination, it is possible to argue that some benefits are achieved through participation. However, possible discomfort associated with participation should perhaps be more emphasised in the consent form.

Altruism was the other main participation motive in the MicroCOPD study (paper II), especially for women. Helping others, which could be regarded as the genuine altruism, was only reported by 15.9% of the participants, but contribution to science and continuation of previous participation added to a broader category of altruism. Altruism was also mentioned in three of the papers (86, 91, 92) from the literature review (paper I) and among healthy people in a later study on pulmonary tuberculosis in Malawi (262), and it was often accompanied by self-interest in an elderly population (91).

The last main participation motive for research bronchoscopy studies is a feeling of obligation (paper I and II). A study found that HIV-positive individuals were motivated by being asked by a physician or that the physician seemed to want them to participate (86). Obligation was more rarely mentioned as a participation motive in the MicroCOPD study. Nevertheless, physicians should be aware that their authority could affect potential participants' choices regarding study participation.

11.1.3 Non-response reasons in research bronchoscopy studies

The invasiveness of a bronchoscopy is the most common reason to decline participation in research bronchoscopy studies (86, 89, 91), and was also the most commonly reported non-response reason in the MicroCOPD study. Invitees feared the discomfort associated with the procedure (86, 89, 91), while some felt participation increased anxiety or perception of risks or complications (92). Several studies have examined patients' perceptions of bronchoscopy and shown that bronchoscopy is associated with discomfort (265-276). Furthermore, studies have reported that bronchoscopy is associated with anxiety (262, 266-268, 272, 276-278), albeit considered unjustified after the procedure (272). Interestingly, findings from the literature indicate that women experience more anxiety before a bronchoscopy (277), which again could explain their lower response rate observed in the MicroCOPD study (see **11.1.1 Response rates in research bronchoscopy studies**). A study by Poi et al. suggested that a more detailed explanation

of what sensations patients should expect to experience during a bronchoscopy might reduce some of the common fears (277), while another emphasised that full disclosure of risks was appreciated among participants (262). This contrasts somewhat to the results in a study by Uzbeck et al., showing that a more detailed pre-procedural risk disclosure given to subjects undergoing bronchoscopy increased anxiety compared to subjects receiving a simple risk information (279). It should be noted that the Uzbeck et al. studied pre-procedural *risk* disclosure, while the study by Poi et al. examined how further details regarding the procedure, for instance expected events or sensations, could alleviate anxiety. Nonetheless, it is expected that study participants are fully informed. Researchers should anyway consider how risk disclosure increases anxiety, and perhaps set aside sufficient time to discuss and relieve fears that potential participants might have, for instance at the time of consent signing. Additionally, each individual has its own preferences and understanding of the information given, which ideally should be accounted for. This is difficult in research with strict legal responsibilities and standardised consent forms and is perhaps easier to consider in the clinical setting.

We observed that subjects that declined participation in the MicroCOPD study after some consideration time more often stated worries/fear compared to subjects declining participation at first encounter (paper II). This suggests that waiting time increases fear of participation and should thus tried to be avoided. Participants with OLD more often reported that disease/health issues prevented them from participation compared to controls, which is not surprisingly considering that OLD are chronic diseases with respiratory symptoms.

How well a bronchoscopy is tolerated, is most often measured by examining how many would accept an additional bronchoscopy in the future, i.e. willingness to return. Despite being an uncomfortable procedure, most patients tolerate a bronchoscopy well (86, 87, 91, 265, 267, 270-274, 276, 280-282). Still, the range in acceptance rates for a future bronchoscopy is fairly wide, ranging from 13% (269) to 100% if probable returners were

included (86, 87, 271). The study with a willingness to return rate of 13% did not add "if necessary" in their willingness to return question and did not offer a sedative, which might explain some of the difference. Being female has been shown to be associated with higher reluctancy (265), while men has shown higher willingness to return for a repeat bronchoscopy (273, 282) and better tolerance in terms of satisfaction and acceptance of discomfort (278). This conforms to the lower response rates among women in research bronchoscopy studies discussed previously. Experienced discomfort has been associated with less patient satisfaction (265, 273, 274, 276). It has been suggested that reluctant subjects recollected more of the procedure (265), and that there is less reluctance to attend a future bronchoscopy among consciously sedated patients (270). Furthermore, others have found relatively low willingness to return among non-sedated patients undergoing bronchoscopy (283), and better tolerance in consciously sedated patients and patients with less pre-procedural anxiety (275, 276, 278). One could speculate that adding sedatives in a research bronchoscopy study design might increase the probability of risks, but analyses on complications and discomfort in the MicroCOPD study has shown that alfentanil reduced overall need for unplanned intervention or early termination of bronchoscopy (216). However, ten cases of drug-induced complications were also seen (216). Measuring tolerance by willingness to return has been criticised since several factors influence the accept of an additional procedure such as a desire for a diagnosis (268). Tolerance is also dependent on when the participant completed the questionnaire, for instance immediately after the procedure, or a few days later when the effect of the sedatives are passed. That bronchoscopy is generally well tolerated is important for researchers planning a study involving bronchoscopy. Knowing that most decline participation due to fear, it seems wise to inform potential participants of the findings regarding tolerance, possibly relieving some of the fear. Additionally, high willingness to return suggests that participation in a study with bronchoscopy does not negatively impact a decision of undergoing a future bronchoscopy if medically indicated, as has been discussed in the literature previously (86).

11.2 The pulmonary mycobiome

The COPD mycobiome is not well studied, and only three studies have used next generation sequencing to study the pulmonary mycobiome in COPD (108, 113, 175). Results from the MicroCOPD study has added knowledge to an understudied research area. To our knowledge, no studies have used next generation sequencing to examine the stability of the lung mycobiome in a COPD population or subjects without lung disease. In the following sections, we will discuss descriptive and longitudinal results from analyses on the pulmonary mycobiome from participants in the MicroCOPD study.

11.2.1 Descriptive analyses on the pulmonary mycobiome

11.2.1.1 Differences between sample types

In our analyses of the mycobiome, we included both OW, BAL, and NCS samples. The OW samples were analysed to account for potential contamination from bronchoscopic carryover. Both the OW and BAL samples in the MicroCOPD study were dominated by *Candida*, and there was significantly more *Candida* in the OW samples compared to BAL samples for both participant categories. OW and BAL samples also differed in Yue-Clayton measures and in beta diversity analyses. Significant differences in *Candida* abundances between OW and BAL has been shown earlier by Cui et al. (113). However, they did not find any difference between OW and BAL in PCoA plots based on OW, BAL, and IS samples from healthy individuals. Their results are somewhat hard to interpret since they did not include any statistical tests to look for differences in beta diversity. The visual interpretation of PCoA plots is subjective. They also rarefied their data to 50 reads per sample. With such low rarefaction depth, samples may end up having no shared features, which would result in a distance of 1.0 with UniFrac. Furthermore, they only included BAL and IS in the PCoA plot made from the entire cohort including HIV-infected and HIV-uninfected individuals with or without normal lung function, and it should be interesting to know where OW samples would have ended up in that plot.

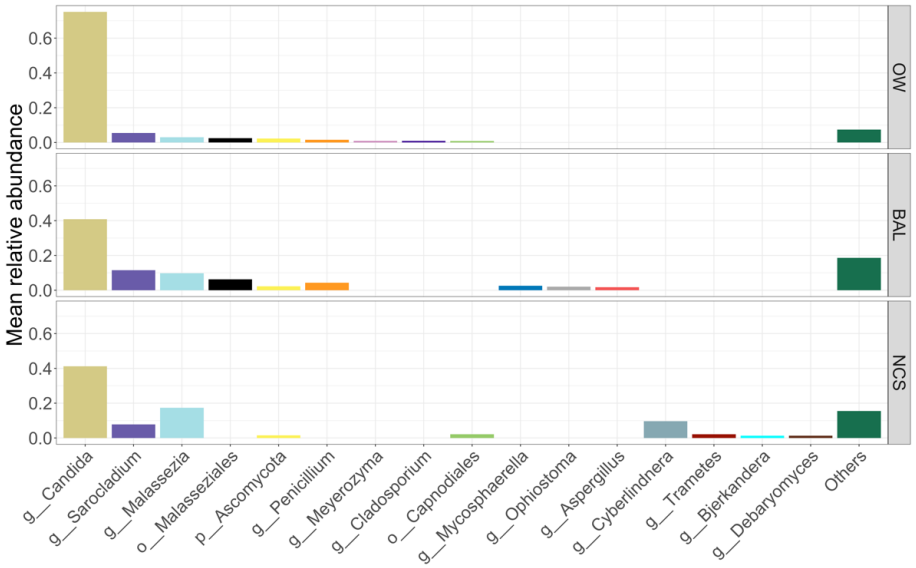
The difference in OW and BAL samples found in the MicroCOPD study suggests a unique mycobiome in the lungs. However, that does not mean that the oral and the lung mycobiome is not associated or influenced by each other. The larynx is the only part that blocks the lungs from being in direct communication with the air and the mouth. Consequently, microorganisms in air, mouth, or even the stomach could be dispersed into the lungs by inhalation or microaspiration, while microorganisms could move the other way through coughing. One study has suggested that most of the microbes in the lungs indeed stem from the oral cavity in healthy subjects, but that local environmental factors and active selection is more important in disease (72). Charlson et al. proposed that there existed a continuity in the respiratory microbiome, ranging from the upper respiratory tract to the lower respiratory tract with decreasing biomass, but that no distinct lung-specific microbiome existed in healthy people (71). The study was limited by a low sample size (n=6). Later, Dickson et al. proposed an adapted island model of lung biogeography, depicting the respiratory system as a function of immigration and extinction of microbes that originate in the upper airway (284). Factors like disease and medication use influence immigration and extinction, thereby changing species richness in a defined area of the respiratory tract. Dickson et al. later showed that the proximal part of the lower respiratory tract more closely resembled the supraglottic space than the distal part, and that the richness decreased with increasing distance from the larynx (42). The question is maybe not whether transfer of microbes between the mouth and lung exists, but rather if lung microbiome studies should focus on authentic microbes with replication in the lung or include transferred microbes and thus analyse the whole community.

The use of BAL in sampling of the lungs has been shown in several studies to sample the airways and not just finding microbiota as a result of oral contamination (42, 285-287), especially when using protected sampling (9). Although these studies have not included fungal data, we believe our BAL samples were representative of the lung environment. Charlson et al. argued that comparing an OW sample and a BAL sample from the same

participant inevitably will detect differences due to rare organisms in the OW sample and suggests use of replicate sampling instead (286). Replicate sampling adds additional costs to already expensive studies and might prove impossible to perform in large-scale studies like the MicroCOPD study. However, only the most abundant genus, *Candida*, was found by the differential abundance/distribution tests to significantly differ between OW and BAL samples, showing less impact from rare taxa.

Figure 5 shows the most abundant fungi found in each of the sample types included in the MicroCOPD study.

Figure 5. Rank abundance plots using most abundant fungi in oral wash, bronchoalveolar lavage, and negative control samples in the MicroCOPD study



OW, oral wash; BAL, bronchoalveolar lavage; NCS, negative control sample. Contaminants identified by *Decontam* are excluded in OW and BAL, but not NCS.

An oral fungal environment rich in *Candida* is well-known from the literature (116, 230, 288). The similarity between BAL and NCS is somewhat concerning, previously also reported by Bittinger et al. (130). Reassuringly though, we did find differences in abundance. *ANCOM v2* reported that *Candida* and *Sarocladium* were different in abundance between OW and NCS, and BAL and NCS. *Candida* was also reported to differ between OW and NCS by *ALDEx2*. Differences were also seen in DNA yields and read counts between NCS and OW/BAL. By looking at relative abundances, such differences are not detected. *Decontam* was used to remove contaminants in OW and BAL samples, but we could have considered to remove OW or BAL samples that resembled NCS. However, we do find this problematic, because it might be that the true constituents of the environmental microbiota, that influence the NCS, are the same as we find in most lung samples.

11.2.1.2 The healthy pulmonary mycobiome

The healthy lung mycobiome has been examined in several studies (100, 113, 128, 130, 175, 289), but sample sizes were relatively small, ranging from 10 participants (289) to 47 participants (175). Studies reported that abundant fungi were *Candida* (175), *Davidiellaceae* (100), *Cladosporium* (100, 128, 130), *Saccharomyces* (113, 175), *Penicillium* (113), *Debaryomyces* (130), *Aspergillus* (100, 289), *Eremothecium* (128), *Systenostrema* (128), and *Malasseziales* (289). BAL samples in the MicroCOPD study contained mainly *Candida*, *Malassezia*, and *Sarocladium*. *Candida* is a well-known pathogen (135), and *C. albicans* is the main pathogen causing candidiasis in most clinical settings. Non-albicans *Candida* species are, nevertheless, increasingly reported as a cause of infections (290). Unfortunately, our analyses were restricted to genus level due to limitations in the ITS barcode in correctly identifying fungi down to species level (117). It would be interesting to have more information on species level to further elucidate *Candida*'s role in maintaining health or causing disease. We observed that *Candida* resided in the lungs of a large fraction of controls with an unknown clinical consequence. There was no difference in *Candida* abundance between healthy controls and participants

with COPD. One could speculate that *Candida* residing in lungs of healthy controls could cause an infection in presence of certain triggers. Such mechanisms have been discussed in the context of the gut mycobiome (291), and it is not unlikely that similar mechanisms could exist in the lungs. Due to the *Candida* dominance, we reported every ASV assigned as *Candida* down to genus and species level in Table 7. The *Candida* taxa reported were found in the finally processed OW and BAL samples included for further analyses both from the first examinations and repeated procedures. The table should be read with some caution, though, due to the uncertainty of species discrimination for ITS1 data (117).

Table 7: Summary of ASVs assigned to *Candida* in the MicroCOPD study

ASV ID	Taxonomic assignment
0336a05ed19277bdf8c25804c187fa0b	k__Fungi;p__Ascomycota;c__Saccharomycetes;o__Saccharomycetales;f__Saccharomycetales_fam_Incertae_sedis;g__Candida;s__Candida_albicans
069929698b007a6ee5210bf7cf150083	k__Fungi;p__Ascomycota;c__Saccharomycetes;o__Saccharomycetales;f__Saccharomycetales_fam_Incertae_sedis;g__Candida;s__Candida_albicans
0bcd4699aaaef91810304329c4000458	k__Fungi;p__Ascomycota;c__Saccharomycetes;o__Saccharomycetales;f__Saccharomycetales_fam_Incertae_sedis;g__Candida;s__Candida_albicans
2ab4a5a14cce5b707ea294316c5f8ad5	k__Fungi;p__Ascomycota;c__Saccharomycetes;o__Saccharomycetales;f__Saccharomycetales_fam_Incertae_sedis;g__Candida;s__Candida_maltosa
326d37aeef9d4ef8f672c8372d5e47b7	k__Fungi;p__Ascomycota;c__Saccharomycetes;o__Saccharomycetales;f__Saccharomycetales_fam_Incertae_sedis;g__Candida;s__Candida_dubliniensis
38a9dc74ed923f0c8683e5f59d25dae5	k__Fungi;p__Ascomycota;c__Saccharomycetes;o__Saccharomycetales;f__Saccharomycetales_fam_Incertae_sedis;g__Candida;s__Candida_maltosa
3fd75501a4b37554390ef45724395bb2	k__Fungi;p__Ascomycota;c__Saccharomycetes;o__Saccharomycetales;f__Saccharomycetales_fam_Incertae_sedis;g__Candida;s__Candida_railenensis
4c5a94cf2792d1e0dad3cbec8262cb62	k__Fungi;p__Ascomycota;c__Saccharomycetes;o__Saccharomycetales;f__Saccharomycetales_fam_Incertae_sedis;g__Candida;s__Candida_albicans
5a484d884f715bd9b91fd5de883f47cf	k__Fungi;p__Ascomycota;c__Saccharomycetes;o__Saccharomycetales;f__Saccharomycetales_fam_Incertae_sedis;g__Candida;s__Candida_maltosa
66bb3b905e859fabca7559698068d546	k__Fungi;p__Ascomycota;c__Saccharomycetes;o__Saccharomycetales;f__Saccharomycetales_fam_Incertae_sedis;g__Candida;s__Candida_maltosa
6740b028fa544055ca8389831d5811fc	k__Fungi;p__Ascomycota;c__Saccharomycetes;o__Saccharomycetales;f__Saccharomycetales_fam_Incertae_sedis;g__Candida;s__Candida_friedrichii
6b79b8cfa44f917109d5c99c6ed1c350	k__Fungi;p__Ascomycota;c__Saccharomycetes;o__Saccharomycetales;f__Saccharomycetales_fam_Incertae_sedis;g__Candida;s__Candida_albicans
7b1133da646399236e40fa903a2fdf12	k__Fungi;p__Ascomycota;c__Saccharomycetes;o__Saccharomycetales;f__Saccharomycetales_fam_Incertae_sedis;g__Candida;s__Candida_tropicalis
803aab77225830ac71f1ce35d71f34c8	k__Fungi;p__Ascomycota;c__Saccharomycetes;o__Saccharomycetales;f__Saccharomycetales_fam_Incertae_sedis;g__Candida;s__Candida_zeylanoides
a0a7fc5a21f96eae94265dc987372c67	k__Fungi;p__Ascomycota;c__Saccharomycetes;o__Saccharomycetales;f__Saccharomycetales_fam_Incertae_sedis;g__Candida;s__Candida_albicans
a72c1820d53342ed63b38c14763d04a1	k__Fungi;p__Ascomycota;c__Saccharomycetes;o__Saccharomycetales;f__Saccharomycetales_fam_Incertae_sedis;g__Candida;s__Candida_maltosa
b2d186cbeb52a7e284be3025c7296ee0	k__Fungi;p__Ascomycota;c__Saccharomycetes;o__Saccharomycetales;f__Saccharomycetales_fam_Incertae_sedis;g__Candida;s__Candida_albicans
bb1800678bd134f4e624eac405e5fcdb	k__Fungi;p__Ascomycota;c__Saccharomycetes;o__Saccharomycetales;f__Saccharomycetales_fam_Incertae_sedis;g__Candida;s__Candida_dubliniensis
c5948bb2019826b1dd03da28e5bc06ea	k__Fungi;p__Ascomycota;c__Saccharomycetes;o__Saccharomycetales;f__Saccharomycetales_fam_Incertae_sedis;g__Candida;s__Candida_dubliniensis
cae63524c934edac25bd15cb100732db	k__Fungi;p__Ascomycota;c__Saccharomycetes;o__Saccharomycetales;f__Saccharomycetales_fam_Incertae_sedis;g__Candida;s__Candida_hyderabadensis
d13eee2d29e795f7a611b571098b5f2f	k__Fungi;p__Ascomycota;c__Saccharomycetes;o__Saccharomycetales;f__Saccharomycetales_fam_Incertae_sedis;g__Candida
d5a3b43015c417b247b0ee32248fddb5	k__Fungi;p__Ascomycota;c__Saccharomycetes;o__Saccharomycetales;f__Saccharomycetales_fam_Incertae_sedis;g__Candida;s__Candida_albicans
fd6afda5d5013896f80abf3d661efa02	k__Fungi;p__Ascomycota;c__Saccharomycetes;o__Saccharomycetales;f__Saccharomycetales_fam_Incertae_sedis;g__Candida
fdb44eafd771df13907fb1e1e0ad0dd2	k__Fungi;p__Ascomycota;c__Saccharomycetes;o__Saccharomycetales;f__Saccharomycetales_fam_Incertae_sedis;g__Candida;s__Candida_albicans

ASV, amplicon sequence variant. The ASVs above were all assigned at least to *Candida* at genus level.

Species of *Malassezia* are common skin commensals, and are associated with a variety of skin disorders, such as pityriasis versicolor, atopic dermatitis, and psoriasis, and can lead to fungemia in susceptible individuals (292). That the observed *Malassezia* in the MicroCOPD study stem from contamination during sample handling cannot be ruled out, as it is commonly found on skin. Efforts were exerted to reduce contamination, including protected BAL sampling and removal of contaminants using *Decontam*, strategies not implemented by the other studies reporting on the healthy lung mycobiome. Dupuy and colleagues emphasise harsh cell lysis methods to detect *Malassezia* (230), and Hoggard et al. found an under-representation of *Malassezia* using primer pair ITS1F – ITS2 (233). Different DNA extraction methods and primers could thus explain the observed gaps in *Malassezia* proportions. *Sarocladium kiliense* is usually found in soil, but do occasionally cause human infections (293), also reported in lungs (294).

11.2.1.3 The COPD pulmonary mycobiome

We observed few differences between the healthy lung mycobiome and the COPD mycobiome in the MicroCOPD study. Concerning abundance, only the *MicrobiomeDDA* algorithm indicated differences between the participant categories. The results from *MicrobiomeDDA* were not replicated by *ANCOM v2* and *ALDEx2*, which questions the reliability of the *MicrobiomeDDA* results. Only three previous studies have explored the lung mycobiome in COPD (108, 113, 175). In contrast to our study, Cui et al. suggested that *P. jirovecii* was associated with COPD (113). However, participants with COPD and the group for comparison were also infected with HIV. This complicates interpretation because *P. jirovecii* is known to be associated with immunosuppression (295), and have previously been detected in the respiratory tract of HIV-infected individuals using nested PCR (296). Another study, however, has shown an association between *Pneumocystis* colonisation and severity of airflow obstruction in smokers (297), possibly suggestion a role of *Pneumocystis* in COPD pathogenesis anyway. We observed no *Pneumocystis* in our data, in line with another large multi-centre study including participants with COPD

(175). The *Pneumocystis* genome only includes one copy of the ITS1 locus, which could result in a negative sequencing result (95).

Su et al. included participants with COPD in their mycobiome study, but participants experienced an exacerbation at the time of inclusion (108). Furthermore, they collected sputum samples, which complicates direct comparison to the BAL samples in the MicroCOPD study. Sputum samples were collected from participants experiencing a COPD exacerbation in the study by Tiew et al. as well (175). Additionally, they included 337 participants with stable COPD, and 47 non-diseased controls. Both participants with COPD and controls showed high abundances of *Candida* (175). In contrast to our study, they observed significantly increased Shannon indexes in the COPD participants compared to controls. Furthermore, they reported several fungi to be COPD specific, including *Trametes*, *Penicillium*, *Mycosphaerella*, *Cryptococcus*, *Cladosporium*, and *Aspergillus*. Our data agreed that *Trametes* and *Cryptococcus* were only found in participants with COPD, in relative abundances of 2 and 0.7%, respectively. However, the remaining taxa listed above were all found in controls in the MicroCOPD study. It should be mentioned that four of the other COPD specific taxa found in the study by Tiew et al., *Wallemia*, *Itersonilia*, *Aureobasidium*, and *Alternaria*, were only found in NCS in the MicroCOPD study, and identified by *Decontam* as contaminants. The *Decontam* algorithm was also employed in Tiew et al.'s study. One can only speculate whether the dissimilar *Wallemia*, *Itersonilia*, *Aureobasidium*, and *Alternaria* classifications represents technical or true differences. One strength of the MicroCOPD study is the inclusion of one NCS per participant. In comparison, Tiew et al. included total of fourteen negative control samples (i.e. PBS) and extraction blanks (i.e. Zymo reagents). Discrepancies in diversity and taxonomy between our and Tiew et al.'s results could perhaps be attributed to the different sample types used, as IS samples have been shown to cluster differently than BAL samples in a PCoA space (113). Still, since similar sample types were utilised inside each of the study designs, the effect from different sample types should be small. Some differences were seen in the fungal taxonomy from

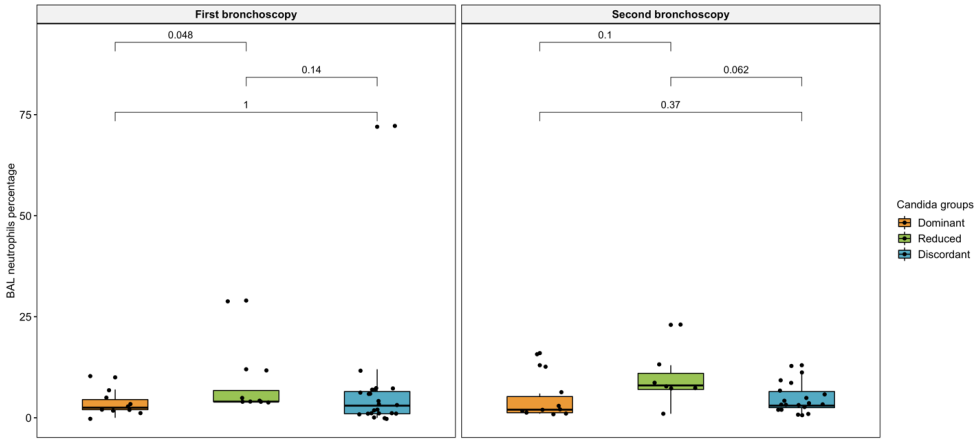
Singapore/Malaysia and Dundee, suggesting that geographical differences in lung mycobiomes exist (175). Geographical differences to our study centre could also explain the observed discrepancies between Tiew et al.'s study and the MicroCOPD study. Nevertheless, mentioned discrepancies are worthy of additional investigations to conclude further. In congruence with our results, they suggested that the lung mycobiomes were unaffected by treatment with ICS in patients with stable COPD (175).

11.2.2 Longitudinal analyses on the pulmonary mycobiome

We have shown in Paper IV that the oral mycobiome showed a high degree of stability, but less so for the pulmonary mycobiome. Intercurrent antibiotic use did not seem to influence the mycobiome. As found in paper III, most sample pairs were dominated by *Candida*, particularly for OW samples. Based on the Yue-Clayton plots, we visually divided each of the participants into three groups: *Candida* dominated, *Candida* reduced, and *Candida* discordant. The dominated and the reduced group had high and low relative abundances of *Candida* both in the first and the second bronchoscopy, respectively. The discordant group were participants with a high relative abundance of *Candida* in one of the bronchoscopies, but low relative abundance of *Candida* in the corresponding bronchoscopy. When participants with asthma were included, no differences were seen between the three groups in terms of C-reactive protein (CRP), thrombocyte count, interleukin 8 (IL-8), COPD assessment test (CAT) score, BAL yield in percentage, BAL macrophages, BAL lymphocytes, BAL eosinophils, sex, participant group, smoking or gastroesophageal reflux disease medications. We did, however, see an increased number of BAL neutrophils in the *Candida* reduced group compared to the *Candida* dominant group for the first bronchoscopy (Figure 6, p-value 0.048). This result should be read with caution. We did not adjust for multiple testing, participants with asthma were not excluded from the analysis, and we could not replicate the finding with blood samples taken from the second bronchoscopy (Figure 6, p-value 0.1). Given that the observation

could be reproduced by other researchers in other populations, one could speculate that *Candida* dominance provides protection against inflammation.

Figure 6: BAL neutrophils between different *Candida* groups in the first and second bronchoscopy in the MicroCOPD study



BAL, bronchoalveolar lavage. Neutrophil content in BAL were compared with Wilcoxon signed-rank test based on the *Candida* content in the sample pairs from participants undergoing two bronchoscopies in the MicroCOPD study.

Three observations strengthened the view of a more stable oral mycobiome compared to the pulmonary mycobiome in the MicroCOPD study. First, the average Yue-Clayton measure in OW bronchoscopy pairs were 0.22 compared to 0.69 in BAL samples. As previously mentioned, the Yue-Clayton measure is based on differences in taxa's relative abundances and is 0 with perfect similarity and 1 with perfect dissimilarity. The lower average Yue-Clayton measure in OW sample pairs is thus suggestive of a more stable mycobiome. Second, pairwise distances between each sample pair were significantly higher in BAL compared to OW using Bray-Curtis as distance metric. The higher Bray-Curtis value, the more dissimilar were the two samples in a sample pair. We did not find any significant differences between pairwise distances using Jaccard. This is probably explained by the high *Candida* dominance in OW samples. A high *Candida* dominance

will affect Bray-Curtis and Jaccard differently since Bray-Curtis takes abundance into account, while Jaccard is based merely on absence-presence, reducing the impact from *Candida*. Significantly higher distances in BAL samples compared to OW samples using the Bray-Curtis distance metric add to the notion of a more stable OW mycobiome. Finally, taxonomy seems to change less between the first and the second bronchoscopy in OW compared to BAL by visual impression. But the high *Candida* dominance could create a visual illusion because of the less coloured OW plot. A differential abundance test could provide a more reliable estimate on taxonomy changes, but we were concerned to include a differential abundance test on group level, as the most interesting changes were intra-individually. We thus had greater confidence using the Yue-Clayton measures, which is calculated for each participant separately.

The literature on the stability of the lung mycobiome is limited, and only three studies have included participants with COPD. Bafadhel et al. collected sputum samples from participants with COPD at baseline and three months later to examine the stability of *Aspergillus fumigatus* cultures (298). They found a poor repeatability of *Aspergillus fumigatus*, but comparison to our study is difficult due to differences in sample type and methodology. The two last studies examined longitudinal changes in the lung mycobiome during exacerbations (108, 175). The study by Su et al. found an unstable mycobiome in included patients. But samples were only collected during participants' hospital stays, which ranged from 7 to 16 days. The time interval is probably too short to draw conclusions on the mycobiome stability. Furthermore, only six participants were included, and no statistics were applied to the results from consecutive collected samples. In Tiew et al.'s study, repeated sputum sampling was performed in 34 participants with COPD before an exacerbation, within 24 hours of an acute exacerbation and again two weeks post exacerbation following treatment with one week of oral antibiotics (either doxycycline or co-amoxiclav) and five days of oral corticosteroids (prednisolone) (175). No significant changes were seen in airway mycobiome profiles, alpha diversity, or beta diversity. Taxonomy was only presented on group level between the different time points,

which could have masked intraindividual differences, as seen in the MicroCOPD study. Results from diversity analyses concurred with results from the MicroCOPD study, though hard to compare due to a low sampling interval and treatment given in the study by Tiew et al. (175). The authors concluded that treatment of an acute exacerbation did not alter the lung mycobiome, but ideally, they should have examined the lung mycobiome at an additional later time point, for instance 3-4 months later.

One could speculate that the bronchoscopic procedure in the MicroCOPD study affected the lung environment more than the sputum sampling performed in the aforementioned exacerbation study by Tiew et al. (175). It has been shown that bronchoscopy and BAL is associated with a potential immune response, exemplified by fever and flu-like symptoms in the following days after the procedure (299, 300). If the immune response is caused by respiratory microorganisms, such immune responses could be involved in the elimination of the given immune response trigger, i.e. the given microorganism. An increased immune response from bronchoscopy could thus have explained the instability in BAL samples, which was not observed in sputum samples during an exacerbation (175). That some of the BAL sample pairs showed identical taxonomy between the two time points challenges the immune response hypothesis but could have several explanations. For instance, that the immune response only happens in a fraction of participants, or that another specie from the same genus has replaced the eliminated microorganism. The higher stability observed in OW samples fits the hypothesis, since the oral cavity is less triggered by the bronchoscopy, and is more accustomed to exposures. We therefore expect less impact from the bronchoscopy on the oral mycobiome. The immune response hypothesis presented here should be examined with multiple sampling time points in larger populations.

Some other studies have performed repeated sampling of the lung mycobiome, but included CF patients (95, 127) or intubated and mechanically ventilated patients with pneumonia (132), which complicates further comparisons to our results.

12 Conclusions

1. There was few publications available on response rates in bronchoscopy studies. Despite a limited literature, we found that response rates were generally high, but varied between healthy and diseased people and with age. The invasive nature of a bronchoscopy was a common reason to decline participation, while responders seemed to be motivated by a combination of personal health benefit and altruism.

2. Just above 50% of invited subjects accepted and underwent a bronchoscopy in the MicroCOPD study, showing that large-scale bronchoscopy studies are feasible. Participants had a somewhat misleading perception of personal health benefit from participation which emphasise the importance of providing accurate study information at recruitment. Detailed information about the procedure, and avoidance of long waiting times, are important to reduce participants' fears and worries.

3. Oral and pulmonary samples differed in taxonomic composition and diversity in the MicroCOPD study, possibly indicating the existence of a pulmonary mycobiome. No consistent differences were found between participants with COPD and controls in terms of differential abundance/distribution, alpha diversity, or beta diversity. ICS use could not be seen to significantly affect the lung mycobiome.

4. The lung mycobiome showed less stability compared to the oral mycobiome in participants from the MicroCOPD study. Neither intercurrent antibiotic use nor time between bronchoscopies seemed to influence the mycobiomes.

13 Future perspectives and implications

The literature review included in the current thesis showed that further research on participation in research bronchoscopy studies is warranted. The analyses on participation in the MicroCOPD study contributed to fill this void, but some topics are still understudied. First, more detailed demographics of non-responders would be useful. We did not apply for any extended ethics approval, preventing detailed examination of non-responders in the MicroCOPD study. Our analyses were thus limited to data collected from their participation in previous studies. Furthermore, an in-depth interview made specifically for a qualitative study would probably be more appropriate to extend our knowledge on participation in research bronchoscopy studies. Additionally, information on participants' understanding of study information also enlightens us on issues regarding bronchoscopy in daily clinical practice, a well-established procedure frequently performed in the Norwegian health system. Newly acquired knowledge on bronchoscopy studies could also facilitate future large-scale studies on the COPD mycobiome.

COPD mycobiome studies are still in their infancy. We have seen in this thesis that COPD mycobiome studies suffer from a lack of a standardised workflow, preventing generalisation. First priority of future studies should thus be to follow a common workflow that researchers have agreed upon so that prior publications, including those in the current thesis, could be validated, and to secure that newly generated results are reliable. Specifically, studies should include larger sample sizes, measure DNA quantitively, and include both negative and positive controls and relevant mock communities. Furthermore, DNA extractions should be performed similarly across studies, and the processing of at least a fraction of the samples should be replicated. Studies should preferably include the same bioinformatical principles. It is however common practice to publish the sequencing data from microbiome studies, so if we could agree on a common workflow prior to the bioinformatical processing, reliable sequencing data could be bioinformatically processed and analysed by different research teams.

When we feel confident that we have characterised the COPD mycobiome using reliable data, studies could move on to examine the COPD mycobiome further, for instance their role in exacerbations.

Exacerbations is a common feature of the COPD disease. Tiew et al. has shown in a recent publication that the COPD lung mycobiome is associated with exacerbations (175). However, they did not observe any changes in their longitudinal analysis during an exacerbation, and future studies should repeat the analyses with increased time intervals or potentially further sampling time points. Considering the dominance of *Candida* in Tiew et al.'s exacerbation study and our results, and our suggestion of *Candida* as a stabilising factor of the lung mycobiome, it should be interesting to reveal *Candida*'s role in exacerbations in more detail.

The majority of studies on the COPD microbiome have focused on bacteria, but recently also the viral part, or the virome, has been examined by metagenomic NGS in participants with a COPD exacerbation (301). Although most studies on COPD so far have examined the bacterial, fungal and viral part separately, less is known on how they interact with each other. We know that fungal-bacterial interactions can be beneficial or detrimental for the host (302), and it seems reasonable to propose that further studies should look more into inter-kingdom interactions in diseased COPD lungs. Furthermore, future studies should relate COPD mycobiomes with the host response to further elucidate the mycobiome's role in disease. Targeted amplicon sequencing is unable to describe mycobiomes' functional effects, but metagenomic sequencing approaches are more well-suited. A metagenomic study of the COPD mycobiome has identified fungi that are responsible for allergic sensitisation in COPD and shown that the sensitisation associates with frequent exacerbations (303). The study is an example of how new and advanced sequencing techniques can create new knowledge.

New knowledge carries hope for future implications. Hypothetically, knowledge on "helpful" and "exacerbating" fungi in the airways could lead to future screening of COPD patients. More directly, information on the architecture of the airway mycobiome might reveal an advantageous composition of fungi, enabling us to design pro- or prebiotic treatment, or detect specific species or strains with detrimental effect on lung health, enabling effective antimycotic treatment. Testing existing antimycotics on identified species and strains are definitely possible, whereas pro- or prebiotic treatment might require some sort of collaboration with industrial partners. The load from COPD on Norwegian hospitals is huge in terms of personnel and economics, and the disease inflicts serious inconveniences on COPD patients' life through several hospital stays, unpleasant symptoms and fear. Clearly, more information on new and promising treatment options that reduce the incidence of COPD exacerbations, and even prevent disease, would have resulted in beneficial implications for the Norwegian health system. In the end, it is such information that one day might benefit patients with COPD as well.

14 Minor errata

Paper I

The study objective of Schook et al.'s study was reported as "Examine whether participation in a smoking cessation trial could influence smoking cessation" (90) in Table 3. It should have been "Examine whether participation in a chemoprevention study could influence smoking cessation".

Additionally, the citation of the study with the lowest response rate should have been reference 13 (Kye et al., reference (89) in this thesis), and not reference 11 (Chudleigh et al., reference (92) in this thesis). The citation is found in the "Response rates" undersection in the "Results".

Paper II

One missing value for number of cigarettes were included as 99 (the default missing value used in data entry) and thus interpreted as 99 cigarettes, and one participant had incorrect smoking amount, which was later changed from 50 to 7. Additionally, non-smokers were not included in the mean calculations, creating falsely high means. The control and COPD groups did not differ much after reanalyses, but mean pack years changed from 20.9 to 10.4 in the asthma group due to the low number of participants with asthma. Pack years and smoking amount was not included in any analysis other than demographics.

Two controls had some minor errors in age from the data entry, which also affected the predicted values of FEV_1 and FVC. We reanalysed age, and FEV_1 and FVC in percentage of predicted for the demographic table. The only observed difference was a change in the standard deviation in FVC in percentage of predicted for the controls from 13.5 to 13.4. No further re-analyses were done as any impact on the results seemed unlikely.

15 References

1. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017;390(10100):1151-210.

2. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med. 2006;3(11):e442.

3. Nielsen R, Johannessen A, Benediktsdottir B, Gislason T, Buist AS, Gulsvik A, et al. Present and future costs of COPD in Iceland and Norway: results from the BOLD study. Eur Respir J. 2009;34(4):850-7.

4. Bhowmik A, Seemungal TA, Sapsford RJ, Wedzicha JA. Relation of sputum inflammatory markers to symptoms and lung function changes in COPD exacerbations. Thorax. 2000;55(2):114-20.

5. Patel IS, Seemungal TA, Wilks M, Lloyd-Owen SJ, Donaldson GC, Wedzicha JA. Relationship between bacterial colonisation and the frequency, character, and severity of COPD exacerbations. Thorax. 2002;57(9):759-64.

6. Sethi S, Murphy TF. Infection in the pathogenesis and course of chronic obstructive pulmonary disease. N Engl J Med. 2008;359(22):2355-65.

7. 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.

8. Dickson RP, Erb-Downward JR, Martinez FJ, Huffnagle GB. The Microbiome and the Respiratory Tract. Annu Rev Physiol. 2016;78:481-504.

9. Grønseth R, Drengenes C, Wiker HG, Tangedal S, Xue Y, Husebø GR, et al. Protected sampling is preferable in bronchoscopic studies of the airway microbiome. ERJ Open Res. 2017;3(3).

10. Kong HH, Morris A. The emerging importance and challenges of the human mycobiome. Virulence. 2017;8(3):310-2.

11. Vestbo J. COPD: definition and phenotypes. Clin Chest Med. 2014;35(1):1-6.

12. Vogelmeier CF, Román-Rodríguez M, Singh D, Han MK, Rodríguez-Roisin R, Ferguson GT. Goals of COPD treatment: Focus on symptoms and exacerbations. Respir Med. 2020;166:105938.

13. Bhatt SP, Balte PP, Schwartz JE, Cassano PA, Couper D, Jacobs DR, Jr., et al. Discriminative Accuracy of FEV1:FVC Thresholds for COPD-Related Hospitalization and Mortality. Jama. 2019;321(24):2438-47.

14. Global Initiative for Chronic Obstructive Lung Disease. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease (2020 REPORT). https://goldcopd.org/wp-content/uploads/2019/12/GOLD-2020-FINAL-ver1.2-03Dec19_WMV.pdf Last accessed: November 12, 2020.

15. López-Campos JL, Tan W, Soriano JB. Global burden of COPD. Respirology. 2016;21(1):14-23.

16. Lamprecht B, McBurnie MA, Vollmer WM, Gudmundsson G, Welte T, Nizankowska-Mogilnicka E, et al. COPD in never smokers: results from the population-based burden of obstructive lung disease study. Chest. 2011;139(4):752-63.

17. Brightling C, Greening N. Airway inflammation in COPD: progress to precision medicine. Eur Respir J. 2019;54(2).

18. Barnes PJ. Inflammatory mechanisms in patients with chronic obstructive pulmonary disease. J Allergy Clin Immunol. 2016;138(1):16-27.

19. Polosukhin VV, Richmond BW, Du RH, Cates JM, Wu P, Nian H, et al. Secretory IgA Deficiency in Individual Small Airways Is Associated with Persistent Inflammation and Remodeling. Am J Respir Crit Care Med. 2017;195(8):1010-21.

20. Hogg JC, Chu F, Utokaparch S, Woods R, Elliott WM, Buzatu L, et al. The nature of small-airway obstruction in chronic obstructive pulmonary disease. N Engl J Med. 2004;350(26):2645-53.

21. McDonough JE, Yuan R, Suzuki M, Seyednejad N, Elliott WM, Sanchez PG, et al. Small-airway obstruction and emphysema in chronic obstructive pulmonary disease. N Engl J Med. 2011;365(17):1567-75.

22. Barnes PJ. Cellular and molecular mechanisms of chronic obstructive pulmonary disease. Clin Chest Med. 2014;35(1):71-86.
23. Ofir D, Laveneziana P, Webb KA, Lam YM, O'Donnell DE. Mechanisms of dyspnea during cycle exercise in symptomatic patients with GOLD stage I chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2008;177(6):622-9.
24. Elbehairy AF, Ciavaglia CE, Webb KA, Guenette JA, Jensen D, Mourad SM, et al. Pulmonary Gas Exchange Abnormalities in Mild Chronic Obstructive Pulmonary Disease. Implications for Dyspnea and Exercise Intolerance. Am J Respir Crit Care Med. 2015;191(12):1384-94.
25. Fletcher CM. Chronic bronchitis. Its prevalence, nature, and pathogenesis. Am Rev Respir Dis. 1959;80:483-94.
26. Postma DS, Weiss ST, van den Berge M, Kerstjens HA, Koppelman GH. Revisiting the Dutch hypothesis. J Allergy Clin Immunol. 2015;136(3):521-9.
27. Barnes PJ. Against the Dutch hypothesis: asthma and chronic obstructive pulmonary disease are distinct diseases. Am J Respir Crit Care Med. 2006;174(3):240-3; discussion 3-4.
28. Kraft M. Asthma and chronic obstructive pulmonary disease exhibit common origins in any country! Am J Respir Crit Care Med. 2006;174(3):238-40; discussion 43-4.
29. Ghebre MA, Bafadhel M, Desai D, Cohen SE, Newbold P, Rapley L, et al. Biological clustering supports both "Dutch" and "British" hypotheses of asthma and chronic obstructive pulmonary disease. J Allergy Clin Immunol. 2015;135(1):63-72.
30. 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.
31. Kanner RE, Anthonisen NR, Connett JE. Lower respiratory illnesses promote FEV(1) decline in current smokers but not ex-smokers with mild chronic obstructive pulmonary disease: results from the lung health study. Am J Respir Crit Care Med. 2001;164(3):358-64.

32. Donaldson GC, Seemungal TA, Bhowmik A, Wedzicha JA. Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. Thorax. 2002;57(10):847-52.

33. Vestbo J, Edwards LD, Scanlon PD, Yates JC, Agusti A, Bakke P, et al. Changes in forced expiratory volume in 1 second over time in COPD. N Engl J Med. 2011;365(13):1184-92.

34. Soler N, Ewig S, Torres A, Filella X, Gonzalez J, Zaubet A. Airway inflammation and bronchial microbial patterns in patients with stable chronic obstructive pulmonary disease. Eur Respir J. 1999;14(5):1015-22.

35. Hill AT, Campbell EJ, Hill SL, Bayley DL, Stockley RA. Association between airway bacterial load and markers of airway inflammation in patients with stable chronic bronchitis. Am J Med. 2000;109(4):288-95.

36. Müllerova H, Maselli DJ, Locantore N, Vestbo J, Hurst JR, Wedzicha JA, et al. Hospitalized exacerbations of COPD: risk factors and outcomes in the ECLIPSE cohort. Chest. 2015;147(4):999-1007.

37. Sapey E, Stockley RA. COPD exacerbations . 2: aetiology. Thorax. 2006;61(3):250-8.

38. Seemungal T, Harper-Owen R, Bhowmik A, Moric I, Sanderson G, Message S, et al. Respiratory viruses, symptoms, and inflammatory markers in acute exacerbations and stable chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2001;164(9):1618-23.

39. Glezen WP, Greenberg SB, Atmar RL, Piedra PA, Couch RB. Impact of respiratory virus infections on persons with chronic underlying conditions. Jama. 2000;283(4):499-505.

40. Rohde G, Wiethege A, Borg I, Kauth M, Bauer TT, Gillissen A, et al. Respiratory viruses in exacerbations of chronic obstructive pulmonary disease requiring hospitalisation: a case-control study. Thorax. 2003;58(1):37-42.

41. Tan WC, Xiang X, Qiu D, Ng TP, Lam SF, Hegele RG. Epidemiology of respiratory viruses in patients hospitalized with near-fatal asthma, acute exacerbations of asthma, or chronic obstructive pulmonary disease. Am J Med. 2003;115(4):272-7.
42. Dickson RP, Erb-Downward JR, Freeman CM, McCloskey L, Beck JM, Huffnagle GB, et al. Spatial Variation in the Healthy Human Lung Microbiome and the Adapted Island Model of Lung Biogeography. Ann Am Thorac Soc. 2015;12(6):821-30.
43. Jones FS. The source of the microorganisms in the lungs of normal animals. J Exp Med. 1922;36(3):317-28.
44. Quinn LH, Meyer OO. The relationship of sinusitis and bronchiectasis. Archives of Otolaryngology. 1929;10(2):152-65.
45. Dickson RP, Erb-Downward JR, Prescott HC, Martinez FJ, Curtis JL, Lama VN, et al. Analysis of culture-dependent versus culture-independent techniques for identification of bacteria in clinically obtained bronchoalveolar lavage fluid. J Clin Microbiol. 2014;52(10):3605-13.
46. Dickson RP, Erb-Downward JR, Huffnagle GB. The role of the bacterial microbiome in lung disease. Expert Rev Respir Med. 2013;7(3):245-57.
47. Faner R, Sibila O, Agustí A, Bernasconi E, Chalmers JD, Huffnagle GB, et al. The microbiome in respiratory medicine: current challenges and future perspectives. Eur Respir J. 2017;49(4).
48. Cho I, Blaser MJ. The human microbiome: at the interface of health and disease. Nat Rev Genet. 2012;13(4):260-70.
49. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. Nature. 2007;449(7164):804-10.
50. Ehrlich S, Consortium T. MetaHIT: The European Union Project on Metagenomics of the Human Intestinal Tract. 2010. p. 307-16.
51. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature. 2006;444(7122):1027-31.

52. Ley RE, Bäckhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI. Obesity alters gut microbial ecology. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(31):11070-5.

53. Castellarin M, Warren RL, Freeman JD, Dreolini L, Krzywinski M, Strauss J, et al. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. Genome Res. 2012;22(2):299-306.

54. Kostic AD, Gevers D, Pedamallu CS, Michaud M, Duke F, Earl AM, et al. Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. Genome Res. 2012;22(2):292-8.

55. Garrett WS, Gallini CA, Yatsunenko T, Michaud M, DuBois A, Delaney ML, et al. Enterobacteriaceae act in concert with the gut microbiota to induce spontaneous and maternally transmitted colitis. Cell Host Microbe. 2010;8(3):292-300.

56. Di Bella JM, Bao Y, Gloor GB, Burton JP, Reid G. High throughput sequencing methods and analysis for microbiome research. J Microbiol Methods. 2013;95(3):401-14.

57. Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ. At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. Appl Environ Microbiol. 2005;71(12):7724-36.

58. Cui L, Morris A, Ghedin E. The human mycobiome in health and disease. Genome Med. 2013;5(7):63.

59. Limon JJ, Skalski JH, Underhill DM. Commensal Fungi in Health and Disease. Cell Host Microbe. 2017;22(2):156-65.

60. Grada A, Weinbrecht K. Next-generation sequencing: methodology and application. J Invest Dermatol. 2013;133(8):e11.

61. 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.

62. 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.

63. R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ Last accessed: October 9, 2020.

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. Kõljalg U, Larsson KH, Abarenkov K, Nilsson RH, Alexander IJ, Eberhardt U, et al. UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. New Phytol. 2005;166(3):1063-8.

66. Nilsson RH, Larsson K-H, Taylor AF S, Bengtsson-Palme J, Jeppesen TS, Schigel D, et al. The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. Nucleic Acids Research. 2018;47(D1):D259-D64.

67. 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-.

68. Huang YJ, Erb-Downward JR, Dickson RP, Curtis JL, Huffnagle GB, Han MK. Understanding the role of the microbiome in chronic obstructive pulmonary disease: principles, challenges, and future directions. Transl Res. 2017;179:71-83.

69. Tipton L, Ghedin E, Morris A. The lung mycobiome in the next-generation sequencing era. Virulence. 2017;8(3):334-41.

70. 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):e00037-15.

71. Charlson ES, Bittinger K, Haas AR, Fitzgerald AS, Frank I, Yadav A, et al. Topographical continuity of bacterial populations in the healthy human respiratory tract. Am J Respir Crit Care Med. 2011;184(8):957-63.

72. Venkataraman A, Bassis CM, Beck JM, Young VB, Curtis JL, Huffnagle GB, et al. Application of a neutral community model to assess structuring of the human lung microbiome. mBio. 2015;6(1).
73. Morris A, Beck JM, Schloss PD, Campbell TB, Crothers K, Curtis JL, et al. Comparison of the respiratory microbiome in healthy nonsmokers and smokers. Am J Respir Crit Care Med. 2013;187(10):1067-75.
74. Segal LN, Alekseyenko AV, Clemente JC, Kulkarni R, Wu B, Gao Z, et al. Enrichment of lung microbiome with supraglottic taxa is associated with increased pulmonary inflammation. Microbiome. 2013;1(1):19.
75. Cabrera-Rubio R, Garcia-Núñez M, Setó L, Antó JM, Moya A, Monsó E, et al. Microbiome diversity in the bronchial tracts of patients with chronic obstructive pulmonary disease. J Clin Microbiol. 2012;50(11):3562-8.
76. Einarsson GG, Comer DM, McIlreavey L, Parkhill J, Ennis M, Tunney MM, et al. Community dynamics and the lower airway microbiota in stable chronic obstructive pulmonary disease, smokers and healthy non-smokers. Thorax. 2016;71(9):795-803.
77. 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.
78. 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.
79. 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.
80. 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.

81. 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.
82. Feigelman R, Kahlert CR, Baty F, Rassouli F, Kleiner RL, Kohler P, et al. Sputum DNA sequencing in cystic fibrosis: non-invasive access to the lung microbiome and to pathogen details. Microbiome. 2017;5(1):20.
83. Khalid-de Bakker C, Jonkers D, Smits K, Mesters I, Masclee A, Stockbrügger R. Participation in colorectal cancer screening trials after first-time invitation: a systematic review. Endoscopy. 2011;43(12):1059-86.
84. Gulsvik A, Humerfelt S, Bakke PS, Omenaas ER, Lehmann S. Norwegian population surveys on respiratory health in adults: objectives, design, methods, quality controls and response rates. Clin Respir J. 2008;2 Suppl 1:10-25.
85. Abrahamsen R, Svendsen MV, Henneberger PK, Gundersen GF, Torén K, Kongerud J, et al. Non-response in a cross-sectional study of respiratory health in Norway. BMJ Open. 2016;6(1):e009912.
86. Lipman MC, Stobbs D, Madge S, Miller R, Johnson MA. Research bronchoscopies do not adversely affect HIV-infected individuals' future health-care decisions. Chest. 1998;114(1):284-90.
87. Mtunthama N, Malamba R, French N, Molyneux ME, Zijlstra EE, Gordon SB. Malawians permit research bronchoscopy due to perceived need for healthcare. J Med Ethics. 2008;34(4):303-7.
88. Kerrison S, Laws S, Cane M, Thompson A. The patient's experience of being a human subject. J R Soc Med. 2008;101(8):416-22.
89. Kye SH, Tashkin DP, Roth MD, Adams B, Nie WX, Mao JT. Recruitment strategies for a lung cancer chemoprevention trial involving ex-smokers. Contemp Clin Trials. 2009;30(5):464-72.
90. Schook RM, Postmus BB, van den Berg RM, Sutedja TG, Man de FS, Smit EF, et al. The finding of premalignant lesions is not associated with smoking cessation in chemoprevention study volunteers. J Thorac Oncol. 2010;5(8):1240-5.

91. Patel D, Akporobaro A, Chinyanganya N, Hackshaw A, Seale C, Spiro SG, et al. Attitudes to participation in a lung cancer screening trial: a qualitative study. Thorax. 2012;67(5):418-25.

92. Chudleigh J, Hoo AF, Ahmed D, Prasad A, Sheehan D, Francis J, et al. Positive parental attitudes to participating in research involving newborn screened infants with CF. J Cyst Fibros. 2013;12(3):234-40.

93. Systematic Reviews: Synthesis of Best Evidence for Clinical Decisions. Annals of Internal Medicine. 1997;126(5):376-80.

94. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. ACP J Club. 1995;123(3):A12-3.

95. Delhaes L, Monchy S, Fréalle E, Hubans C, Salleron J, Leroy S, et al. The airway microbiota in cystic fibrosis: a complex fungal and bacterial community--implications for therapeutic management. PLoS One. 2012;7(4):e36313.

96. Rudkjøbing VB, Thomsen TR, Alhede M, Kragh KN, Nielsen PH, Johansen UR, et al. True microbiota involved in chronic lung infection of cystic fibrosis patients found by culturing and 16S rRNA gene analysis. J Clin Microbiol. 2011;49(12):4352-5.

97. Boutin S, Graeber SY, Weitnauer M, Panitz J, Stahl M, Clausznitzer D, et al. Comparison of microbiomes from different niches of upper and lower airways in children and adolescents with cystic fibrosis. PLoS One. 2015;10(1):e0116029.

98. Huang YJ, Nelson CE, Brodie EL, Desantis TZ, Baek MS, Liu J, et al. Airway microbiota and bronchial hyperresponsiveness in patients with suboptimally controlled asthma. J Allergy Clin Immunol. 2011;127(2):372-81.e1-3.

99. Garzoni C, Brugger SD, Qi W, Wasmer S, Cusini A, Dumont P, et al. Microbial communities in the respiratory tract of patients with interstitial lung disease. Thorax. 2013;68(12):1150-6.

100. Charlson ES, Diamond JM, Bittinger K, Fitzgerald AS, Yadav A, Haas AR, et al. Lung-enriched organisms and aberrant bacterial and fungal respiratory microbiota after lung transplant. Am J Respir Crit Care Med. 2012;186(6):536-45.

101. Huang YJ, Kim E, Cox MJ, Brodie EL, Brown R, Wiener-Kronish JP, et al. A persistent and diverse airway microbiota present during chronic obstructive pulmonary disease exacerbations. Omics. 2010;14(1):9-59.
102. 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.
103. Millares L, Ferrari R, Gallego M, Garcia-Nuñez M, Pérez-Brocal V, Espasa M, et al. Bronchial microbiome of severe COPD patients colonised by Pseudomonas aeruginosa. Eur J Clin Microbiol Infect Dis. 2014;33(7):1101-11.
104. Aguirre E, Galiana A, Mira A, Guardiola R, Sánchez-Guillén L, Garcia-Pachon E, et al. Analysis of microbiota in stable patients with chronic obstructive pulmonary disease. Apmis. 2015;123(5):427-32.
105. Wang Z, Bafadhel M, Haldar K, Spivak A, Mayhew D, Miller BE, et al. Lung microbiome dynamics in COPD exacerbations. European Respiratory Journal. 2016;47(4):1082-92.
106. Garcia-Nuñez M, Millares L, Pomares X, Ferrari R, Pérez-Brocal V, Gallego M, et al. Severity-related changes of bronchial microbiome in chronic obstructive pulmonary disease. J Clin Microbiol. 2014;52(12):4217-23.
107. Millares L, Pérez-Brocal V, Ferrari R, Gallego M, Pomares X, García-Núñez M, et al. Functional Metagenomics of the Bronchial Microbiome in COPD. PLoS One. 2015;10(12):e0144448.
108. Su J, Liu H-y, Tan X-l, Ji Y, Jiang Y-x, Prabhakar M, et al. Sputum Bacterial and Fungal Dynamics during Exacerbations of Severe COPD. PLOS ONE. 2015;10(7):e0130736.
109. Galiana A, Aguirre E, Rodriguez JC, Mira A, Santibañez M, Candela I, et al. Sputum microbiota in moderate versus severe patients with COPD. Eur Respir J. 2014;43(6):1787-90.

110. 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.

111. Segal LN, Clemente JC, Wu BG, Wikoff WR, Gao Z, Li Y, et al. Randomised, double-blind, placebo-controlled trial with azithromycin selects for anti-inflammatory microbial metabolites in the emphysematous lung. Thorax. 2017;72(1):13-22.

112. Engel M, Endesfelder D, Schloter-Hai B, Kublik S, Granitsiotis MS, Boschetto P, et al. Influence of lung CT changes in chronic obstructive pulmonary disease (COPD) on the human lung microbiome. PLoS One. 2017;12(7):e0180859.

113. Cui L, Lucht L, Tipton L, Rogers MB, Fitch A, Kessinger C, et al. Topographic diversity of the respiratory tract mycobiome and alteration in HIV and lung disease. Am J Respir Crit Care Med. 2015;191(8):932-42.

114. Ghannoum M. The Mycobiome. The Scientist. 2016.

115. Moyes DL, Naglik JR. The mycobiome: influencing IBD severity. Cell Host Microbe. 2012;11(6):551-2.

116. Ghannoum MA, Jurevic RJ, Mukherjee PK, Cui F, Sikaroodi M, Naqvi A, et al. Characterization of the oral fungal microbiome (mycobiome) in healthy individuals. PLoS Pathog. 2010;6(1):e1000713.

117. Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, et al. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc Natl Acad Sci U S A. 2012;109(16):6241-6.

118. Tang J, Iliev ID, Brown J, Underhill DM, Funari VA. Mycobiome: Approaches to analysis of intestinal fungi. J Immunol Methods. 2015;421:112-21.

119. Seed PC. The human mycobiome. Cold Spring Harb Perspect Med. 2014;5(5):a019810.

120. Noverr MC, Falkowski NR, McDonald RA, McKenzie AN, Huffnagle GB. Development of allergic airway disease in mice following antibiotic therapy and fungal microbiota increase: role of host genetics, antigen, and interleukin-13. Infect Immun. 2005;73(1):30-8.

121. Morris A, Hillenbrand M, Finkelman M, George MP, Singh V, Kessinger C, et al. Serum (1→3)-β-D-glucan levels in HIV-infected individuals are associated with immunosuppression, inflammation, and cardiopulmonary function. J Acquir Immune Defic Syndr. 2012;61(4):462-8.

122. Köhler JR, Casadevall A, Perfect J. The spectrum of fungi that infects humans. Cold Spring Harb Perspect Med. 2014;5(1):a019273.

123. Prohic A, Jovovic Sadikovic T, Krupalija-Fazlic M, Kuskunovic-Vlahovljak S. Malassezia species in healthy skin and in dermatological conditions. Int J Dermatol. 2016;55(5):494-504.

124. Kalan L, Loesche M, Hodkinson BP, Heilmann K, Ruthel G, Gardner SE, et al. Redefining the Chronic-Wound Microbiome: Fungal Communities Are Prevalent, Dynamic, and Associated with Delayed Healing. mBio. 2016;7(5).

125. Sokol H, Leducq V, Aschard H, Pham HP, Jegou S, Landman C, et al. Fungal microbiota dysbiosis in IBD. Gut. 2017;66(6):1039-48.

126. Yang AM, Inamine T, Hochrath K, Chen P, Wang L, Llorente C, et al. Intestinal fungi contribute to development of alcoholic liver disease. J Clin Invest. 2017;127(7):2829-41.

127. Willger SD, Grim SL, Dolben EL, Shipunova A, Hampton TH, Morrison HG, et al. Characterization and quantification of the fungal microbiome in serial samples from individuals with cystic fibrosis. Microbiome. 2014;2:40.

128. van Woerden HC, Gregory C, Brown R, Marchesi JR, Hoogendoorn B, Matthews IP. Differences in fungi present in induced sputum samples from asthma patients and non-atopic controls: a community based case control study. BMC Infect Dis. 2013;13:69.

129. Bousbia S, Papazian L, Saux P, Forel JM, Auffray JP, Martin C, et al. Repertoire of intensive care unit pneumonia microbiota. PLoS One. 2012;7(2):e32486.

130. Bittinger K, Charlson ES, Loy E, Shirley DJ, Haas AR, Laughlin A, et al. Improved characterization of medically relevant fungi in the human respiratory tract using next-generation sequencing. Genome Biol. 2014;15(10):487.

131. Kramer R, Sauer-Heilborn A, Welte T, Guzman CA, Abraham WR, Höfle MG. Cohort Study of Airway Mycobiome in Adult Cystic Fibrosis Patients: Differences in Community Structure between Fungi and Bacteria Reveal Predominance of Transient Fungal Elements. J Clin Microbiol. 2015;53(9):2900-7.
132. Krause R, Halwachs B, Thallinger GG, Klymiuk I, Gorkiewicz G, Hoenigl M, et al. Characterisation of Candida within the Mycobiome/Microbiome of the Lower Respiratory Tract of ICU Patients. PLoS One. 2016;11(5):e0155033.
133. Nguyen LD, Deschaght P, Merlin S, Loywick A, Audebert C, Van Daele S, et al. Effects of Propidium Monoazide (PMA) Treatment on Mycobiome and Bacteriome Analysis of Cystic Fibrosis Airways during Exacerbation. PLoS One. 2016;11(12):e0168860.
134. Botterel F, Angebault C, Cabaret O, Stressmann FA, Costa JM, Wallet F, et al. Fungal and Bacterial Diversity of Airway Microbiota in Adults with Cystic Fibrosis: Concordance Between Conventional Methods and Ultra-Deep Sequencing, and Their Practical use in the Clinical Laboratory. Mycopathologia. 2018;183(1):171-83.
135. Nguyen LD, Viscogliosi E, Delhaes L. The lung mycobiome: an emerging field of the human respiratory microbiome. Front Microbiol. 2015;6:89.
136. Ellepola AN, Samaranayake LP. Inhalational and topical steroids, and oral candidosis: a mini review. Oral Dis. 2001;7(4):211-6.
137. Shipley TW, Kling HM, Morris A, Patil S, Kristoff J, Guyach SE, et al. Persistent pneumocystis colonization leads to the development of chronic obstructive pulmonary disease in a nonhuman primate model of AIDS. J Infect Dis. 2010;202(2):302-12.
138. Christensen PJ, Preston AM, Ling T, Du M, Fields WB, Curtis JL, et al. Pneumocystis murina infection and cigarette smoke exposure interact to cause increased organism burden, development of airspace enlargement, and pulmonary inflammation in mice. Infect Immun. 2008;76(8):3481-90.
139. Morris A, Alexander T, Radhi S, Lucht L, Sciurba FC, Kolls JK, et al. Airway obstruction is increased in pneumocystis-colonized human immunodeficiency virus-infected outpatients. J Clin Microbiol. 2009;47(11):3773-6.

140. Grønseth R, Haaland I, Wiker HG, Martinsen EM, Leiten EO, Husebø G, et al. The Bergen COPD microbiome study (MicroCOPD): rationale, design, and initial experiences. Eur Clin Respir J. 2014;1.

141. PubMed. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information. https://pubmed.ncbi.nlm.nih.gov/ Last accessed: October 10, 2020.

142. Elsevier. Embase. https://www.elsevier.com/solutions/embase-biomedical-research Last accessed: October 10, 2020.

143. MeSH Database. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information. http://www.ncbi.nlm.nih.gov/mesh Last accessed: October 10, 2020.

144. Chang AA, Heskett KM, Davidson TM. Searching the literature using medical subject headings versus text word with PubMed. Laryngoscope. 2006;116(2):336-40.

145. Eagan TM, Ueland T, Wagner PD, Hardie JA, Mollnes TE, Damås JK, et al. Systemic inflammatory markers in COPD: results from the Bergen COPD Cohort Study. Eur Respir J. 2010;35(3):540-8.

146. Sørheim IC, Johannessen A, Grydeland TB, Omenaas ER, Gulsvik A, Bakke PS. Case-control studies on risk factors for chronic obstructive pulmonary disease: how does the sampling of the cases and controls affect the results? Clin Respir J. 2010;4(2):89-96.

147. Anthonisen NR, Manfreda J, Warren CP, Hershfield ES, Harding GK, Nelson NA. Antibiotic therapy in exacerbations of chronic obstructive pulmonary disease. Ann Intern Med. 1987;106(2):196-204.

148. Jones PW, Harding G, Berry P, Wiklund I, Chen WH, Kline Leidy N. Development and first validation of the COPD Assessment Test. Eur Respir J. 2009;34(3):648-54.

149. Mahler DA, Wells CK. Evaluation of clinical methods for rating dyspnea. Chest. 1988;93(3):580-6.

150. Borg, G., and Borg, E. (2010) The Borg CR Scales® folder. Hasselby, Sweden, Borg Perception.

151. Johannessen A, Lehmann S, Omenaas ER, Eide GE, Bakke PS, Gulsvik A. Post-bronchodilator spirometry reference values in adults and implications for disease management. Am J Respir Crit Care Med. 2006;173(12):1316-25.
152. Hoang T, Wiker H, Eagan TML, Drengenes C (2019). protocols.io. 16S Amplicon PCR for the V3-V4 region for the MicroCOPD samples. https://dx.doi.org/10.17504/protocols.io.2sygefw Last accessed: March 3, 2021.
153. Hoang T, Wiker H, Eagan TML, Drengenes C. 16S Amplicon PCR for the V3-V4 region for the MicroCOPD samples. Published protocol 2019; version 1, 1-16.
154. Usyk M, Zolnik CP, Patel H, Levi MH, Burk RD. Novel ITS1 Fungal Primers for Characterization of the Mycobiome. mSphere. 2017;2(6).
155. 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.
156. Bushnell B, Rood J, Singer E. BBMerge – Accurate paired shotgun read merging via overlap. PLOS ONE. 2017;12(10):e0185056.
157. Rivers AR, Weber KC, Gardner TG, Liu S, Armstrong SD. ITSxpress: Software to rapidly trim internally transcribed spacer sequences with quality scores for marker gene analysis. F1000Res. 2018;7:1418.
158. 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.
159. Callahan B. DADA2 ITS Pipeline Workflow (1.8). https://benjjneb.github.io/dada2/ITS_workflow.html Last accessed: February 21, 2021.
160. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, et al. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome. 2018;6(1):90.
161. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. Journal of machine learning research. 2011;12(Oct):2825–2830.

162. UNITE Community (2019): UNITE QIIME release for Fungi. UNITE Community. 10.15156/BIO/786334.

163. Zhang Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA sequences. J Comput Biol. 2000;7(1-2):203-14.

164. Nucleotide. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/nucleotide/ Last accessed: January 5, 2021.

165. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2018;46(D1):D8-d13.

166. Batch Entrez. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/sites/batchentrez Last accessed: October 7, 2020.

167. UNITE Community (2019): UNITE QIIME release for Fungi 2. UNITE Community. 10.15156/BIO/786349.

168. UNITE Community (2019): UNITE QIIME release for eukaryotes. UNITE Community. 10.15156/BIO/786335.

169. UNITE Community (2019): UNITE QIIME release for eukaryotes 2. UNITE Community. 10.15156/BIO/786350.

170. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10(1):421.

171. Frøslev TG, Kjøller R, Bruun HH, Ejrnæs R, Brunbjerg AK, Pietroni C, et al. Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. Nat Commun. 2017;8(1):1188.

172. Weaver D, Gago S, Bromley M, Bowyer P. The Human Lung Mycobiome in Chronic Respiratory Disease: Limitations of Methods and Our Current Understanding. Current Fungal Infection Reports. 2019;13(3):109-19.

173. 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.

174. Drengenes C, Wiker HG, Kalananthan T, Nordeide E, Eagan TML, Nielsen R. Laboratory contamination in airway microbiome studies. BMC Microbiol. 2019;19(1):187.

175. Tiew PY, Dicker AJ, Keir HR, Poh ME, Pang SL, Mac Aogáin M, et al. A high-risk airway mycobiome is associated with frequent exacerbation and mortality in COPD. Eur Respir J 2020; in press.

176. StataCorp. 2015. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP.

177. StataCorp. 2017. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC.

178. Wickham H, Hester J, Francois R (2018). readr: Read Rectangular Text Data. R package version 1.3.1. https://CRAN.R-project.org/package=readr Last accessed: October 9, 2020.

179. Firke S (2020). janitor: Simple Tools for Examining and Cleaning Dirty Data. R package version 2.0.1. https://CRAN.R-project.org/package=janitor Last accessed: October 9, 2020.

180. Müller K, Wickham H (2020). tibble: Simple Data Frames. R package version 3.0.1. https://CRAN.R-project.org/package=tibble Last accessed: October 9, 2020.

181. Mahto A, Jota, Morton E (2020). SOfun: Functions From Answers to R Questions on Stack Overflow. http://mrdwab.github.io/SOfun Last accessed: October 9, 2020.

182. Wickham H, François R, Henry L, Müller K (2020). dplyr: A Grammar of Data Manipulation. R package version 1.0.0. https://CRAN.R-project.org/package=dplyr Last accessed: October 9, 2020.

183. Wickham H, Henry L (2020). tidyr: Tidy Messy Data. R package version 1.1.0. https://CRAN.R-project.org/package=tidyr Last accessed: October 9, 2020.

184. Friedman J, Hastie T, Tibshirani R (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, 33(1), 1-22. URL http://www.jstatsoft.org/v33/i01/ Last accessed: October 9, 2020.

185. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
186. Kassambara A (2020). ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.3.0. https://CRAN.R-project.org/package=ggpubr Last accessed: October 9, 2020.
187. Aphalo PJ (2019). gginnards: Explore the Innards of 'ggplot2' Objects. R package version 0.0.3. https://CRAN.R-project.org/package=gginnards Last accessed: October 9, 2020.
188. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. (2019). vegan: Community Ecology Package. R package version 2.5-6. https://CRAN.R-project.org/package=vegan Last accessed: October 9, 2020.
189. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 2019;35(3):526-8.
190. Chen J, King E, Deek R, Wei Z, Yu Y, Grill D, et al. An omnibus test for differential distribution analysis of microbiome sequencing data. Bioinformatics. 2018;34(4):643-51.
191. Mandal S, Van Treuren W, White RA, Eggesbø M, Knight R, Peddada SD. Analysis of composition of microbiomes: a novel method for studying microbial composition. Microb Ecol Health Dis. 2015;26:27663.
192. Fernandes AD, Reid JN, Macklaim JM, McMurrough TA, Edgell DR, Gloor GB. Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. Microbiome. 2014;2:15.
193. Gloor GB, Macklaim JM, Fernandes AD. Displaying Variation in Large Datasets: Plotting a Visual Summary of Effect Sizes. Journal of Computational and Graphical Statistics. 2016;25(3):971-9.
194. Fernandes AD, Macklaim JM, Linn TG, Reid G, Gloor GB. ANOVA-Like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq. PLOS ONE. 2013;8(7):e67019.

195. Yue JC, Clayton MK. A Similarity Measure Based on Species Proportions. Communications in Statistics - Theory and Methods. 2005;34(11):2123-31.
196. Wickham H, Averick M, Bryan J, Chang W, McGowan LDA, François R (2019). Welcome to the Tidyverse. Journal of Open Source Software, 4(43), 1686.
197. Wickham H. Reshaping Data with the reshape Package. Journal Of Statistical Software. 2007;21:1-20.
198. Ammar R (2019). randomcoloR: Generate Attractive Random Colors. R package version 1.1.0.1. https://CRAN.R-project.org/package=randomcoloR Last accessed: October 10, 2020.
199. Kassambara A. ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.3.0. https://CRAN.R-project.org/package=ggpubr Last accessed: February 23, 2021.
200. StataCorp. 2019. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC.
201. Bokulich NA, Dillon MR, Zhang Y, Rideout JR, Bolyen E, Li H, et al. q2-longitudinal: Longitudinal and Paired-Sample Analyses of Microbiome Data. mSystems. 2018;3(6).
202. Silberzahn R, Uhlmann EL, Martin DP, Anselmi P, Aust F, Awtrey E, et al. Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results. Advances in Methods and Practices in Psychological Science. 2018;1(3):337-56.
203. Parab S, Bhalerao S. Study designs. Int J Ayurveda Res. 2010;1(2):128-31.
204. Ranganathan P, Aggarwal R. Study designs: Part 1 - An overview and classification. Perspect Clin Res. 2018;9(4):184-6.
205. Heale R, Twycross A. Validity and reliability in quantitative studies. Evid Based Nurs. 2015;18(3):66-7.
206. Tripepi G, Jager KJ, Dekker FW, Zoccali C. Selection bias and information bias in clinical research. Nephron Clin Pract. 2010;115(2):c94-9.

207. Catalogue of bias collaboration. Bankhead CR, Spencer EA, Nunan D. Information bias. In: Sackett Catalogue Of Biases 2019. https://catalogofbias.org/biases/information-bias/ Last accessed: February 16, 2021.
208. Thiese MS. Observational and interventional study design types; an overview. Biochem Med (Zagreb). 2014;24(2):199-210.
209. Thompson EA, Gann LB, Cressman ENK. Learning to successfully search the scientific and medical literature. Cell Stress Chaperones. 2019;24(2):289-93.
210. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6(7):e1000097.
211. Henderson LK, Craig JC, Willis NS, Tovey D, Webster AC. How to write a Cochrane systematic review. Nephrology (Carlton). 2010;15(6):617-24.
212. Lange P, Halpin DM, O'Donnell DE, MacNee W. Diagnosis, assessment, and phenotyping of COPD: beyond FEV_1. Int J Chron Obstruct Pulmon Dis. 2016;11 Spec Iss(Spec Iss):3-12.
213. Mannino DM, Sonia Buist A, Vollmer WM. Chronic obstructive pulmonary disease in the older adult: what defines abnormal lung function? Thorax. 2007;62(3):237-41.
214. Price DB, Yawn BP, Jones RC. Improving the differential diagnosis of chronic obstructive pulmonary disease in primary care. Mayo Clin Proc. 2010;85(12):1122-9.
215. Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, et al. Interpretative strategies for lung function tests. Eur Respir J. 2005;26(5):948-68.
216. Leiten EO, Eagan TML, Martinsen EMH, Nordeide E, Husebø GR, Knudsen KS, et al. Complications and discomfort after research bronchoscopy in the MicroCOPD study. BMJ Open Respiratory Research. 2020;7(1):e000449.
217. Leiten EO, Martinsen EM, Bakke PS, Eagan TM, Grønseth R. Complications and discomfort of bronchoscopy: a systematic review. Eur Clin Respir J. 2016;3:33324.

218. Minich JJ, Sanders JG, Amir A, Humphrey G, Gilbert JA, Knight R. Quantifying and Understanding Well-to-Well Contamination in Microbiome Research. mSystems. 2019;4(4):e00186-19.
219. Dickson RP, Erb-Downward JR, Freeman CM, McCloskey L, Falkowski NR, Huffnagle GB, et al. Bacterial Topography of the Healthy Human Lower Respiratory Tract. mBio. 2017;8(1):e02287-16.
220. Tangedal S, Aanerud M, Grønseth 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.
221. Czurda S, Smelik S, Preuner-Stix S, Nogueira F, Lion T. Occurrence of Fungal DNA Contamination in PCR Reagents: Approaches to Control and Decontamination. J Clin Microbiol. 2016;54(1):148-52.
222. Hanson B, Zhou Y, Bautista EJ, Urch B, Speck M, Silverman F, et al. Characterization of the bacterial and fungal microbiome in indoor dust and outdoor air samples: a pilot study. Environ Sci Process Impacts. 2016;18(6):713-24.
223. Halwachs B, Madhusudhan N, Krause R, Nilsson RH, Moissl-Eichinger C, Högenauer C, et al. Critical Issues in Mycobiota Analysis. Front Microbiol. 2017;8:180.
224. MacConaill LE, Burns RT, Nag A, Coleman HA, Slevin MK, Giorda K, et al. Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. BMC Genomics. 2018;19(1):30.
225. 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.
226. Kircher M, Sawyer S, Meyer M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res. 2012;40(1):e3.
227. Wright ES, Vetsigian KH. Quality filtering of Illumina index reads mitigates sample cross-talk. BMC Genomics. 2016;17(1):876.

228. Bowman SM, Free SJ. The structure and synthesis of the fungal cell wall. Bioessays. 2006;28(8):799-808.
229. Benito-León J, Laurence M. The Role of Fungi in the Etiology of Multiple Sclerosis. Frontiers in Neurology. 2017;8(535).
230. Dupuy AK, David MS, Li L, Heider TN, Peterson JD, Montano EA, et al. Redefining the human oral mycobiome with improved practices in amplicon-based taxonomy: discovery of Malassezia as a prominent commensal. PLoS One. 2014;9(3):e90899.
231. Pérez-Brocal V, Magne F, Ruiz-Ruiz S, Ponce CA, Bustamante R, Martin VS, et al. Optimized DNA extraction and purification method for characterization of bacterial and fungal communities in lung tissue samples. Sci Rep. 2020;10(1):17377.
232. 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.
233. Hoggard M, Vesty A, Wong G, Montgomery JM, Fourie C, Douglas RG, et al. Characterizing the Human Mycobiota: A Comparison of Small Subunit rRNA, ITS1, ITS2, and Large Subunit rRNA Genomic Targets. Front Microbiol. 2018;9:2208.
234. McTaggart LR, Copeland JK, Surendra A, Wang PW, Husain S, Coburn B, et al. Mycobiome Sequencing and Analysis Applied to Fungal Community Profiling of the Lower Respiratory Tract During Fungal Pathogenesis. Front Microbiol. 2019;10:512.
235. Nilsson RH, Anslan S, Bahram M, Wurzbacher C, Baldrian P, Tedersoo L. Mycobiome diversity: high-throughput sequencing and identification of fungi. Nat Rev Microbiol. 2019;17(2):95-109.
236. Bellemain E, Carlsen T, Brochmann C, Coissac E, Taberlet P, Kauserud H. ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. BMC Microbiol. 2010;10:189.
237. Ali N, Mac Aogáin M, Morales RF, Tiew PY, Chotirmall SH. Optimisation and Benchmarking of Targeted Amplicon Sequencing for Mycobiome Analysis of Respiratory Specimens. Int J Mol Sci. 2019;20(20).

238. Wang Y, LêCao KA. Managing batch effects in microbiome data. Brief Bioinform. 2020;21(6):1954-70.

239. Nilsson RH, Ryberg M, Abarenkov K, Sjökvist E, Kristiansson E. The ITS region as a target for characterization of fungal communities using emerging sequencing technologies. FEMS Microbiology Letters. 2009;296(1):97-101.

240. Rivers A (2018). Q2-ITSxpress: A tutorial on a QIIME 2 plugin to trim ITS sequences. https://forum.qiime2.org/t/q2-itsxpress-a-tutorial-on-a-qiime-2-plugin-to-trim-its-sequences/5780 Last accessed: February 19, 2021.

241. Schirmer M, D'Amore R, Ijaz UZ, Hall N, Quince C. Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data. BMC Bioinformatics. 2016;17:125.

242. Illumina. Quality Scores for Next-Generation Sequencing. https://www.illumina.com/documents/products/technotes/technote_Q-Scores.pdf Last accessed: February 21, 2021.

243. Bokulich N. Fungal ITS analysis tutorial. https://forum.qiime2.org/t/fungal-its-analysis-tutorial/7351 Last accessed: February 21, 2021.

244. Nearing JT, Douglas GM, Comeau AM, Langille MGI. Denoising the Denoisers: an independent evaluation of microbiome sequence error-correction approaches. PeerJ. 2018;6:e5364.

245. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome Datasets Are Compositional: And This Is Not Optional. Front Microbiol. 2017;8:2224.

246. Morton JT, Marotz C, Washburne A, Silverman J, Zaramela LS, Edlund A, et al. Establishing microbial composition measurement standards with reference frames. Nat Commun. 2019;10(1):2719.

247. McMurdie PJ, Holmes S. Waste not, want not: why rarefying microbiome data is inadmissible. PLoS Comput Biol. 2014;10(4):e1003531.

248. QIIME 2. “Moving Pictures” tutorial. https://docs.qiime2.org/2021.2/tutorials/moving-pictures/ Last accessed March 3, 2021.

249. Diaz PI, Hong BY, Dupuy AK, Strausbaugh LD. Mining the oral mycobiome: Methods, components, and meaning. Virulence. 2017;8(3):313-23.
250. Zhang I, Pletcher SD, Goldberg AN, Barker BM, Cope EK. Fungal Microbiota in Chronic Airway Inflammatory Disease and Emerging Relationships with the Host Immune Response. Front Microbiol. 2017;8:2477.
251. Irinyi L, Serena C, Garcia-Hermoso D, Arabatzis M, Desnos-Ollivier M, Vu D, et al. International Society of Human and Animal Mycology (ISHAM)-ITS reference DNA barcoding database--the quality controlled standard tool for routine identification of human and animal pathogenic fungi. Med Mycol. 2015;53(4):313-37.
252. Nilsson RH, Ryberg M, Kristiansson E, Abarenkov K, Larsson KH, Kõljalg U. Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. PLoS One. 2006;1(1):e59.
253. Prakash PY, Irinyi L, Halliday C, Chen S, Robert V, Meyer W. Online Databases for Taxonomy and Identification of Pathogenic Fungi and Proposal for a Cloud-Based Dynamic Data Network Platform. Journal of Clinical Microbiology. 2017;55(4):1011-24.
254. Lavrinienko A, Jernfors T, Koskimäki JJ, Pirttilä AM, Watts PC. Does Intraspecific Variation in rDNA Copy Number Affect Analysis of Microbial Communities? Trends Microbiol. 2021;29(1):19-27.
255. Ramette A. Multivariate analyses in microbial ecology. FEMS Microbiol Ecol. 2007;62(2):142-60.
256. Sinha R, Abnet CC, White O, Knight R, Huttenhower C. The microbiome quality control project: baseline study design and future directions. Genome Biology. 2015;16(1):276.
257. Amos GCA, Logan A, Anwar S, Fritzsche M, Mate R, Bleazard T, et al. Developing standards for the microbiome field. Microbiome. 2020;8(1):98.
258. Murthy VH, Krumholz HM, Gross CP. Participation in Cancer Clinical TrialsRace-, Sex-, and Age-Based Disparities. JAMA. 2004;291(22):2720-6.
259. Lee E, Wen P. Gender and sex disparity in cancer trials. ESMO Open. 2020;5(Suppl 4).

260. Harris DJ, Douglas PS. Enrollment of women in cardiovascular clinical trials funded by the National Heart, Lung, and Blood Institute. N Engl J Med. 2000;343(7):475-80.
261. Tandon A, Murray C, Lauer J, Evans D. Measuring Overall Health System Performance for 191 Countries. Global Programme on Evidence forHealth Policy Discussion Paper No 30. 2000.
262. McCallum AD, Nyirenda D, Lora W, Khoo SH, Sloan DJ, Mwandumba HC, et al. Perceptions of Research Bronchoscopy in Malawian Adults with Pulmonary Tuberculosis: A Cross-Sectional Study. PLoS One. 2016;11(10):e0165734.
263. Gargoum FS, O'Keeffe ST. Readability and content of patient information leaflets for endoscopic procedures. Ir J Med Sci. 2014;183(3):429-32.
264. Kincaid J, Fishburne RP, Rogers RL. Chissom BS. Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel. 1975.
265. Fujimoto K, Ishiwata T, Kasai H, Terada J, Shionoya Y, Ikari J, et al. Identification of factors during bronchoscopy that affect patient reluctance to undergo repeat examination: Questionnaire analysis after initial bronchoscopy. PLoS One. 2018;13(12):e0208495.
266. Saxon C, Fulbrook P, Fong KM, Ski CF. High-risk respiratory patients' experiences of bronchoscopy with conscious sedation and analgesia: A qualitative study. J Clin Nurs. 2018;27(13-14):2740-51.
267. Andrychiewicz A, Konarska K, Gorka K, Bartyzel S, Salek M, Biedron G, et al. Evaluation of factors that influence anxiety and satisfaction in patients undergoing bronchofiberoscopy with analgosedation. Clin Respir J. 2017;11(5):566-73.
268. Pink KL, Woolley J, Ionescu AA. Does "patient-reported discomfort" from bronchoscopy differ over time? J Bronchology Interv Pulmonol. 2012;19(4):288-93.
269. Park JS, Ryu JS, Lee SM, Yim JJ, Yoo CG, Kim YW, et al. Influence of additional post-bronchoscopy visit on patient satisfaction after flexible bronchoscopy. Korean J Intern Med. 2010;25(4):392-8.

270. Ni YL, Lo YL, Lin TY, Fang YF, Kuo HP. Conscious sedation reduces patient discomfort and improves satisfaction in flexible bronchoscopy. Chang Gung Med J. 2010;33(4):443-52.
271. Steinfort DP, Irving LB. Patient satisfaction during endobronchial ultrasound-guided transbronchial needle aspiration performed under conscious sedation. Respir Care. 2010;55(6):702-6.
272. Bernasconi M, Chhajed PN, Müller P, Borer H. Patients' satisfaction with flexible bronchoscopy in a hospital-based community practice. Respiration. 2009;78(4):440-5.
273. Hirose T, Okuda K, Ishida H, Sugiyama T, Kusumoto S, Nakashima M, et al. Patient satisfaction with sedation for flexible bronchoscopy. Respirology. 2008;13(5):722-7.
274. Lechtzin N, Rubin HR, White P, Jr., Jenckes M, Diette GB. Patient satisfaction with bronchoscopy. Am J Respir Crit Care Med. 2002;166(10):1326-31.
275. Maguire GP, Rubinfeld AR, Trembath PW, Pain MC. Patients prefer sedation for fibreoptic bronchoscopy. Respirology. 1998;3(2):81-5.
276. Mitsumune T, Senoh E, Adachi M. Prediction of patient discomfort during fibreoptic bronchoscopy. Respirology. 2005;10(1):92-6.
277. Poi PJ, Chuah SY, Srinivas P, Liam CK. Common fears of patients undergoing bronchoscopy. Eur Respir J. 1998;11(5):1147-9.
278. Putinati S, Ballerin L, Corbetta L, Trevisani L, Potena A. Patient satisfaction with conscious sedation for bronchoscopy. Chest. 1999;115(5):1437-40.
279. Uzbeck M, Quinn C, Saleem I, Cotter P, Gilmartin JJ, O'Keeffe ST. Randomised controlled trial of the effect of standard and detailed risk disclosure prior to bronchoscopy on peri-procedure anxiety and satisfaction. Thorax. 2009;64(3):224-7.
280. Bodington R, Meghjee S, Thirumaran M, Faruqi S. Patient experience surrounding bronchoscopy. Pneumologia. 2016;65(3):134-7.
281. Hehn BT, Haponik E, Rubin HR, Lechtzin N, Diette GB. The relationship between age and process of care and patient tolerance of bronchoscopy. J Am Geriatr Soc. 2003;51(7):917-22.

282. McLean AN, Semple PA, Franklin DH, Petrie G, Millar EA, Douglas JG. The Scottish multi-centre prospective study of bronchoscopy for bronchial carcinoma and suggested audit standards. Respir Med. 1998;92(9):1110-5.
283. López FJR, del Mar Valdivia Salas M, Pérez JL, Lucas JAR, Suarez BF, Gascón FS, et al. Flexible Bronchoscopy With Only Topical Anesthesia. Journal of Bronchology & Interventional Pulmonology. 2006;13(2):54-7.
284. Dickson RP, Erb-Downward JR, Huffnagle GB. Towards an ecology of the lung: new conceptual models of pulmonary microbiology and pneumonia pathogenesis. Lancet Respir Med. 2014;2(3):238-46.
285. Berger G, Wunderink RG. Lung microbiota: genuine or artifact? Isr Med Assoc J. 2013;15(12):731-3.
286. Charlson ES, Bittinger K, Chen J, Diamond JM, Li H, Collman RG, et al. Assessing bacterial populations in the lung by replicate analysis of samples from the upper and lower respiratory tracts. PLoS One. 2012;7(9):e42786.
287. Dickson RP, Erb-Downward JR, Freeman CM, Walker N, Scales BS, Beck JM, et al. Changes in the lung microbiome following lung transplantation include the emergence of two distinct Pseudomonas species with distinct clinical associations. PLoS One. 2014;9(5):e97214.
288. Mukherjee PK, Chandra J, Retuerto M, Sikaroodi M, Brown RE, Jurevic R, et al. Oral mycobiome analysis of HIV-infected patients: identification of Pichia as an antagonist of opportunistic fungi. PLoS Pathog. 2014;10(3):e1003996.
289. Fraczek MG, Chishimba L, Niven RM, Bromley M, Simpson A, Smyth L, et al. Corticosteroid treatment is associated with increased filamentous fungal burden in allergic fungal disease. J Allergy Clin Immunol. 2018;142(2):407-14.
290. Sanguinetti M, Posteraro B, Lass-Flörl C. Antifungal drug resistance among Candida species: mechanisms and clinical impact. Mycoses. 2015;58 Suppl 2:2-13.
291. Lagunes L, Rello J. Invasive candidiasis: from mycobiome to infection, therapy, and prevention. Eur J Clin Microbiol Infect Dis. 2016;35(8):1221-6.

292. Theelen B, Cafarchia C, Gaitanis G, Bassukas ID, Boekhout T, Dawson TL, Jr. Malassezia ecology, pathophysiology, and treatment. Med Mycol. 2018;56(suppl_1):S10-s25.
293. Bougnoux ME, Brun S, Zahar JR. Healthcare-associated fungal outbreaks: New and uncommon species, New molecular tools for investigation and prevention. Antimicrob Resist Infect Control. 2018;7:45.
294. Júnior MC, de Moraes Arantes A, Silva HM, Costa CR, Silva Mdo R. Acremonium kiliense: case report and review of published studies. Mycopathologia. 2013;176(5-6):417-21.
295. Gottlieb MS, Schroff R, Schanker HM, Weisman JD, Fan PT, Wolf RA, et al. Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. N Engl J Med. 1981;305(24):1425-31.
296. Morris A, Kingsley LA, Groner G, Lebedeva IP, Beard CB, Norris KA. Prevalence and clinical predictors of Pneumocystis colonization among HIV-infected men. Aids. 2004;18(5):793-8.
297. Morris A, Sciurba FC, Lebedeva IP, Githaiga A, Elliott WM, Hogg JC, et al. Association of chronic obstructive pulmonary disease severity and Pneumocystis colonization. Am J Respir Crit Care Med. 2004;170(4):408-13.
298. Bafadhel M, McKenna S, Agbetile J, Fairs A, Desai D, Mistry V, et al. Aspergillus fumigatus during stable state and exacerbations of COPD. Eur Respir J. 2014;43(1):64-71.
299. Standiford TJ, Kunkel SL, Strieter RM. Elevated serum levels of tumor necrosis factor-alpha after bronchoscopy and bronchoalveolar lavage. Chest. 1991;99(6):1529-30.
300. Krause A, Hohberg B, Heine F, John M, Burmester GR, Witt C. Cytokines derived from alveolar macrophages induce fever after bronchoscopy and bronchoalveolar lavage. Am J Respir Crit Care Med. 1997;155(5):1793-7.

301. van Rijn AL, van Boheemen S, Sidorov I, Carbo EC, Pappas N, Mei H, et al. The respiratory virome and exacerbations in patients with chronic obstructive pulmonary disease. PLoS One. 2019;14(10):e0223952.
302. Krüger W, Vielreicher S, Kapitan M, Jacobsen ID, Niemiec MJ. Fungal-Bacterial Interactions in Health and Disease. Pathogens. 2019;8(2).
303. Tiew PY, Ko FWS, Pang SL, Matta SA, Sio YY, Poh ME, et al. Environmental fungal sensitisation associates with poorer clinical outcomes in COPD. Eur Respir J. 2020;56(2).

16 Appendices and papers

16.1 Appendix A. Questionnaire, MicroCOPD study

ID number, MicroCOPD: _____

Recruitment form and personal information. MicroCOPD.

First name, last name	*1.1*		Sex	*1.1a*	
Norwegian national identity number	*1.2a, 1.2b*				
ID number BergenCOPD	*1.3*	GeneCOPD ID	*1.3a*	ID, MicroCOPD	*1.4*
Recruitment source	*1.5*				
Address, postcode and place	*1.6a, 1.6b, 1.6c*				
Telephone number	*1.7a, 1.7b, 1.7c, 1.7d*				

Re-bronchoscopy (tick): *1.8*

Participant group (tick):

COPD	**Control, never-smoke**	**Control, smoke**	**Asthma**	**Other** *1.9*

Critical information regarding recruitment

Informed about fast	*1.10*
Informed about non-ability to drive if sedation is given	*1.11*
Questioned about anticoagulants, dual antiplatelet therapy	*1.12*
Questioned about artificial heart valve	*1.13*
Questioned about antibiotic usage, steroids and exacerbation	*1.14*

Bronchoscopic sampling completed?

Sample	**Tick if completed**	**Note what is missing, and reason**
Brushes*2	*1.32*	*1.28*
SVL	*1.33*	*1.29*
BAL	*1.34*	*1.30*
Biopsies	*1.35*	*1.31*

Plotted and controlled, date and signature: _____ *1.27a, 1.27b*

ID number, MicroCOPD: _____

Contraindications.

	Yes	No
BLEEDING RISK		
Known haemophilia		2.1
Blood samples		2.2
tpc < $75*10^9$		2.3
INR > 2,0		2.4
Anticoagulants (Marevan, Pradaxa, Xarelto, Eliquis)		2.5
Dual antiplatelet therapy, or clopidogrel/plavix/ticagrelor/brilique last 5 days.		2.6
Low molecular weight heparin last 24 hours.		2.7
OTHER		
Artificial heart valve		2.8
Myocardial infarction, acute coronary syndrome/unstable angina last 6 weeks		2.9
Known severe pulmonary hypertension		2.10
SpO2 below 90%, with supplemental oxygen		2.11
STABLE COPD?		
Hospitalised for COPD last 2 weeks		2.12
Antibiotic usage last 2 weeks		2.13
Oral glucocorticoids last 2 weeks		2.14
Ongoing exacerbation – 2 major or 1 major + 1 minor, 2 subsequent days (relative)		
MAJ: Increased dyspnoea		2.15
MAJ: Increased sputum		2.16
MAJ: Colour change sputum		2.17
MIN: Stuffed/runny nose		2.18
MIN: Increased cough or sore throat		2.19
MIN: Asthenia		2.20
MIN: Increased wheezing sounds from the chest		2.21
THE PATIENT INTENDS TO DRIVE HOME AFTER THE PROCEDURE		2.22
FAST NOT COMPLETED		2.23

In the case of any "yes", the project physician is contacted for individual consideration. If the procedure is conducted despite "yes", the reason is documented here:

2.24

_____

ID number, MicroCOPD: _____

Medication and vaccination:

Note all drugs the patient uses, both as-needed and regular medication. No medication: *3.0.*

Medication name	Administration form	Dose (unit)	Dosage (*number of doses per 24 hours, or B for as-needed*)	Last dosage given (time in last 24h)	Start date if cure
3.1a	*3.1b*	*3.1c, 3.1d*	*3.1e*	*3.1f*	*3.1g*
3.2a	*3.2b*	*3.2c, 3.2d*	*3.2e*	*3.2f*	*3.2g*
3.3a	*3.3b*	*3.3c, 3.3d*	*3.3e*	*3.3f*	*3.3g*
3.4a	*3.4b*	*3.4c, 3.4d*	*3.4e*	*3.4f*	*3.4g*
3.5a	*3.5b*	*3.5c, 3.5d*	*3.5e*	*3.5f*	*3.5g*
3.6a	*3.6b*	*3.6c, 3.6d*	*3.6e*	*3.6f*	*3.6g*
3.7a	*3.7b*	*3.7c, 3.7d*	*3.7e*	*3.7f*	*3.7g*
3.8a	*3.8b*	*3.8c, 3.8d*	*3.8e*	*3.8f*	*3.8g*
3.9a	*3.9b*	*3.9c, 3.9d*	*3.9e*	*3.9f*	*3.9g*
3.10a	*3.10b*	*3.10c, 3.10d*	*3.10e*	*3.10f*	*3.10g*
3.11a	*3.11b*	*3.11c, 3.11d*	*3.11e*	*3.11f*	*3.11g*

	Approximate month/year
When did you last receive a flu shot?	*3.12*
When did you last receive antibiotic treatment?	*3.13*
When was your latest cortisone/prednisolone cure?	*3.14*

ID number, MicroCOPD: _____

Conditions/diseases (active treatment, current symptoms, sequelae etc)

Disease Yes	Disease No	Diagnosis	When diagnosed (years since)
	4.1	Chronic obstructive pulmonary disease (COPD)	4.1a
	4.2	Emphysema	4.2a
	4.3	Chronic bronchitis	4.3a
	4.4	Asthma	4.4a
	4.5	Lung fibrosis	4.5a
	4.6	Cystic fibrosis	4.6a
	4.7	Sarcoidosis	4.7a
	4.8	Lung cancer	4.8a
	4.9	Tuberculosis	4.9a

Conditions in airways are to be verified through medical history taking/spirometry/medical records and are regarded as diagnoses given/verified at the time of examination.

CONDITION	Yes	No	*CONDITION*	Yes	No
Diabetes mellitus		4.10	Depression with regular use of medication		4.33
Myocardial infarction		4.11	Other psychiatric illness		4.34
Angina		4.12	which?		4.35
Intermittent claudication		4.13	Muscle disease with regular use of medication		4.36
Heart valve condition		4.14	which?		4.37
Heart failure		4.15	**Active known cancer (diagnosed/treated last 5 years)**		
Cerebral infarction or bleeding		4.16	Lung cancer		4.38
Other known neurological disease		4.17	GI cancer		4.39
which?		4.18	Breast cancer		4.40
Gastric ulcer		4.19	Endometrial cancer (NOT dysplasia only)		4.41
Hepatic disease		4.20	Cancer in gonads (testes/ovaries)		4.42
which?		4.21	Prostate cancer		4.43
Kidney disease		4.22	Blood cancer, leukaemia		4.44
which?		4.23	Lymphoma		4.45
High blood pressure, treated hypertension		4.24	Skin cancer (not including treated basalioma)		4.46
			Other type of cancer		4.47
Inflammatory diseases in need of therapy			which		4.49
Rheumatoid arthritis		4.26	**Other diseases (active treatment, physician-given diagnose) – write here:**		
Psoriasis arthritis		4.27			4.50
Systemic lupus erythematosus		4.28			4.51
Polymyalgia rheumatica		4.29			4.52
Ulcerous colitis/Mb Crohn		4.30			4.53
Disease in skeleton or joints with regular use of medication, including osteoporosis		4.31			4.54
Which?		4.32			4.55

Comorbidities should, to the greatest extent possible, be verified through medical history taking or medical record review.

ID number, MicroCOPD: _____

Marital status, children, education, menopause, domestic animals

1. Are you (one tick):		Married/Registered partner	5.1a
		Widow/widower	5.1b
		Cohabitant	5.1c
		Divorced, live alone	5.1d
		Unmarried/single	5.1e
2. If you have children, how many?		*number*	5.2
3. Which education level best suits you?		Compulsory education	5.3a
		High school/vocational training	5.3b
		3 years of higher education/university	5.3c
		≥4 years of higher education/university	5.3d
4. For women, do you still experience regular periods?	Yes	No	5.4
4. a. If no, when did you reach menopause?	*years ago*		5.5
5. Do you keep domestic animals or birds?	Yes	No	5.6
5. b. Which domestic animal(s)/bird(s) do you have?			5.6a1 – 5.6a4
5. c. Have you kept domestic animals/birds at home before?	Yes	No	5.6b
5. d Which domestic animal(s)/bird(s) did you have before?			5.6c1 - 5.6c4

Arterial blood gas and pulse oximetry

	Yes	No	
Does the patient receive continuous oxygen			5.7
Oxygen supplied in the 30 minutes prior to puncture (litres/min)			5.8
Blood gas results			
FiO_2, oxygen fraction, room air = 0,21			5.9
pH			5.10
Oxygen tension (PaO_2, kPa)			5.11
Carbon acid tension ($PaCO_2$, kPa)			5.12
arterial saturation (%)			5.13
bicarbonate, mmol/l			5.14
carbon monoxide, %			5.15
haemoglobin, g/dl			5.16
Blood gas not performed because (perform pulse oximetry, note in bronchoscopy form)			
Refuses			5.17
> 6 attempts			5.18
Apparatus failure			5.19

ID number, MicroCOPD: _____

Lung function testing, height, weight

1. Chosen spirometer: _____ *6.1*

2. Technician (four-character code): ☐☐☐☐ *6.2*

3. Weight of participant (kg): ☐☐☐.☐ *6.3*

4. Height of participant (in whole cm) ☐☐☐ *6.4*

5. Given Ventolin? Yes No *6.5*

5.1 Time: ☐☐:☐☐ *6.6*

6. Time test start: ☐☐:☐☐ *6.7*

7.a Best FEV1 _____ (litres) *6.8* 7b. Best FEV_1 % of predicted_____ *6.9*

8.a Best FVC _____ (litres) *6.10* 8b. Best FVC % of predicted_____ *6.11*

9. Which reference values were applied? _____ *6.12*

ID number, MicroCOPD: _____

COPD exacerbations, smoking habits, alcohol

1. Number of exacerbations in the last 12 months requiring antibiotics/steroids or hospital admission?

_____ (number of exacerbations) *7.1*

If yes, answer question 2, if not, move on to question 3.

2. If one or more therapy requiring exacerbations in the last 12 months – how many of them required acute hospitalisation?

_____ (number of admissions) *7.2*

3a Do you smoke daily now? Yes No
If yes, answer 3b, if no move to 3d *7.3*

3b Do you smoke <u>cigarettes</u> daily? (roll-your-own or manufactured)? Yes No
If yes, move to 3f, if no move to 3c *7.4*

3c What do you use to smoke tobacco? pipe / cigar
Move to 3f *7.5*

3d Have you smoked daily before? Yes No
If yes move to 3e, if no move to 3i *7.6*

3e How long since you quit?
Less than three months
Between three months and one year
One to five years
More than five years *7.7*

3f How many years have you smoked daily? ⌊_⌊_⌋ *7.8* Number of years

3g How many cigarettes do you smoke or did you smoke daily? (give the number per day, both roll-your-own and manufactured) **Move to question 4.** ⌊_⌊_⌋ *7.9* Number of units

3i Do you smoke cigarettes *once in a while*, or have you smoked cigarettes once in a while before? **If yes, move to 3j and 3k, if not move to question 4.** Yes Before No *7.13*

3j For how long have you smoked *once in a while*? ⌊_⌊_⌋ *7.14* Number of years

3k How many cigarettes do/did you smoke in the course of a regular week? ⌊_⌊_⌋ *7.15* Number of units

4. Number of cigarettes or tobacco units (not snus) in the last 24 hours _____ number of units. *7.10*

5. Number of hours since you smoked or used tobacco _____ (number of hours) *7.11*

6. How many units of alcohol do you consume in the course of an average week?

_____ (number of alcohol units) *7.12*

ID number, MicroCOPD: _____

CAT COPD assessment test (copyright GSK):

For each item below, place a mark (X) in the box that best describes your current situation. Please ensure that you only select one response for each question

Example:

I am very happy	0	1	2	3	4	5	I am very sad

								SCORE
I never cough	0	1	2	3	4	5	I cough all the time	8.1
I have no phlegm (mucus) in my chest at all	0	1	2	3	4	5	My chest is full of phlegm (mucus)	8.2
My chest does not feel tight at all	0	1	2	3	4	5	My chest feels very tight	8.3
When I walk up a hill or a flight of stairs, I am not out of breath	0	1	2	3	4	5	When I walk up a hill or a flight of stairs, I am completely out of breath	8.4
I am not limited to doing any activities at home	0	1	2	3	4	5	I am completely limited to doing all activities at home	8.5
I am confident leaving my home despite my lung condition	0	1	2	3	4	5	I am not confident leaving my home at all because of my lung condition	8.6
I sleep soundly	0	1	2	3	4	5	I do not sleep soundly because of my lung condition	8.7
I have lots of energy	0	1	2	3	4	5	I have no energy at all	8.8
							TOTAL SCORE	8.9

Motivation for participation

Why did you wish to take part in this project (open question, answer in free-form text)?

8.10

Expectations

On a scale from zero to 10, where 10 is the worst you can imagine, and 0 is nothing. How much do you dread this examination? (whole numbers, **no comment**)

8.11

ID number, MicroCOPD: _____

Assessment of chronic dyspnoea (mMRC scale)

First
Are you restricted from walking due to other condition than breathlessness?

Yes ☐ No ☐ *9.0*

If yes, move pass the next question, if no:

Give the one answer that is correct for you (mark only *one*) 9.1

☐ I am too breathless to leave the house or I am breathless when dressing.

☐ I stop for breath after walking about 100 metres or after a few minutes on level ground

☐ On level ground, I walk slower than people of the same age because of breathlessness, or have to stop for breath when walking at my own pace.

☐ I get short of breath when hurrying on level ground or walking up a slight hill ☐ I only get breathless with strenuous exercise

Assessment of dyspnoea before bronchoscopy (Borg scale) *9.2*

How severe is your breathlessness?			Tick
0	Nothing at all	"No intensity"	
0,3			
0,5	Very, very slight	Just noticeable	
0,7			
1	Very slight		
1,5			
2	Slight	Light	
2,5			
3	Moderate		
4			
5	Severe	Heavy	
6			
7	Very severe		
8			
9			
10	Extremely severe	"Strongest intensity"	
11			
*	Absolute maximum	Highest possible	

Borg CR10 scale.

ID number, MicroCOPD: _____

Bronchoscopy date: ____/____/20___ 10.0

Safety									
Contraindications checked	10.1	Allergies, asked*	10.2	Blood samples date (safety)	10.2a				
Tpc *10^9	10.3	Hb, mg/l	10.4	INR	10.5				
SpO2 WITHOUT supplemental oxygen (before start, %)	10.7b	SpO2 WITH supplemental oxygen (before start, %)	10.6	O2 supplied (l/min	10.7				
BP before anaesthesia	10.8	BP after anaesthesia	10.9						
Operators, equipment									
Operator 1, four-character code	10.10	Nurse 1, four-character code	10.11						
Bronchoscope 1	10.12	Rack	10.13						
Drugs									
Lidocaine 10 mg/spray, number of applications	10.14								
Bronchodilation prior to procedure	*Drug(s)*	*Amount + unit*	*Indication*						
Fill in here if given:	10.15	10.16, 10.16a	10.17						
Alfentanil preop., mg	10.18	Supplemental alfentanil perop, mg	10.19						
Midazolam preop, mg	10.20	Supplemental midazolam perop, mg	10.21						
Applied lidocaine during procedure, in millilitre (20 mg/ml)	10.22	Applied adrenaline, 0,1 mg/ml, place, amount	10.23						
Procedure start and end									
Time, start (passing of **vocal cords**)	10.24	Time end (scope withdrawn)	10.25						

ID number, MicroCOPD: _____

Sampling							
Negative controls of fluid?	*11.1*	**Inspection** completed**	*11.2*	**Normal** (not norm. next page)	*11.3*		
Gingival samples taken.	*11.3a*	Comment gingiva	*11.3b*				
Oral wash, type & fluid amount	*11.4a, 11.4b,*	**Oral wash return (ml)**	*11.5*				
Brushes	**Order**	**Lobe**	**Segment**	**Number of brushes**			
Right	*11.6a*	*11.6*	*11.7*	*11.8*			
Left	*11.9a*	*11.9*	*11.10*	*11.11*			
Lavage	**Order**	**Lobe**	**Segm.**	**Type of fluid**	**BAL or SVL?**	**Installed (ml)**	**Return (ml)**
Right	*11.18a*	*11.18*	*11.19*	*11.20*	*11.21*	*11.22*	*11.23*
Left	*11.12a*	*11.12*	*11.13*	*11.14*	*11.15*	*11.16*	*11.17*

Endobrochial biopsy	**Place (lobe, carina level and segment)**	**Type of forceps**	**Sent to (GMA, freezer, mitoc., other)**
number 1	*11.24*	*11a*	*11a1*
number 2	*11.27*	*11b*	*11b1*
number 3	*11.30*	*11c*	*11c1*
number 4	*11.33*	*11d*	*11d1*
number 5	*11.36*	*11e*	*11e1*
number 6	*11.39*	*11f*	*11f1*
number 7	*11.41a*	*11g*	*11g1*
number 8	*11.41c*	*11h*	*11h1*

Termination		
Complicated	*11.42, 11.43*	*Must fill out complication form*
When can the patient eat (time)	*11.44*	
Observation time completed, time, four-character code	*11.45*	*11.46*

* anaesthesia, sedation

** vocal cords, carina, inspection of all lobes and segmental ostia

ID number, MicroCOPD: _____

Bronchoscopy event form

Study personnel present: _____ *12.1a - h*

Event:	**Time, duration, assumed cause and intervention**
No event (mark) *12.2*	
Cough *12.3, 12.3a*	
Dyspnoea *12.4, 12.4a*	
Oxygen desaturation >4 % or to <90 % *12.5, 12.5a*	
Change in BP/Heart rate *12.6, 12.6a*	
Bleeding *12.7*	
Other serious complication *12.8*	
Other events *12.9*	

ID number, MicroCOPD: _____

Subject experience of bronchoscopy

Question	Immediately after	After the observation	After 1 week
A. What do you think of the procedure now? Have you had any discomforts (which?) **(Open question)**	*13.1a*	*13.1b*	*13.1c*
B. How **uncomfortable** did you find this experience, taking into consideration everything that has happened until now, on a scale from 0 to 10, where 10 is the most uncomfortable you can imagine, and 0 is no discomfort.	*13.2a*	*13.2b*	*13.2c*
C. For how long do you think the procedure lasted?	*13.3a*		
1. How short of breath are you now? (Borg scale), 0-10 (show Borg scale)	*13.4a*	*13.4b*	*13.4c*
2. If you were asked to participate in a **new** research project involving the **same procedure**, would you participate? (if yes, go to question 4, if no, ask question 3)	*13.5a*	*13.5b*	*13.5c*
3. If your doctor advised you to undergo this type of procedure, would you then have done it **again**?	*13.6a*	*13.6b*	*13.6c*
4. Do you have **a sensation of fever** in your body now, or have you had fever/fever sensation in relation to bronchoscopy or after the procedure?	*13.7a*	*13.7b*	*13.7c*
5. Did you cough **blood** or red/light red saliva?	*13.8a*	*13.8b*	*13.8c*
6. Have you, after the procedure, experienced			
a. Increased breathlessness, dyspnoea or tightening of the chest? (*Synonymous words*)?	*13.9a*	*13.9b*	*13.9c*
b. Increased sputum?	*13.10a*	*13.10b*	*13.10c*
c. Sputum colour change?	*13.11a*	*13.11b*	*13.11c*
d. Increased rhinitis/stuffed nose?	*13.12a*	*13.12b*	*13.12c*
e. Increased wheezing chest sounds?	*13.13a*	*13.13b*	*13.13c*
f. Sore throat/increased cough?	*13.14a*	*13.14b*	*13.14c*
g. Increased fatigue/lack of initiative?	*13.15a*	*13.15b*	*13.15c*
f. flu symptoms (fever, muscle/joint ache, headache, reduced general condition)?	*13.16a*	*13.16b*	*13.16c*
7. Have you, after the procedure, needed to **seek a doctor/call** (unscheduled), use **antibiotics**, receive cortisone/**prednisolone** or be admitted to **hospital**? (if yes, note what, and cause)			*13.17, 13.17a*
8. Have you, in relation to the procedure or after received any new treatment? In case, which? (type of treatment, if drug – dosage etc)			*13.18*
9. Note if the participant has been in contact with a physician or the study personnel outside of standard follow-up – reason for contact, date and intervention.			*13.19*

16.2 Paper I

EUROPEAN
CLINICAL RESPIRATORY
JOURNAL

co-action PUBLISHING

REVIEW ARTICLE

Participation in research bronchoscopy: a literature review

Einar Marius Hjellestad Martinsen[1]*, Elise Orvedal Leiten[1], Per Sigvald Bakke[1], Tomas Mikal Lind Eagan[1,2] and Rune Grønseth[2]

[1]Department of Clinical Science, University of Bergen, Bergen, Norway; [2]Department of Thoracic Medicine, Haukeland University Hospital, Bergen, Norway

Bronchoscopy is the preferred method for collecting biological samples from the lower airways of subjects in clinical research. However, ensuring participation in clinical research can be challenging when the research includes an invasive procedure. For this report we reviewed the literature to look for information on participation in research bronchoscopy studies to better design our own study, the Bergen COPD Microbiome study (MicroCOPD). We performed a systematic literature search on participation in research bronchoscopy studies in February 2014 using the search engines of PubMed and EMBASE. The literature search resulted in seven relevant papers. Motivation was an end point in six of the seven papers, but reasons for declining participation and recruitment strategies also seemed important. Human subjects participate in research bronchoscopy studies for personal benefit and altruistic reasons. Inconvenience associated with research, in addition to fear of procedures, is considered a barrier. Radio, especially news stations, generated the most inquiries for a clinical study involving bronchoscopy. There is a lack of information on participation in research bronchoscopy studies in the literature. A bronchoscopy study has been initiated at Haukeland University Hospital, Bergen, Norway, to examine the role of the microbiome in COPD, and participation will be explored as a substudy.

Keywords: *COPD*; *clinical research*; *research subjects*; *human volunteers*; *motivation*; *refusal to participate*

Responsible Editor: Göran Eriksson, Lund University, Sweden.

*Correspondence to: Einar Marius Hjellestad Martinsen, Department of Clinical Science, University of Bergen, NO-5021 Bergen, Norway, Email: einar.martinsen@student.uib.no

Received: 24 August 2015; Revised: 21 December 2015; Accepted: 7 January 2016; Published: 3 February 2016

Chronic obstructive pulmonary disease (COPD) will be the third leading cause of death in 2030, according to estimates by the World Health Organization (1). The mechanisms explaining why only a limited fraction of individuals exposed to tobacco and other air pollutants develop COPD remain unknown. Recent advances in the field of metagenomics have indicated that airway microbiota might differ between subjects with and without COPD (2). To sample the airway microbiota, it is necessary to have a feasible method, yet with minimal contamination. Although induced sputum is a possibility (3), this method is prone to contamination from the oral microbiota. Furthermore, the accuracy in predicting which segments of the airways are being sampled is uncertain.

Bronchoscopy is a safe procedure with a low complication rate (4) and is the ideal procedure both to ensure minimal contamination as well as to enable mapping of different areas of the airways. However, a semi-invasive procedure such as bronchoscopy can be associated with discomfort. Together with pre-procedural anxiety, this discomfort might lower participation in studies that sample the airways by bronchoscopy. Previous studies on the airway microbiota in asthma and COPD patients with bronchoscopic sampling had low numbers of participants (2, 5, 6). There is a need for studies with more statistical strength to secure reliable and reproducible data. However, large-scale bronchoscopy studies would require attention to logistic challenges, including recruitment and participation.

The views expressed in this report are that of the authors and not the official position of the institution.

Citation: European Clinical Respiratory Journal 2016, **3**: 29511 - http://dx.doi.org/10.3402/ecrj.v3.29511

More information on response rates and participation motives could lead to better-targeted recruitment for clinical studies. Furthermore, by revealing common preprocedural concerns and anxieties, it might be possible to also improve patient information and compliance in regular clinical practice. The aim of the current report was to perform a systematic review of the current literature on participation motives, response rates, and recruitment strategies in research bronchoscopy studies with an emphasis on studies including COPD patients.

Methods

Search strategy

Two separate literature searches were performed using the PubMed search engine of the US National Library of Medicine (7) and the Excerpta Medica Database (EMBASE) provided by the medical publisher Elsevier (8).

PubMed papers are indexed by keywords called *medical subject headings* (MeSH) (9). Due to the hierarchical organization, generalized MeSH terms include papers classified by specific MeSH term. We identified MeSH terms from the indexed papers in initial searches, supplied by qualified suggestions from a collegial brainstorming session using a modification of a population–intervention–comparison–outcome (PICO) scheme (10). Most search terms were included as both MeSH terms and text words to increase search sensitivity. The columns were combined with 'OR', and rows were combined with 'AND'.

EMBASE has similar functions as PubMed, though MeSH terms are replaced by Emtree terms. We used the same modified PICO scheme (Table 1) for the EMBASE search. MeSH terms were replaced by explosion search, and text words were replaced by multipurpose (mp) terms.

Titles and abstracts were sifted and classified by prespecified exclusion and inclusion criteria (Table 2). Only papers concerning recruitment to studies including bronchoscopy were included. Reports of motives or perceived benefits of participation in studies with respiratory invasive procedures, reasons for non-response, recruitment sources, and response rates in studies involving respiratory invasive procedures were included. Papers not written in English or a Scandinavian language were excluded, together with non-human studies, case studies, and secondary publications, including literature reviews, reports, comments, letters, guidelines, newspaper articles, books, or book chapters. Studies that did not have participation as a main objective or as a study end point were also excluded. Similar criteria were used in evaluating retrieved papers found from both the PubMed and EMBASE searches.

Table 1. Modified PICO scheme used for a literature review on participation in research bronchoscopy studies

P1	P2	I	O
COPD (MeSH)	Patients (MeSH)	Bronchoscopy (MeSH)	Patient participation (MeSH)
COPD (tw)	Patients (tw)	Bronchoscopy (tw)	Participation (tw)
Chronic obstructive pulmonary disease (tw)	Participants (tw)		Response (tw)
	Human volunteers (MeSH)		Non-response (tw)
	Volunteers (tw)		Attitude (MeSH)
	Study (MeSH major topic)		Attitude (tw)
	Trial (MeSH major topic)		Motivation (MeSH)
	Research subjects (MeSH)		Motivation (tw)
	Research subjects/psychology (MeSH)		Refusal to participate (MeSH)
	Clinical research (MeSH)		Refusal to participate (tw)
			Informed consent (MeSH)
			Participation rate, patient (MeSH)
			Participation rate (tw)
			Patient selection (MeSH)
			Patient selection (tw)
			Advertising as topic (MeSH)
			Advertising (tw)
			Risk assessment (MeSH)
			Risk assessment (tw)
			Altruism (MeSH)
			Altruism (tw)
			Perception (tw)

The P1 column was excluded in the final search due to a paucity of results.
P, population; I, intervention; C, comparison; O, outcomes; MeSH, medical subject headings; tw, text words

Citation: European Clinical Respiratory Journal 2016, **3**: 29511 - http://dx.doi.org/10.3402/ecrj.v3.29511

Table 2. Number of retrieved papers for a literature search in the databases PubMed and EMBASE on participation in research bronchoscopy studies according to classification criteria

Classification criteria	PubMed	EMBASE
Total number retrieved	**989**	**987**
Non-English/non-Scandinavian language	102	None
Case studies/series	82	427
Secondary publications	116	108
Non-human studies	7	None
Participation not a major topic	674	443
Papers included in review	**8**	**9**
Papers common to PubMed and EMBASE searches	7	

Secondary publications included reviews, expert panels, letters, guidelines, and so on. The EMBASE search excluded studies in languages other than English and Scandinavian languages, as well as non-human studies

Results

Results from the two literature searches were classified as shown in Table 2. The majority of papers, 1,117, were excluded due to their lack of participation as a main objective or study end point. The PubMed search yielded eight relevant papers, and the EMBASE search yielded nine relevant papers. Seven out of nine articles from the EMBASE search were also found in the PubMed search. Thus, 10 individual papers were included for in-depth review. Three of these 10 papers did not report participation and were excluded.

Table 3 provides an in-depth overview of the final seven included papers (11–17). Four of the papers were published in the last five years, and six of the studies were conducted in Europe. Six papers focused on motives or perceived benefits of participation for studies involving research bronchoscopies (11, 12, 14–17). One of these also reported reasons to decline participation (16), and one studied predictors for the decision to consent to a second bronchoscopy (14). Further, one study evaluated the recruitment process in a lung cancer chemoprevention study that included a research bronchoscopy (13). Research bronchoscopies were carried out in all the studies (Table 3).

Five of the reviewed studies were prospective (11, 13–16). The largest included 146 participants in a smoking cessation trial (17), whereas the smallest examined 18 subjects (12). Three studies were limited to current or ex-smokers as study subjects (13, 16, 17), and none of these studies compared results with a healthy control population. Only one study included COPD patients (16), but the study emphasized lung cancer. Thus, none of the studies published results generalizable to a COPD population.

The most frequently used method of obtaining information was interviews (12, 13, 15–17), which were conducted by telephone in three of the studies (12, 13, 17). The remaining two studies made use of self-completed questionnaires for data collection (11, 14). Statistical methods were not reported in three of the included studies (11–13). The effects of demographic variables on participation were examined in five of the papers (13–17).

Motivation and benefits of study participation

We identified four main groups of motives for participation in bronchoscopy studies – *personal benefit* (11, 12, 14–17), *altruism* (11, 14, 16), *perceived importance of research* (11, 12), and *obedience to the authority of the researchers* (14).

Personal benefit was found as a participation motive in all six papers that examined participation motives (11, 12, 14–17). The benefits appeared to take various forms, but were mainly defined as interest in their own health (11, 14, 17), getting a proper health assessment (15, 16), or treatment or surveillance of their health (12, 15). In the study of parents of children with cystic fibrosis (CF), personal benefit was more important when they accepted participation on behalf of their children compared to parents of healthy control infants (98 vs 25%) (11). In the Malawian study, new volunteers expected participation to be of benefit to them, perceived as health assessment and prompt treatment (15). In the study on an HIV-infected population in the United Kingdom, two-thirds of participants stated personal benefit as important, but only 51% gave their own health as their main motive for participation (14).

In the latter study (14), *altruism* was considered the main reason for participation by HIV-infected patients. The same was true for parents of healthy controls in the CF case–control study (11). In the study by Patel et al., four of seven elderly participants (age >70) gave altruistic reasons for participation. However, this motive was often accompanied by self-interest. No participants stated altruism as the only motive for participation (16).

Participation as a result of *physician's authority* was rare, but could be seen in the study by Lipman et al. (14), which found that participants were motivated by being asked by a physician or that the physician seemed to want them to participate. Kerrison et al. (12) found a

Table 3. Overview of the included papers after a literature search on participation in research bronchoscopy studies

Author, year of publication, country (reference)	Study objective	Main outcomes	*N*	Population	Inclusion criteria	Design	Purpose of bronchoscopy	Bronchoscopy procedure	Relevant findings
Lipman, 1998, United Kingdom and United States (14)	Examine acceptability of research bronchoscopy in asymptomatic HIV subjects	Pre- and post-bronchoscopic participation motives, willingness to undergo second procedure. Response rates	Responders: 75 Non-responders: 23	HIV-infected individuals, 18–60 years	No acute/chronic respiratory disease, IV drug use or antiretroviral therapy	Prospective semistructured, self-completed questionnaires	Harvest cell populations from the lung	Bronchoscopy with BAL	Response rate: 70% Participation motives: personal benefits, altruism Non-responders younger 79% would accept second research bronchoscopy
Mtunthama, 2008, Malawi and United Kingdom (15)	Examine adequacy of study information, motives for participation, and complication rate in a research bronchoscopy study	Participation motives, perceived sufficiency of study information, self-reported adverse events	100	Malawian volunteers	None given	Prospective. Interviews with both open and closed questions	Determine factors underlying susceptibility to respiratory infection among adults	Bronchoscopy with BAL	Response rate: not given Participation motives: personal benefit, study enthusiasm All agreed to a second bronchoscopy. 94% considered pre-information adequate and useful
Kerrison, 2008, United Kingdom (12)	Examine patients' experience of clinical research	Participation motives, importance of referral and consent procedures, sufficiency of study information	18 in bronchoscopy study	Six clinical studies	None given	Retrospective qualitative study. Telephone interviews, questionnaires or focus groups	Identify and sample suspicious lesions	Fluorescence bronchoscopy and optical biopsy	Response rate: 64% in bronchoscopy substudy Participation benefits: extra care, increased surveillance, see an expert and help researchers and others Most encountered studies through the caregiving hospital
Kye, 2009, USA (13)	Examine various recruitment strategies	Response rates and cost calculations	137	Healthy ex-smokers, ≥30 pky	No contraindications for bronchoscopy or celecoxib, no use of certain medications or no cancer	Prospective study. Telephone interviews	Examination and sampling	White light and fluorescence bronchoscopy with BAL and bronchial biopsies	Response rate: 3.1% Radio advertisements generated most inquiries, followed by Internet, print media, posted flyers, and mass mailing
Schook, 2010, Netherlands (17)	Examine whether participation in a smoking cessation trial could influence smoking cessation	Smoking cessation rate and influence of participation on cessation	146	Healthy current or former smokers, ≥20 pky	No serious comorbid disease, FEV1 below 1000 mL or use of systemic or inhaled corticosteroids in previous year	Retrospective study. Questionnaire-based telephone interviews	Examination	Autofluorescence or white light bronchoscopy with biopsies	Response rate: 73% Participation motives: personal benefit

Table 3 (*Continued*)

Author, year of publication, country (reference)	Study objective	Main outcomes	*N*	Population	Inclusion criteria	Design	Purpose of bronchoscopy	Bronchoscopy procedure	Relevant findings
Patel, 2012, United Kingdom (16)	Examine methods and participation in a lung cancer screening study	Participation motives, declining reasons, and views on screening method	60	Current or ex-smokers, ≥20 pky and or ≥20 years, with mild to moderate COPD	No serious comorbid disease, life expectancy ≥5 years	Prospective qualitative study. Semi-structured interviews	Examination	Fluorescence bronchoscopy with biopsies	Response rate: not given Participation motives: personal benefit, altruism Participation barriers: fear, bad experiences, and travel
Chudleigh, 2013, United Kingdom (11)	Examine recruitment and retention of CF infants and healthy controls, as well as parental attitudes to participation	Participation motives, benefits, and disadvantages. Response rates	Cases: 85 Volunteers: 56	Infants with and without CF, including parents	CF: no contraindicated disorders, no preterms. Healthy controls: no medical and/or social contraindications, ≥2500 g	Prospective longitudinal, observational study. Self-completed questionnaires	Examination	Bronchoscopy with BAL	Response rate: 69% (CF), 21% (healthy controls) Participation motives: personal benefit, altruism

BAL, bronchoalveolar lavage; COPD, chronic obstructive pulmonary disease; CF, cystic fibrosis; FEV1, Forced expiratory volume; pky, pack-years.

somewhat similar motive when participants described their participation as an *important investment in scientific progress*. This motive was also mentioned by 32% of the parents in the CF case–control study (11).

Response rates

Response rate, defined as number of enrolled divided by approached or prescreened individuals, was a main objective in three of the reviewed papers (11, 13, 14), whereas response rates could be found or derived from two additional papers (12, 17). Response rates varied from 3 to 73% (11, 17) and seemed to be higher in studies involving individuals that were affected by the index disease (11, 14). Chudleigh et al. showed that recruitment of healthy controls was feasible, but more challenging than recruiting CF patients (11). Only Lipman et al. looked into predictors for participation and found that participants were significantly older (14).

Refusal

Reasons for declining primary participation or a second bronchoscopy was reported in three articles (13, 14, 16), and disadvantages of participation were examined in one study (11). All studies exploring refusal to participate listed a negative view on bronchoscopy as a main reason for non-response (13, 14, 16), and the severity and duration of previous experienced post-bronchoscopy symptoms was the most common reason to refuse a second research bronchoscopy in an HIV population (14). Patients that refused or were unsure (21% of participants who already had undergone a bronchoscopy) had more clinically advanced HIV infection. However, all of the participants did agree to a second bronchoscopy if medically indicated (14). Patel et al. identified barriers to participation as disadvantages of involvement exemplified by travel inconvenience, bad experiences, and negative perceptions of bronchoscopy (16), whereas Chudleigh reported anxiety and perceived risk for complications as negative aspects of participation (11).

Recruitment strategies

The study by Kye et al. was the only one reporting the efficacy of various recruitment strategies, and in a US lung cancer screening trial they found that radio advertisement was the most effective, especially information on the news stations, followed by Internet posting, print media, posted and racked flyers, and mass mailings (13).

Discussion

We have shown that the literature on participation in research bronchoscopy studies is somewhat limited. Nevertheless, investigators planning new studies might benefit from some inferences. First, it seems that both control subjects and younger individuals have lower response rates (11, 14). The highest response rates were seen in a study involving current or former smokers (17)

and in a study of an HIV population (14). Conversely, healthy subjects in a chemoprevention study had the lowest participation rate (3.1%) (13), possibly suggesting that more advanced diseases result in higher participation fractions. However, the latter study also pointed out the challenge of recruiting healthy subjects, as these require strict entry criteria and minimal comorbidities, which may also have contributed to their very low response rate. An earlier review conducted on participation in COPD studies without an invasive procedure also examined response rates (18). Sohanpal et al. found that study participation rates were higher than expected, and 81% of the studies included had a study participation rate above 50%. This finding conforms to the current report. The average participation rate was 77.8% in Sohanpal et al.'s review, whereas our review had an average of 55.8%, possibly suggesting higher participation rates in studies without invasive procedures. However, our material is limited and caution needs to be taken when comparing these results. Second, it seems that the main motivation for participation lies somewhere in a balance between perceived personal health benefit and altruism. This warrants some caution from investigators in not portraying participation as a substitution for otherwise inadequate healthcare access and some modesty in what results might be expected from the study. In particular, professionals should be aware that perceived authority results in undue pressure on invitees (14). Third, the main reason for non-response or declination of participation was fear related to the invasive procedure. It would be logical to assume that both content and deliverance of study information influences participation, but no study examined these factors in detail. Fourth, participation in research bronchoscopy does not seem to negatively influence patient consent for medically indicated procedures (14), which is of key importance when deciding whether or not to include bronchoscopy in a clinical study. Finally, researchers considering media as a recruitment source should know that radio and Internet advertising seem to be the most effective sources (13).

The distinction between non-therapeutic and therapeutic studies could possibly give rise to different participation motives. In the current review, we defined three of the articles as therapeutic (13, 16, 17), defined by any perceived direct benefit to the participants involved, and four as non-therapeutic (11, 12, 14, 15). Interestingly, all six papers that focused on motivation (11, 12, 14–17) reported personal benefit to be important. The perceived benefit from a non-therapeutic study could reflect a lack of understanding among the participants, thus emphasizing the importance of providing adequate and detailed information to eligible subjects.

In comparison to our area of study, participation in colorectal cancer screening trials was reviewed by Bakker et al. (19). They found that participation rates and completion of the fecal occult blood test as a screening procedure was higher when the researcher added the screening kit to the invitations. The addition was compared to invitations in which participants had to request the kit if interested or visit their general practitioner or a screening center to obtain the kit. In addition, a higher test completion rate was observed if participants received the test by post before a health check rather than being offered the test at the health check. General practitioner involvement and face-to-face invitation also resulted in higher participation. Furthermore, long travel distances from the screening facility were considered a participation barrier, consistent with results reported in the study by Patel et al. included in the current review (16).

A review by Ellis assessed patient and physician participation in randomized clinical trials in oncology (20). In line with our findings, the review confirmed that altruism, personal benefit, and scientific contribution were important motives for participants. Physician's authority was not emphasized, but patients that trusted their doctors seemed more willing to participate in clinical trials. Males, older patients, less well-educated persons or persons from lower socioeconomic backgrounds were also more inclined to participate. Among the reasons for non-response the authors emphasized fear of randomization, suggesting the process to be unfair, and loss of freedom to make their own decisions. Some individuals reported the feeling of being a guinea pig as unpleasant. Lack of information and distrust of the medical profession also appeared as barriers. Further, large difference in the treatment offered negatively influenced the decision to participate.

Our review has also revealed some limitations in the existing literature and some fields that warrant further research. The bulk of literature was on participation motives, whereas only two studies presented information on non-responders. Only one of these two presented a response rate, but this study had fewer than 100 individuals in their final analysis, and only 52% of the participants had actually undergone a bronchoscopy at the time of analysis (14).

We have summarized the existing literature on recruitment strategies, response rates, and participation motives in studies including bronchoscopy as a part of their design. In particular we set out to identify studies including COPD patients, but found there was very little data on this patient group. Inclusion for an observational bronchoscopy study was completed in June 2015 at Haukeland University Hospital, Bergen, Norway (MicroCOPD). The aim of the MicroCOPD study is to shed light on the role of the microbiome in COPD (21). Participants underwent a bronchoscopy with collection of protected specimen brushes, small-volume lavage, bronchoalveolar lavage, and bronchial biopsies. The respiratory microbiome in subjects with and without COPD will be investigated

Citation: European Clinical Respiratory Journal 2016, **3**: 29511 - http://dx.doi.org/10.3402/ecrj.v3.29511

relative to disease progression and development and are expected to provide insight in a new and promising research field. Participation will be examined as a substudy. Motives for participation will be asked as an open question before the bronchoscopy. Response rates will be estimated, and predictors for participation are targeted to be revealed. It is anticipated that these results will contribute to later research on COPD and facilitate the conduction of other bronchoscopy studies.

Conclusions

A literature search performed between December 2013 and February 2014 exploring participation in clinical studies involving research bronchoscopies yielded seven relevant articles. Conducting bronchoscopy studies involves difficulties in recruiting control subjects and younger individuals, as well as the invasive nature of the procedure. Responders seem driven by a combination of personal health benefit and altruistic motives. However, we found no solid evidence on recruitment for COPD studies, and the characterization of non-responders had major limitations. Thus, further research on participation in bronchoscopy studies is warranted.

Conflict of interest and funding

EMM, EOL, and PSB have nothing to disclose. TMLE has received personal fees from GlaxoSmithKline, Boehringer-Ingelheim, and InterMune, outside the submitted work. RG has received grants and personal fees from GlaxoSmithKline and personal fees from Boehringer-Ingelheim and Astra Zeneca, outside the submitted work.

EMM is a participant in the Medical Student Research Programme at the University of Bergen, and the current work is funded by his fellowship from the university.

Funding sources: EMM and EOL received scholarships from the Medical Student Research Programme at the University of Bergen, Norway.

References

1. World Health Organization. Projections of mortality and causes of death, 2015 and 2030. Available from: http://www.who.int/healthinfo/global_burden_disease/projections/en/ [cited 13 July 2015].
2. 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.
3. Keatings VM, Nightingale JA. Induced sputum: whole sample. Methods Mol Med. 2001; 56: 93–8.
4. Du Rand IA, Blaikley J, Booton R, Chaudhuri N, Gupta V, Khalid S, et al. British Thoracic Society guideline for diagnostic flexible bronchoscopy in adults: accredited by NICE. Thorax. 2013; 68(Suppl 1): i1–i44.
5. 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.
6. 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.
7. PubMed [Internet]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/
8. EMBASE [Internet]. Available from: http://www.elsevier.com/online-tools/embase.
9. U.S. National Library of Medicine. Fact Sheet Medical Subject Headings (MeSH®) 1999 [updated 9 December 2013]. Available from: http://www.nlm.nih.gov/pubs/factsheets/mesh.html [cited 5 January 2015].
10. Wikipedia. PICO process: Wikipedia; 2014 [updated 19 May 2014]. Available from: http://en.wikipedia.org/wiki/PICO_process [cited 11 November 2014].
11. Chudleigh J, Hoo AF, Ahmed D, Prasad A, Sheehan D, Francis J, et al. Positive parental attitudes to participating in research involving newborn screened infants with CF. J Cyst Fibros. 2013; 12(3): 234–40.
12. Kerrison S, Laws S, Cane M, Thompson A. The patient's experience of being a human subject. J R Soc Med. 2008; 101(8): 416–22.
13. Kye SH, Tashkin DP, Roth MD, Adams B, Nie WX, Mao JT. Recruitment strategies for a lung cancer chemoprevention trial involving ex-smokers. Contemporary Clin Trials. 2009; 30(5): 464–72.
14. Lipman MC, Stobbs D, Madge S, Miller R, Johnson MA. Research bronchoscopies do not adversely affect HIV-infected individuals' future health-care decisions. Chest. 1998; 114(1): 284–90.
15. Mtunthama N, Malamba R, French N, Molyneux ME, Zijlstra EE, Gordon SB. Malawians permit research bronchoscopy due to perceived need for healthcare. J Med Ethics. 2008; 34(4): 303–7.
16. Patel D, Akporobaro A, Chinyanganya N, Hackshaw A, Seale C, Spiro SG, et al. Attitudes to participation in a lung cancer screening trial: a qualitative study. Thorax. 2012; 67(5): 418–25.
17. Schook RM, Postmus BB, van den Berg RM, Sutedja TG, Man de FS, Smit EF, et al. The finding of premalignant lesions is not associated with smoking cessation in chemoprevention study volunteers. J Thorac Oncol. 2010; 5(8): 1240–5.
18. Sohanpal R, Hooper R, Hames R, Priebe S, Taylor S. Reporting participation rates in studies of non-pharmacological interventions for patients with chronic obstructive pulmonary disease: a systematic review. Syst Rev. 2012; 1: 66.
19. Khalid-de Bakker C, Jonkers D, Smits K, Mesters I, Masclee A, Stockbrugger R. Participation in colorectal cancer screening trials after first-time invitation: a systematic review. Endoscopy. 2011; 43(12): 1059–86.
20. Ellis PM. Attitudes towards and participation in randomised clinical trials in oncology: a review of the literature. Ann Oncol. 2000; 11(8): 939–45.
21. Grønseth R, Haaland I, Wiker HG, Martinsen EMH, Leiten EO, Husebø G, et al. The Bergen COPD microbiome study (MicroCOPD): rationale, design, and initial experiences. Eur Clin Respir J. 2014; 1: 26196, doi: http://dx.doi.org/10.3402/ecrj.v1.26196

16.3 Paper II

Martinsen et al. Multidisciplinary Respiratory Medicine (2019) 14:14
https://doi.org/10.1186/s40248-019-0178-3

Multidisciplinary Respiratory Medicine

ORIGINAL RESEARCH ARTICLE Open Access

Motivation and response rates in bronchoscopy studies

Einar M. H. Martinsen[1*], Tomas M. L. Eagan[1,2], Elise O. Leiten[1], Eli Nordeide[2], Per S. Bakke[1], Sverre Lehmann[1,2] and Rune Nielsen[1,2]

Abstract

Background: Bronchoscopy is frequently used to sample the lower airways in lung microbiome studies. Despite being a safe procedure, it is associated with discomfort that may result in reservations regarding participation in research bronchoscopy studies. Information on participation in research bronchoscopy studies is limited. We report response rates, reasons for non-response, motivation for participation, and predictors of participation in a large-scale single-centre bronchoscopy study ("MicroCOPD").

Methods: Two hundred forty-nine participants underwent at least one bronchoscopy in addition to being examined by a physician, having lung function tested, and being offered a CT scan of the heart and lungs (subjects > 40 years). Each participant was asked an open question regarding motivation. Non-response reasons were gathered, and response rates were calculated.

Results: The study had a response rate just above 50%, and men had a significantly higher response rate than women (56.5% vs. 44.8%, $p = 0.01$). Procedural fear was the most common non-response reason. Most participants participated due to perceived personal benefit, but a large proportion did also participate to help others and contribute to science. Men were less likely to give exclusive altruistic motives, whereas subjects with asthma were more likely to report exclusive personal benefit as main motive.

Conclusion: Response rates of about 50% in bronchoscopy studies make large bronchoscopy studies feasible, but the fact that participants are motivated by their own health status places a large responsibility on the investigators regarding the accuracy of the provided study information.

Keywords: COPD, Clinical research, Motivation, Response rate, Non-response

Introduction

Although primarily employed as a diagnostic tool, bronchoscopy is useful in studies on airway inflammation, bronchial remodelling, and the airway microbiota. Studies on airway microbiota have so far been relatively low-powered [1–3], and future studies will depend on larger samples. Even if bronchoscopy is associated with a low complication rate [4], some discomfort is inevitable, and potential participants may therefore have reservations [5]. Knowledge on motives for participation and response rates in bronchoscopy studies have the potential to optimise the recruitment process.

However, there are few studies providing reliable response rates and motives for participation in bronchoscopy studies. A literature review on research bronchoscopy studies included seven relevant studies, and found personal benefit and altruistic reasons to be the most important participation motives, whereas fear of the bronchoscopy was reported as a participation barrier [6]. Response rates from the seven studies varied from 3 to 73%, and no study examined participation among subjects with chronic obstructive pulmonary disease (COPD) in particular [6].

The Bergen COPD Microbiome study ("MicroCOPD") is a large single-centre study of the airway microbiota, with bronchoscopic sampling of all participants. Data was collected at the Department of Thoracic Medicine, Haukeland University Hospital in Bergen, Norway, between April 2013 and June 2015. The main objective in

* Correspondence: einar.martinsen@uib.no
[1]Department of Clinical Science, University of Bergen, N-5021 Bergen, Norway
Full list of author information is available at the end of the article

the MicroCOPD study was to examine and compare airway microbiota from subjects without COPD (controls) and subjects with COPD. Some subjects with asthma were also included. The current paper reports response rates, reasons for non-response, motivation for participation, and predictors of participation in the MicroCOPD study.

Methods

Study design and population

The MicroCOPD study was a single-centre prospective, observational study carried out in Bergen, Western Norway. The study design has been described previously [7]. A pilot study of eight subjects with COPD was conducted in 2012 for protocol improvement. Participants from the pilot and main study were included in the current analyses. The main study included its first participant on April 11th, 2013, with the final study bronchoscopy performed June 5th, 2015. The study was conducted in accordance with the declaration of Helsinki and guidelines for good clinical practice. The regional committee of medical ethics approved the project (project number 2011/1307), and all participants provided informed written consent.

Controls and subjects with COPD or asthma were mainly recruited among participants of two previous studies performed by our research groups; the GeneCOPD study from 2003 to 2004 [8] and the Bergen COPD cohort study from 2006 to 2009 [9–12]. In addition, 6 subjects were recruited from outpatient clinics, and 8 subjects were recruited by their own initiative through attention from local media and hospital staff.

All subjects from the two previous studies who still lived in Bergen or the closest surrounding municipalities were eligible for participation. Potential participants were screened by an interview performed by a study physician regarding exclusion criteria for bronchoscopy before giving informed consent. We did not include subjects with increased bleeding risk, subjects with unstable cardiac conditions, or subjects with hypercapnia or hypoxaemia when receiving oxygen supplement [7]. Elderly subjects judged frail by the study physician were excluded. Participation was postponed for subjects that had used antibiotics or oral corticosteroids in the last 14 days, as well as subjects with symptoms of acute exacerbation of COPD.

Data collection

Subjects that declined participation at the screening interview were asked about their non-response reason. Participants attended the outpatient clinic over one or two days depending on the availability of computed tomography (CT) scanning. A pulmonary and coronary CT scan was offered as part of a concurrent study, and this would be performed prior to bronchoscopy if the participants were scheduled for both procedures. At the day of bronchoscopy, prior to the procedure, participants underwent a structured interview regarding their medical history, respiratory symptoms, smoking habits, medication use, motivation, and exacerbation frequency if they had obstructive lung disease. An open question on motivation was first included in the study questionnaire from the fifth pilot patient, asked immediately prior to the procedure. Additionally, post-bronchodilator spirometry was performed and blood samples were collected.

Diagnoses of COPD and asthma were evaluated by the study physician, based on medical history, symptoms, pulmonary CT scan, and post-bronchodilator spirometry [13, 14]. Controls were judged to have no sign of airway or lung disease, based on the same information. After all participants were included, a panel of three physicians evaluated the diagnoses for a quality control regarding possible misclassification between controls and subjects with COPD or asthma.

The bronchoscopy procedure was explained in detail to each participant by the study physician immediately prior to the procedure. The procedure was performed with the participant in the supine position, with the option of light sedation (alfentanil, potentially combined with midazolam). Samples were gathered by sterile brushes and bronchoalveolar lavage (BAL) after application of a local anaesthetic agent. Additionally, gathering of bronchial biopsies began in May 2014. The details of bronchoscopic sampling have been previously published [7]. The average length of the bronchoscopy procedures was 15 min, including bronchoscopies with bronchial biopsies.

Outcomes

Responders were subjects who accepted the invitation and underwent a bronchoscopy. Non-responders were subjects who did not undergo a bronchoscopy. Late non-responders were subjects who reconsidered an initial decision to participate, or when bronchoscopies were terminated before sampling due to participant discomfort. All non-responders were asked about their reason to decline participation. The response rate was defined as the number of bronchoscopies performed divided by the number of invited subjects.

Participation motives were collected from an open question before bronchoscopy started: "Why did you wish to take part in this project?". Participants could provide more than one motive for participation, thus the overall numbers of motives exceeded the numbers of participants. At the time of analysis we initially merged the unique motives into 16 more principal motives, and then further classified these into three main groups: 1) *Altruism* was motivation by a wish to help others or a

wish to continue participation from previous studies, as well as desire to contribute to science. 2) *Personal benefit* was motivation by a wish to somehow improve own health by participating in the project. 3) *Obligation* was a subjective feeling of being bound to participate. Subjects without specific reasons were labelled missing. We constructed binary variables "exclusive altruism" and "exclusive personal benefit" by coding them as '1' if the participant only gave altruism or personal benefit as main motive, respectively. Participants stating both altruism and personal benefit, or altruism/personal benefit and obligation, were coded '0' on these variables.

Statistical analyses

All analyses were performed using Stata version 14 [15]. Response rates were stratified by sex and study category (control/obstructive lung disease). Chi-square test or Fisher's exact test was used to compare frequencies of non-response reasons.

Bivariate analyses of responders and non-responders, as well as initial and late non-responders, were performed using parametric (t-test) and non-parametric tests (chi-square test or Fisher's exact test), when judged appropriate. Bivariate logistic regression models were fitted with "exclusive altruism" or "exclusive personal benefit" as outcome. Covariates with p less than 0.20 before adjustment were included in multivariate models. In the logistic regression, age and FEV_1 were treated as continuous variables, but divided by 10 to provide ratios for an increase of 10 units. Smoking habits were grouped according to current smoking status (never-, ex-, current-smokers), and we calculated number of pack/years (cigarettes per day divided by 20, multiplied by years smoking). Never-smokers and ex-smokers were merged into one category in the logistic regression analysis. Lung function was analysed using the percentage of predicted values of FEV_1 and FVC, as well as the FEV_1/FVC-ratio. Dyspnoea was classified according to the modified Medical Research Council (mMRC) dyspnoea scale [16].

Results

Flow chart (Fig. 1)

In total, 2,205 subjects from the two previous COPD cohorts were considered potential participants for the MicroCOPD study. 1,743 were ineligible, mainly due to death or that the MicroCOPD inclusion period ended (see Additional file 1: Table S1 for details). The total number of invited individuals for bronchoscopy was 462, of whom 323 subjects accepted the invitation. 85 subjects reconsidered their decision to participate, and further three bronchoscopies were terminated before sampling due to participant discomfort.

Response rates

Since the denominator of the response rate for subjects recruited from outpatient clinics, local media, and hospital staff was unknown, these 14 subjects were excluded from the response-rate analyses. Final response rate for the main study was 50.9%. The response rates in women and men were 44.8% (100/223) and 56.5% (135/239), respectively ($p = 0.01$). No significant difference in attendance was seen between subjects without obstructive lung disease and subjects with COPD or asthma.

Demographics of responders and non-responders (Table 1)

There was no significant difference in age between responders and non-responders or between early or late responders. Whereas responders and initial non-responders did not differ by study category, there was a larger number of patients among the late non-responders compared with initial non-responders.

Non-response reasons (Table 2)

Most initial non-responders stated that they feared the discomfort of a bronchoscopy (23.7%), and together with unspecific fear and worries related to study participation this accounted for 40.2% of all the initial non-response. The percentage of worries and fears in late non-responders was more than twice as high ($p < 0.01$). Among the initial non-responders there was a higher expression of study fatigue (10.1% vs. 2.3%, $p = 0.03$). A considerable number of the non-responders felt that their own health prevented participation (17.3% in initial non-responders, and 26.1% in late non-responders), and most so in subjects with obstructive lung diseases ($p < 0.01$).

Detailed demographics of responders (Table 3)

The majority of responders were ex-smokers (68.3%), and a minority were never-smokers (9.6%). Age and sex were not significantly different between the controls and subjects with COPD. However, there were more ex-smokers and higher number of pack/years, less education, fewer married, more drug use, more comorbidities, as well as higher symptom burden and lower lung function among the subjects with COPD ($p \leq 0.01$, tests not shown).

Motivation (Table 4)

Personal health benefit was the most common stated *principal motive* for participation (49.0%), followed by contribution to science (39.2%). 39 subjects (15.9%) also mentioned helping others as motivation. After merging into broader categories, primarily altruism was the *main motive* stated by most participants (67.3%), while 52.2% gave motives considered to be of personal benefit. Only 2.0% participated out of a sense of obligation.

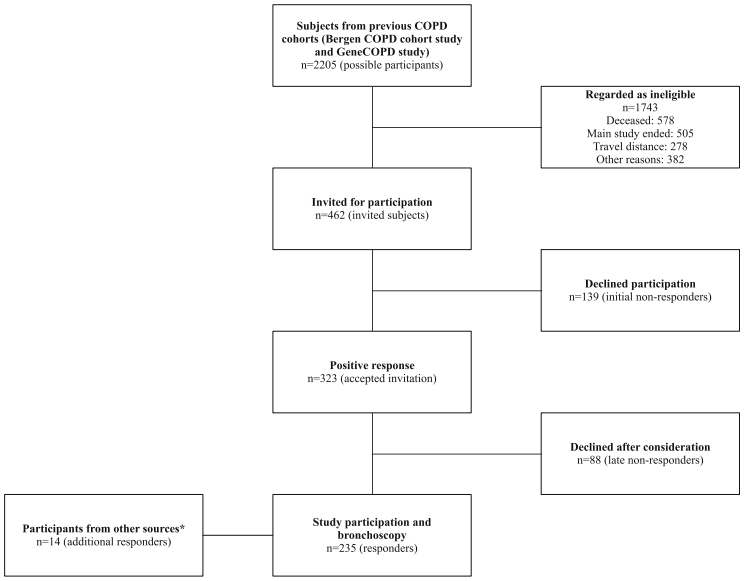


Fig. 1 Flow chart of an observational research bronchoscopy study. * Local media, hospital staff, and outpatient clinics were regarded as other sources

Frequencies on motives were also stratified by participant group, i.e. control, COPD, and asthma (see Additional file 1: Table S2 for details).

Men were less likely to state altruism as their main motive for participation (Fig. 2a, odds ratio (OR) 0.6, 95% confidence interval (0.3, 0.9)). This effect was more pronounced in the adjusted model, OR = 0.5 (0.3, 0.9). More subjects with asthma stated personal benefit motives than controls, unadjusted OR = 4.4 (1.5, 13.3), adjusted OR = 5.1 (1.6, 16.0) (Fig. 2b). No significant effect was observed by FEV_1 in percentage of predicted, age, number of comorbidities, education, or smoking status.

Discussion

We have reported response rates, non-response reasons, and motives for participation in a large single-centre bronchoscopy study. Response rates were about 50%, and did not differ between controls and subjects with COPD or asthma. The main reasons for non-response were fear of discomfort from the bronchoscopic procedure, and a subjective feeling of being diseased or too bothered from health issues to participate, especially among subjects with COPD or asthma. Participants were most frequently motivated by altruistic motives, but less so for men.

Table 1 Demographics of responders and non-responders in an observational research bronchoscopy study

Variable	Responders $n = 235$	Non-responders		p-*	p**
		Initial, $n = 139$	Late, $n = 88$		
Age (SD)	66.9 (7.6)	67.9 (8.0)	67.4 (7.5)	0.3	0.6
Sex				0.01	0.5
Women (%)	42.6	56.1	51.1		
Men (%)	57.4	43.9	48.9		
Study category				0.2	0.06
Controls (%)	43.0	41.7	29.6		
Obstructive lung disease (%)	57.0	58.3	70.4		

*Difference between responders and all non-responders
**Difference between initial and late non-responders

Table 2 Self-reported non response-reasons in an observational research bronchoscopy study

Reasons	Non-responders			Study category		
	Initial, $n = 139$ Frequency (%)	Late, $n = 88$ Frequency (%)	p	Controls, $n = 84$ Frequency (%)	OLD, $n = 143$ Frequency (%)	p
Discomfort	33 (23.7)	16 (18.2)	0.3	19 (22.6)	30 (21.0)	0.2
Unspecified worries/fear concerning participation	23 (16.5)	30 (34.1)	< 0.01	19 (22.6)	34 (23.8)	0.08
Disease/health issues	24 (17.3)	23 (26.1)	0.1	8 (9.5)	39 (27.3)	< 0.01
Study fatigue	14 (10.1)	2 (2.3)	0.03	5 (6.0)	11 (7.7)	0.2
Time constraint	5 (3.6)	5 (5.7)	0.5	7 (8.3)	3 (2.1)	0.2
Practical[a]	5 (3.6)	6 (6.8)	0.3	5 (6.0)	6 (4.2)	0.9
Not satisfied with previous study participation	2 (1.4)	0	0.5	0	2 (1.4)	0.5
Feeling too old	2 (1.4)	0	0.5	2 (2.4)	0	0.2
Refuse to specify	1 (0.7)	0	1.0	1 (1.2)	0	0.5
Personal reason	0	3 (3.4)	0.06	1 (1.2)	2 (1.4)	1.0
Not specified	30 (21.6)	3 (3.4)	< 0.01	17 (20.2)	16 (11.2)	0.7

OLD obstructive lung disease
[a]Practical reflects practical issues for researcher or patient

Table 3 Demographics of participants in an observational research bronchoscopy study

Variable	All, $n = 249$	Control, $n = 103$	COPD, $n = 130$	Asthma, $n = 16$
Age (SD)	66.3 (8.3)	65.3 (8.6)	67.2 (7.3)	65.5 (12.6)
Sex (men)	143 (57.4)	60 (58.3)	76 (58.5)	7 (43.8)
Number of medications (SD)	3.8 (3.2)	1.8 (1.7)	5.4 (3.3)	3.6 (2.4)
Number of comorbidities (SD)	1.1 (1.1)	0.8 (1.0)	1.4 (1.2)	0.8 (0.9)
FEV_1, % of predicted (SD)	78.3 (28.2)	103.9 (12.3)	56.5 (19.2)	90.7 (13.3)
FVC, % of predicted (SD)	102.7 (18.7)	111.7 (13.5)	95.0 (19.2)	107.5 (16.5)
FEV_1/FVC-ratio (SD)	0.6 (0.2)	0.7 (0.1)	0.5 (0.1)	0.7 (0.1)
Pack/years (SD)[a]	28.6 (20.1)	21.6 (16.9)	33.7 (20.5)	20.9 (22.0)
Smoking status (%)				
Daily	55 (22.1)	25 (24.3)	30 (23.1)	0 (0.0)
Ex-smokers	170 (68.3)	59 (57.3)	99 (76.2)	12 (75.0)
Never	24 (9.6)	19 (18.5)	1 (0.8)	4 (25.0)
Marital status (%)[a]				
Married/partner	157 (64.3)	79 (77.5)	72 (57.1)	6 (37.5)
Widowed	22 (9.0)	3 (2.9)	16 (12.7)	3 (18.8)
Cohabitant	20 (8.2)	5 (4.9)	13 (10.3)	2 (12.5)
Divorced, lives alone	33 (13.5)	11 (10.8)	17 (13.5)	5 (31.3)
Single	12 (4.9)	4 (3.9)	8 (6.4)	0 (0.0)
Education (%)[a]				
Primary school	48 (19.8)	12 (11.8)	35 (27.8)	1 (6.7)
Upper secondary/high school	125 (51.4)	52 (51.0)	67 (53.2)	6 (40.0)
3 years or more of higher education	70 (28.8)	38 (37.3)	24 (19.1)	8 (53.3)
mMRC Grade 2 and higher (%)				
Grade 2 level ground	30 (54.6)	3 (100)	27 (52.9)	0 (0.0)
Grade 3100 m	18 (32.7)	0 (0.0)	17 (33.3)	1 (100)
Grade 4 resting dyspnoea	7 (12.7)	0 (0.0)	7 (13.7)	0 (0.0)

FEV_1 Forced Expiratory Volume in 1 Second, *FVC* Forced Vital Capacity, and *mMRC* modified Medical Research Council dyspnoea scale
[a]Due to some missing information, the sum of participants is not 249

Table 4 Motives reported by the 245[a] participants who gave motives in an observational research bronchoscopy study

Motives	n	Percentage
Primarily altruism[b]	*165*	*67.3*
Previous participation	23	9.4
Contribute to science	96	39.2
Help others	39	15.9
Give back (for previous participation)	7	2.9
Generally positive (to examination or participation and "yes-human")	6	2.4
Social responsibility	3	1.2
COPD in family/among friends (including risk of COPD in family)	19	7.8
Available time	3	1.2
Primarily personal benefit[b]	*128*	*52.2*
Personal health benefit	120	49.0
Experience the discomfort of bronchoscopy	1	0.4
Challenge	1	0.4
Curiosity	14	5.7
Fun	1	0.4
Primarily obligation[b]	*5*	*2.0*
Acquaintance (in study, working with and was connected to the study or asked by)	4	1.6
Trust in authority/research	1	0.4
Missing	20	8.2

[a]Participation was not part of the questionnaire for the first four participants
[b]Unique motives are categorised into three main motives (in italic) by merging the unique motives listed below the main motive. The frequency (n) of main motives is not equal to the sum of each principal motive because a subject stating both "personal health benefit" and "challenge" would result in two observations in principal motives, but just one after merging

Given the invasiveness of the involved procedures, a response rate of 50% is not remarkably low. Little data exists on participation in research bronchoscopy studies [6]. Neither of seven Norwegian respiratory health surveys studies between 1965 and 1999 included a bronchoscopy, but baseline response rates varied from 68 to 90% [17]. Once attending, only 5% of attendants did not complete their participation [17]. In the current study, 27.2% (88/323) reconsidered their decision, suggesting higher rates of reconsideration with more invasive procedures.

A trend towards lower participation rates has been observed in Norwegian studies over time [17, 18]. This trend could at worst lead to selection bias and compromised external validity. In general, young, single men living in urban areas are the least likely to participate in social science surveys, while older women are the most willing [18]. In the current study, more men than women were responders. Experiences from clinical work suggest that women worry more about clinical procedures, and this could serve as a possible explanation for the observed difference. Additionally, more men were motivated by perceived personal benefit. If this observation stands true, one could speculate that male motivation is more easily satisfied by participation in a clinical study involving an actual diagnostic procedure, than participation in a questionnaire study. We observed no difference in mean age between responders and non-responders, but younger individuals were omitted from the current study, and frail elders were excluded. In another Norwegian study on respiratory health, non-response was related to lower age, rural habitation, and smoking habits [19]. Response rates from the current study will help researchers scale the number of invited subjects, aiming to recruit a sufficient number of participants, in order to avoid type II errors. These numbers can also be of value when investigators seek funding and ethics approval, providing precise information regarding the inclusion process.

Knowledge of reasons for non-response could guide researchers to provide precise information regarding the procedure during recruitment, which in turn might influence the willingness to participate. Observed difference in worries/fear between initial and late non-responders suggests that participants become frightened during the waiting time. Information on relevant discomfort should always be disclosed at first contact to avoid unreasonably procedural fear, and unnecessary waiting time before scheduled procedures should be avoided, both for research and clinical purposes. This will reduce costs and planning of non-performed procedures.

In agreement with the literature, we could categorise motives for participation into three groups, namely *personal benefit, altruism,* and *obligation* [6], although the review stated *obedience to the authority of the researchers* as an own group. Only one subject claimed trust in authority/research to be of importance in the current study. This discrepancy with previous studies might reflect both cultural differences and differences in health care organisation. Furthermore, patients are increasingly making their own health decisions [20], which might have changed the view of physicians as authorities.

We observed that women expressed more altruistic motives than men. In concordance with the current study, a meta-analysis on altruism and gender by Rand et al. showed women to be more intuitively altruistic, and men to be more selfish both intuitively and after consideration [21]. Our observation that subjects with asthma tend to report personal benefit needs to be interpreted with some caution. There were few subjects with asthma in our study, and they were recruited in a non-controlled manner.

Our results indicate that providing information on future implications of research can promote participation

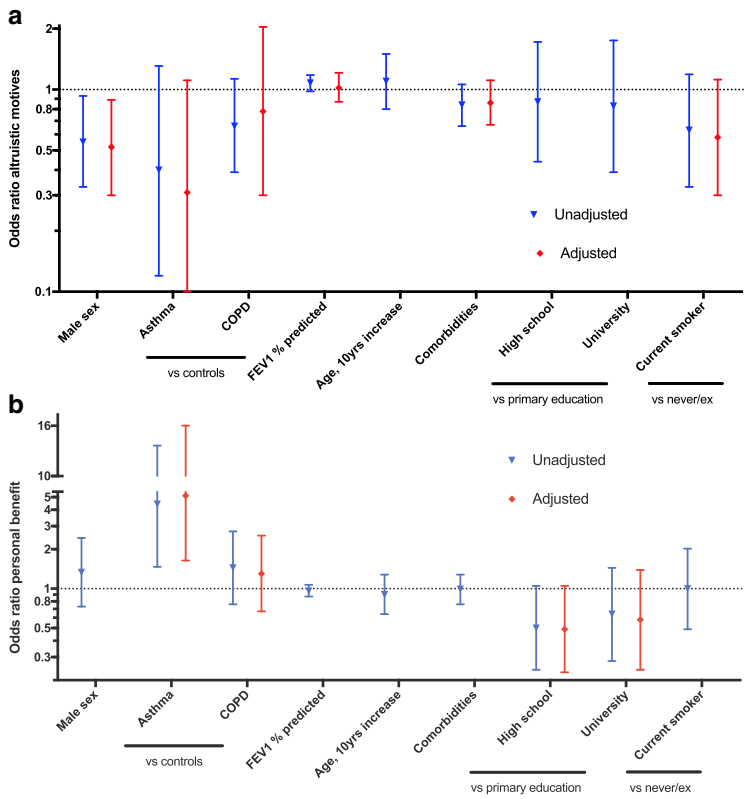


Fig. 2 Logistic regression on **a**) exclusive altruism- and **b**) exclusive personal benefit-variables in an observational research bronchoscopy study

by appealing to a desire to contribute to science and future health care. Emphasising potential health benefits of study participation would probably have an even greater effect, but this warrants caution. Screening effects of bronchoscopy are not known. Also, there is a small complication risk associated with the procedure [4]. Participants were offered participation also in a concurrent study wherein a CT scan was offered, however participation in either study was not dependent on participation in the other. Thus, no exclusive, immediate benefit was received for the participants in the MicroCOPD study. Even though this was clearly stated in the written consent, almost half of the final participants stated personal benefit as an important motive. Participants' expectation of perceived health benefit from participation is well known from the literature, even though no such benefit should be expected [5, 22], also where this is clearly stated by the research team [23]. Thus, we believe that participants´ perceived personal benefit in observational studies should be examined more thoroughly in future studies.

The MicroCOPD study is, to our knowledge, the largest single-centre lung microbiome study performed to date. We had extensive demographics on responders, and reliable results on motivation and non-response reasons. Some potential weaknesses deserve mentioning. Firstly, due to ethical and practical reasons, demographics on non-responders were sparse, and a considerable proportion of non-responders did not give any reason for their decline. Secondly, albeit a large study, the heterogeneity of the participants may have obscured the finding of important predictors of participation and motivation. Thirdly, an in-depth interview could have provided more insight into the details of non-response and motivation. Finally, most participants had already shown a willingness to take part in previous studies. Hence, they might be more prone to take part than a general population, generating some degree of selection bias. On the other hand, "study fatigue" might have lowered the participation rate in the current study.

Conclusions

The response rate for research bronchoscopy in our study was 50%, and did not differ between controls and subjects with COPD or asthma. Non-responders refused participation mainly due to procedural fear. In contrast, responders were driven by perceived personal benefit, but a large proportion did also participate to help others and contribute to science. Our findings underline the importance of providing comprehensive information about the procedures. This might serve to avoid refusal on a possible misunderstood risk assessment, and to secure inclusion of a sufficient number of well-informed participants.

Additional file

Additional file 1: Table S1. Reasons for ineligibility in an observational research bronchoscopy study, $n = 1743$. **Table S2.** Motives reported by the 245[a] participants who gave motives in an observational research bronchoscopy study stratified by participant group. (PDF 96 kb)

Abbreviations

"MicroCOPD": The Bergen COPD Microbiome study; BAL: Bronchoalveolar lavage; COPD: Chronic obstructive pulmonary disease; CT: Computed tomography; OLD: Obstructive lung disease

Acknowledgements

Not applicable

Funding

This work was supported by the Medical Student Research Programme at the University of Bergen, Norway.

Availability of data and materials

The dataset generated and/or analysed during the current study is not publicly available due to confidentiality, but is available from the corresponding author on reasonable request.

Additional information

An early version of this work was presented in poster form last September 10th 2017 at the European Respiratory Society International Congress in Milan.

Authors' contributions

EMHM had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. EMHM, TMLE, EOL, EN, SL, and RN participated in different aspects of data collection. EMHM, TMLE, EOL, EN, PSB, SL, and RN contributed substantially to the study design, data analysis and interpretation, and the writing of the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

The study was conducted in accordance with the declaration of Helsinki and guidelines for good clinical practice. The regional committee of medical ethics approved the project (project number 2011/1307), and all participants provided informed, written consent.

Consent for publication

Not applicable.

Competing interests

EMHM, EOL, EN, PSB and SL declare no conflict of interest. TMLE reports grants from Western Norway Regional Health Authority during the conduct of the study, and personal fees from Boehringer Ingelheim and AstraZeneca outside the submitted work. RN reports grants from GlaxoSmithKline during the conduct of the study, and grants from Boehringer Ingelheim, grants and personal fees from AstraZeneca, grants from Novartis, and personal fees from GlaxoSmithKline outside the submitted work.

Publisher's Note

Author details

[1]Department of Clinical Science, University of Bergen, N-5021 Bergen, Norway. [2]Department of Thoracic Medicine, Haukeland University Hospital, N-5021 Bergen, Norway.

Received: 5 October 2018 Accepted: 5 March 2019
Published online: 02 May 2019

References

1. Beck JM, Schloss PD, Venkataraman A, Twigg H 3rd, Jablonski KA, Bushman FD, et al. Multicenter comparison of lung and Oral microbiomes of HIV-infected and HIV-uninfected individuals. Am J Respir Crit Care Med. 2015; 192(11):1335–44.
2. Morris A, Beck JM, Schloss PD, Campbell TB, Crothers K, Curtis JL, et al. Comparison of the respiratory microbiome in healthy nonsmokers and smokers. Am J Respir Crit Care Med. 2013;187(10):1067–75.
3. Shenoy MK, Iwai S, Lin DL, Worodria W, Ayakaka I, Byanyima P, et al. Immune response and mortality risk relate to distinct lung microbiomes in patients with HIV and pneumonia. Am J Respir Crit Care Med. 2017; 195(1):104–14.
4. Leiten EO, Martinsen EMH, Bakke PS, Eagan TML, Grønseth R. Complications and discomfort of bronchoscopy: a systematic review. European clinical respiratory journal. 2016;3(1):33324.
5. Patel D, Akporobaro A, Chinyanganya N, Hackshaw A, Seale C, Spiro SG, et al. Attitudes to participation in a lung cancer screening trial: a qualitative study. Thorax. 2012;67(5):418–25.
6. Martinsen EM, Leiten EO, Bakke PS, Eagan TM, Gronseth R. Participation in research bronchoscopy: a literature review. European clinical respiratory journal. 2016;3:29511.
7. Grønseth R, Haaland I, Wiker HG, Martinsen EMH, Leiten EO, Husebø G, et al. The Bergen COPD microbiome study (MicroCOPD): rationale, design, and initial experiences. European clinical respiratory journal. 2014;1(1):26196.
8. Sorheim IC, Johannessen A, Grydeland TB, Omenaas ER, Gulsvik A, Bakke PS. Case-control studies on risk factors for chronic obstructive pulmonary disease: how does the sampling of the cases and controls affect the results? Clin Respir J. 2010;4(2):89–96.
9. Eagan TM, Aukrust P, Bakke PS, Damas JK, Skorge TD, Hardie JA, et al. Systemic mannose-binding lectin is not associated with chronic obstructive pulmonary disease. Respir Med. 2010;104(2):283–90.
10. 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.
11. Eagan TM, Damas JK, Ueland T, Voll-Aanerud M, Mollnes TE, Hardie JA, et al. Neutrophil gelatinase-associated lipocalin: a biomarker in COPD. Chest. 2010;138(4):888–95.
12. 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.
13. Global Initiative for Chronic Obstructive Lung Disease. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease, 2019 [cited December 2018]. Available from: https://goldcopd.org/.
14. Global Initiative for Asthma. Global Strategy for Asthma Management and Prevention, 2018 [cited December 2018]. Available from: https://ginasthma.org/.
15. StataCorp. Stata [cited April 2018]. Available from: https://www.stata.com.
16. MDCalc. mMRC (Modified Medical Research Council) Dyspnea Scale [cited September 2018]. Available from: https://www.mdcalc.com/mmrc-modified-medical-research-council-dyspnea-scale.
17. Gulsvik A, Humerfelt S, Bakke PS, Omenaas ER, Lehmann S. Norwegian population surveys on respiratory health in adults: objectives, design, methods, quality controls and response rates. Clin Respir J. 2008;2(Suppl 1):10–25.

18. ScienceNordic. Fewer willing to participate in surveys [cited October 2017]. Available from: http://sciencenordic.com/fewer-willing-participate-surveys.
19. Abrahamsen R, Svendsen MV, Henneberger PK, Gundersen GF, Toren K, Kongerud J, et al. Non-response in a cross-sectional study of respiratory health in Norway. BMJ Open. 2016;6(1):e009912.
20. Charles C, Gafni A, Whelan T. Shared decision-making in the medical encounter: what does it mean? (or it takes at least two to tango). Soc Sci Med. 1997;44(5):681–92.
21. Rand DG, Brescoll VL, Everett JA, Capraro V, Barcelo H. Social heuristics and social roles: intuition favors altruism for women but not for men. J. Exp. Psychol. Gen. 2016;145(4):389–96.
22. Mtunthama N, Malamba R, French N, Molyneux ME, Zijlstra EE, Gordon SB. Malawians permit research bronchoscopy due to perceived need for healthcare. J Med Ethics. 2008;34(4):303–7.
23. Lipman MC, Stobbs D, Madge S, Miller R, Johnson MA. Research bronchoscopies do not adversely affect HIV-infected individuals' future health-care decisions. Chest. 1998;114(1):284–90.

16.4 Paper III

The pulmonary mycobiome - a study of subjects with and without chronic obstructive pulmonary disease

Einar M. H. Martinsen[1], Tomas M. L. Eagan[1,2], Elise O. Leiten[1], Ingvild Haaland[1], Gunnar R. Husebø[1,2], Kristel S. Knudsen[2], Christine Drengenes[1,2], Walter Sanseverino[3], Andreu Paytuví-Gallart[3], and Rune Nielsen[1,2]

[1]Department of Clinical Science, University of Bergen, Bergen, Norway
[2]Department of Thoracic Medicine, Haukeland University Hospital, Bergen, Norway
[3]Sequentia Biotech SL, Barcelona, Spain

Corresponding author:
Einar Marius Hjellestad Martinsen
Mailing address: einar.martinsen@uib.no
Phone number: +47 977 88 414
Postal address:
Department of Clinical Science, University of Bergen
N-5021 Bergen
Norway

Abstract

Background: The fungal part of the pulmonary microbiome (mycobiome) is understudied. We report the composition of the oral and pulmonary mycobiome in participants with COPD compared to controls in a large-scale single-centre bronchoscopy study (MicroCOPD).

Methods: Oral wash and bronchoalveolar lavage (BAL) was collected from 93 participants with COPD and 100 controls. Fungal DNA was extracted before sequencing of the internal transcribed spacer 1 (ITS1) region of the fungal ribosomal RNA gene cluster. Taxonomic barplots were generated, and we compared taxonomic composition, Shannon index, and beta diversity between study groups, and by use of inhaled steroids.

Results: The oral and pulmonary mycobiomes from controls and participants with COPD were dominated by *Candida*, and there were more *Candida* in oral samples compared to BAL for both study groups. *Malassezia* and *Saroclad*ium were also frequently found in pulmonary samples. No consistent differences were found between study groups in terms of differential abundance/distribution. Alpha and beta diversity did not differ between study groups in pulmonary samples, but beta diversity varied with sample type. The mycobiomes did not seem to be affected by use of inhaled steroids.

Conclusion: Oral and pulmonary samples differed in taxonomic composition and diversity, possibly indicating the existence of a pulmonary mycobiome.

Introduction

Fungi are ubiquitous, and are found in indoor and outdoor environments (1). Due to its direct communication with surrounding air, the respiratory tract is constantly exposed to fungal spores through inhalation (2). Healthy airways possess effective removal of such spores through mucociliary clearance and phagocytosis. In contrast, impaired defence mechanisms, use of immunosuppressant, and frequent use of antibiotics probably predispose for increased fungal growth (2), and all factors are quite frequent in chronic obstructive pulmonary disease (COPD).

The fungal part of the microbiome, the mycobiome, of the lungs is understudied (3), and only three studies have used next generation sequencing to study the mycobiome of the respiratory tract in COPD particularly (4-6). Notably, participants in Cui et al.'s study were also HIV infected, and only ten had COPD (4). The study by Su et al. (5) and Tiew et al. (6) used sputum samples, which are vulnerable to contamination from the high-biomass oral cavity. By contrast, mycobiome studies of other respiratory diseases have evolved rapidly. For instance, a study on asthma patients showed higher fungal burdens in participants receiving corticosteroid therapy (7), while another study has revealed associations between *Aspergillus*-specific immunisation and bronchiectasis severity (8). There is clearly a need for large studies of the mycobiome, with a well-characterised COPD disease population and healthy controls.

The Bergen COPD Microbiome study (short name "MicroCOPD") fills this scientific void (9). Samples were collected from the lower airways of participants with and without COPD using bronchoscopy. The aim of the current paper was threefold: 1) To characterise and compare the oral and pulmonary mycobiomes in a large cohort of participants without lung disease (controls). 2) To characterise the oral and pulmonary mycobiomes of participants with COPD, and contrast it to controls, and finally, 3) To examine whether the mycobiomes were affected by the use of inhaled steroids (ICS) in participants with COPD.

Materials and methods

Study design and population

The study design of the MicroCOPD study has previously been published (9). MicroCOPD was a single-centre observational study carried out in Bergen, Western Norway. Study enrolment was between April 11th, 2013, and June 5th, 2015. The study was conducted in accordance with the declaration of Helsinki and guidelines for good clinical practice. The regional committee of medical ethics Norway north division (REK-NORD) approved the project (project number 2011/1307), and all participants provided written consent.

Both subjects with and without COPD were invited to participate. Participants from two previous cohort studies in our vicinity, the GeneCOPD study and the Bergen COPD cohort study, were contacted regarding participation in the current study, and some participants were recruited through media, the local outpatient clinic, or among hospital staff (9). Potential participants were excluded if they had increased bleeding risk, unstable cardiac conditions, hypercapnia, or hypoxaemia when receiving oxygen supplement, as specified in the study protocol (9). We postponed participation for subjects that had used antibiotics or systemic steroids last 2 weeks prior to participation, and COPD patients should not have been admitted to hospital due to COPD last 2 weeks. Furthermore, participants with symptoms of an ongoing systemic or respiratory infection could not attend, but had to postpone participation. COPD was defined as chronic airway obstruction (low FEV_1/FVC) in presence of respiratory symptoms (10), and the diagnosis was verified by experienced pulmonologists based on spirometry, radiologic imaging, respiratory symptoms, and disease history. Subjects without COPD or other lung diseases were defined as control subjects. 22 control subjects had a ratio of FEV_1/FVC lower than 0.7, but did not have symptoms of COPD.

Data collection

All data collection was performed in our outpatient clinic. A post-bronchodilator spirometry was performed before the bronchoscopy. Study personnel conducted a structured interview regarding contraindications, medication use, comorbidities, smoking habits, and evaluation of dyspnoea. A sterile unsealed bottle of phosphate-buffered saline (PBS) was opened prior to the procedure, and the fluid within was used for all sample fluids, including negative control samples, oral wash (OW), and bronchoalveolar lavage (BAL). The OW sample was taken before the bronchoscopy by gargling 10 ml of the PBS water for 1 minute; collected in a sterile Eppendorf tube. The bronchoscopy was performed with the participant in supine position using oral access. Topical anaesthesia was given by a 10 mg/dose lidocaine oral spray pre-procedurally, and 20 mg/ml lidocaine was delivered per-operatively through a catheter within the bronchoscope's working channel. Light sedation with alfentanil was offered to all. The details of bronchoscopic sampling have been published previously (9). The yields of two fractions of protected BAL of 50 mL were collected from the right middle lobe using a sterile catheter (Plastimed Combicath, prod number 58229.19) inserted in the bronchoscope working channel. The second fraction was used for the current mycobiome analysis. Additionally, a sample from the PBS was taken for each participant directly from the bottle used for that particular participant, without entering the bronchoscope or participant. This PBS sample served as a negative control sample.

Laboratory processing

Fungal DNA was extracted using a combination of enzymatic lysis with lysozyme, mutanolysin, and lysostaphin, and mechanical lysis methods using the FastPrep-24 as described by the manufacturers of the FastDNA Spin Kit (MP Biomedicals, LLC, Solon, OH, USA). Libraries were prepared with a modified version of the Illumina 16S Metagenomic Sequencing Library Preparation guide (Part no. 15044223 Rev. B). The internal transcribed spacer (ITS) 1 region was PCR amplified (increased from 25 to 28 cycles) using primer set ITS1-30F/ITS1-217R, which sequences are

GTCCCTGCCCTTTGTACACA and TTTCGCTGCGTTCTTCATCG (11). A subsequent index PCR was performed with 9 cycles instead of 8. Samples underwent 2x250 cycles of paired-end sequencing in three separate sequencing runs on Illumina HiSeq (Illumina Inc., San Diego, CA, USA).

Bioinformatics

Quantitative Insights into Microbial Ecology (QIIME) 2 (12) version 2019.01 and 2019.10 was chosen as the main pipeline for bioinformatic analyses, and additional R packages were utilised as suited (13). FASTQ files containing all fungal reads were trimmed using the q2-itsxpress plugin (14). Trimmed reads were then denoised, i.e., identification and removal of low-quality reads and chimeric sequences, using the Divisive Amplicon Denoising Algorithm version 2 (DADA2) q2-dada2 plugin (15). DADA2 also generated exact amplicon sequence variants (ASVs). LULU, an R package to curate DNA amplicon data post clustering, was used to exclude artefactual ASVs (16). ASVs present in only one sample, and ASVs with a total abundance less than 10 sequences across all samples, were filtered out. Presumed contaminants were identified using the R package Decontam (17) with the prevalence-based approach (user defined threshold = 0.5), and then removed. Taxonomic assignments were made using a UNITE database for fungi with clustering at 99% threshold level (18) (via q2-feature-classifier (19) classify-sklearn (20)). Resulting ASVs assigned only as *Fungi* at kingdom level, or *Fungi* at kingdom level with unidentified phylum, were manually investigated using the BLASTN program in NCBI (21). ASVs with unambiguous BLASTN results with a high max score were repeatedly assigned to new taxonomic assignments using UNITE databases with fungi or all eukaryotes with different threshold levels (18, 22-24) (via q2-feature-classifier (19) classify-sklearn (20) and classify-consensus-blast (25)), and included for further analyses if the new assignments matched the BLASTN result. ASVs with ambiguous or non-fungal BLASTN results were discarded. Alpha diversity was calculated using Shannon index, and beta diversity metrics (Bray-Curtis dissimilarity and Jaccard similarity coefficient) were estimated using q2-diversity after samples were

rarefied (subsampled without replacement) to 1000 sequences per sample. The rarefaction depth was chosen based on testing with multiple different values and resulting alpha rarefaction plots. We aimed to find a rarefaction depth as high as possible while excluding a minimum of samples.

Data analyses

Statistical analyses of demographical data were analysed with Stata version 15 (26). A flow chart of the bioinformatic process was generated using Flowchart Designer version 3 (http://flowchart.lofter.com). Alpha- and beta diversity analyses including participants with COPD were stratified by GOLD stage (10). Statistical differences in alpha diversity measured with Shannon index were tested with R using Kruskal-Wallis for unpaired variables, and Wilcoxon signed-rank test for paired analyses. Differences in beta diversity between study groups and ICS use were tested with permuted analysis of variance (PERMANOVA) adjusted for sex, age, and percentage of predicted forced expiratory volume in 1 second (FEV_1), and differences in spread with permuted multivariate analysis of beta-dispersion (PERMDISP). Procrustes analyses were performed to check for differences in beta diversity between sample types. PERMANOVA, PERMDISP, and Procrustes were analysed using the Vegan package in R (27). We used both Bray-Curtis and Jaccard distances for the beta diversity analyses. To compare taxonomic composition between pairs of samples we calculated the Yue-Clayton measure of dissimilarity (1-θYC) (28). Furthermore, differences in distributions and relative abundances were evaluated by the Microbiome Differential Distribution Analysis (MicrobiomeDDA) omnibus test (29), the second version of analysis of composition of microbiomes (ANCOM v2, https://github.com/FrederickHuangLin/ANCOM) (30), and the second version of ANOVA-Like Differential Expression (ALDEx2) (31-33) at genus level. A significance level of 0.05 was used in all analyses.

Results

Demographics of participants

The majority of participants with COPD was regular ICS users, and presented with more comorbidities, higher medication use, and poorer lung function measurements (FEV_1/FVC-ratio and mMRC) than controls (Table 1).

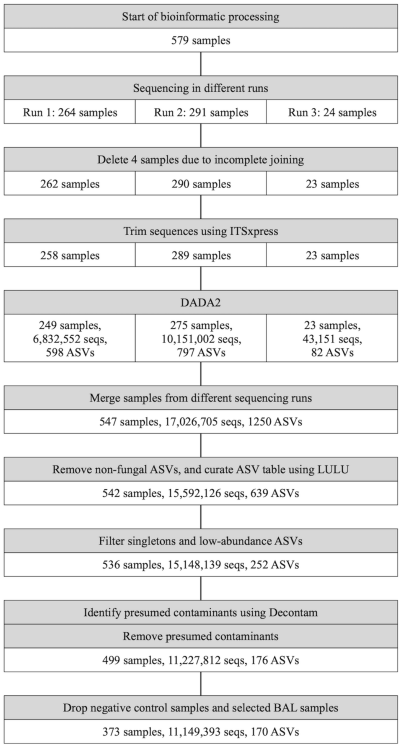
Table 1. Demographics of participants providing fungal samples in the MicroCOPD study

Variable	Control, n=100	COPD, n=93
Age, mean years (SD)	65.6 (8.5)	67.5 (7.6)
Male, sex (%)	57 (57.0)	50 (53.8)
Number of medications, mean (SD)	1.8 (1.7)	5.2 (3.1)
Use of inhaled steroids (%)	-	56 (60.2)
Number of comorbidities, mean (SD)	0.8 (1.0)	1.4 (1.2)
FEV_1, mean % of predicted (SD)	104.0 (12.3)	61.1 (17.3)
FVC, mean % of predicted (SD)	111.7 (13.6)	98.6 (18.7)
FEV_1/FVC-ratio, mean (SD)	0.7 (0.1)	0.5 (0.1)
Pack years, mean (SD)	16.6 (14.3)	30.0 (17.9)
Smoking status (%)		
Daily	24 (24.0)	22 (23.7)
Ex-smokers	58 (58.0)	70 (75.3)
Never	18 (18.0)	1 (1.1)
mMRC Grade 2 and higher (%)*		
Grade 2: Dyspnoea when walking at level ground	3 (3.0)	16 (17.6)
Grade 3: Dyspnoea when walking 100 meters	-	12 (13.2)
Grade 4: Dyspnoea at rest	-	2 (2.2)

FEV_1, forced expiratory volume in 1 second; FVC, forced vital capacity; mMRC, modified medical research council dyspnoea scale. * Two participants with COPD missed information on mMRC.

Flow chart

Fig 1. Flow chart of fungal samples, sequences, and fungal ASVs in the MicroCOPD study

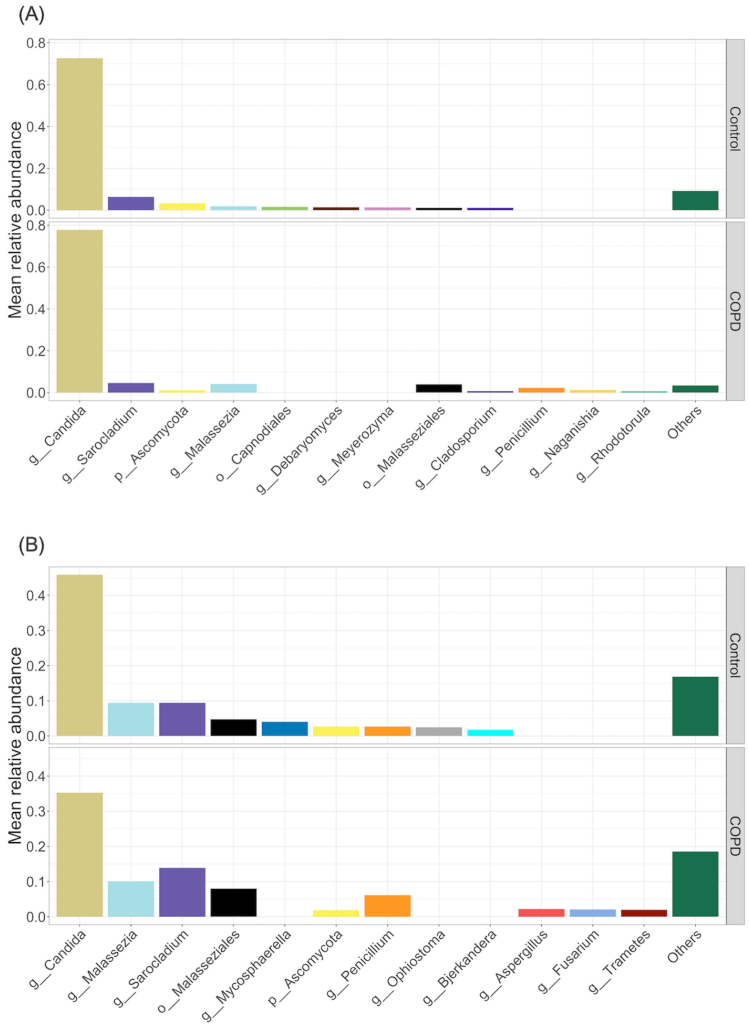


DADA2, Divisive Amplicon Denoising Algorithm version 2; seqs, sequences; ASV, amplicon sequence variant. Samples were sequenced in three different runs before trimming and denoising. Data from different sequencing runs were merged, and then further processed to exclude presumed contaminants and negative control samples prior to analyses.

The bioinformatic processing is shown in Fig 1, and details are given in S1 File. ASVs identified by Decontam as presumed contaminants are listed in S1 Table.

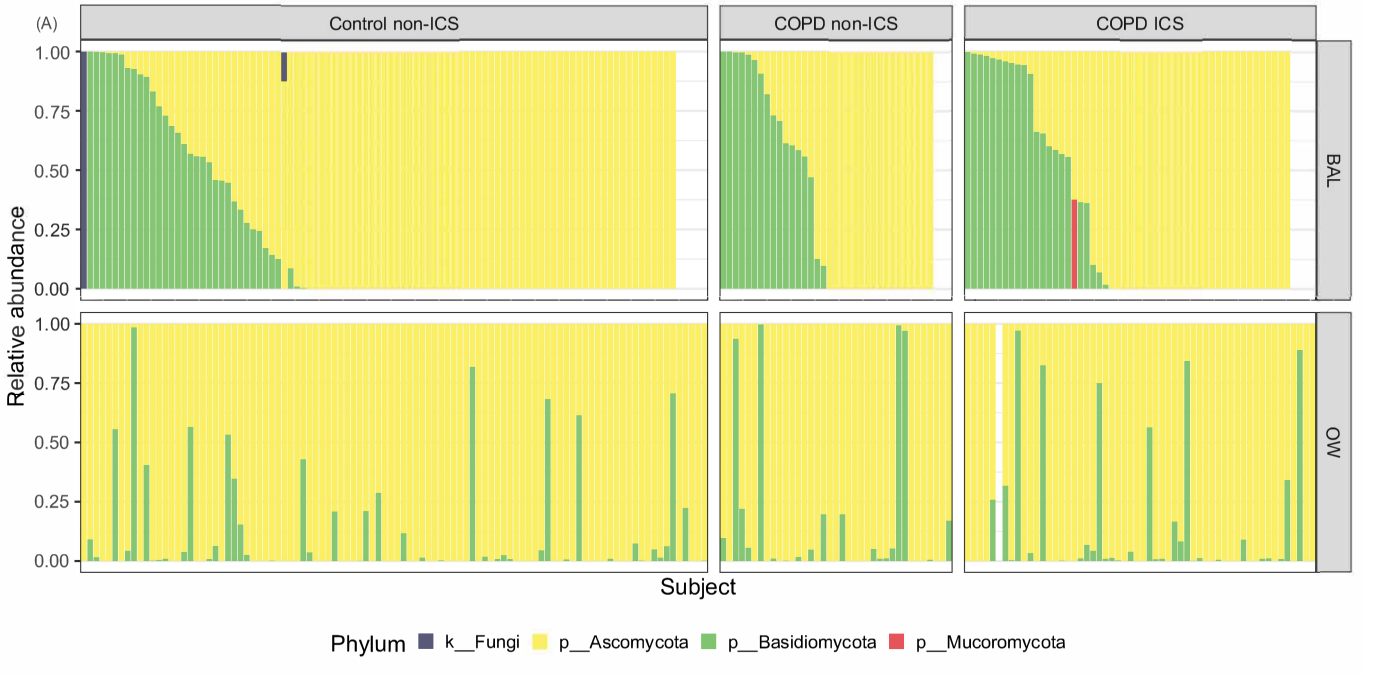
Taxonomy and abundance/distribution testing

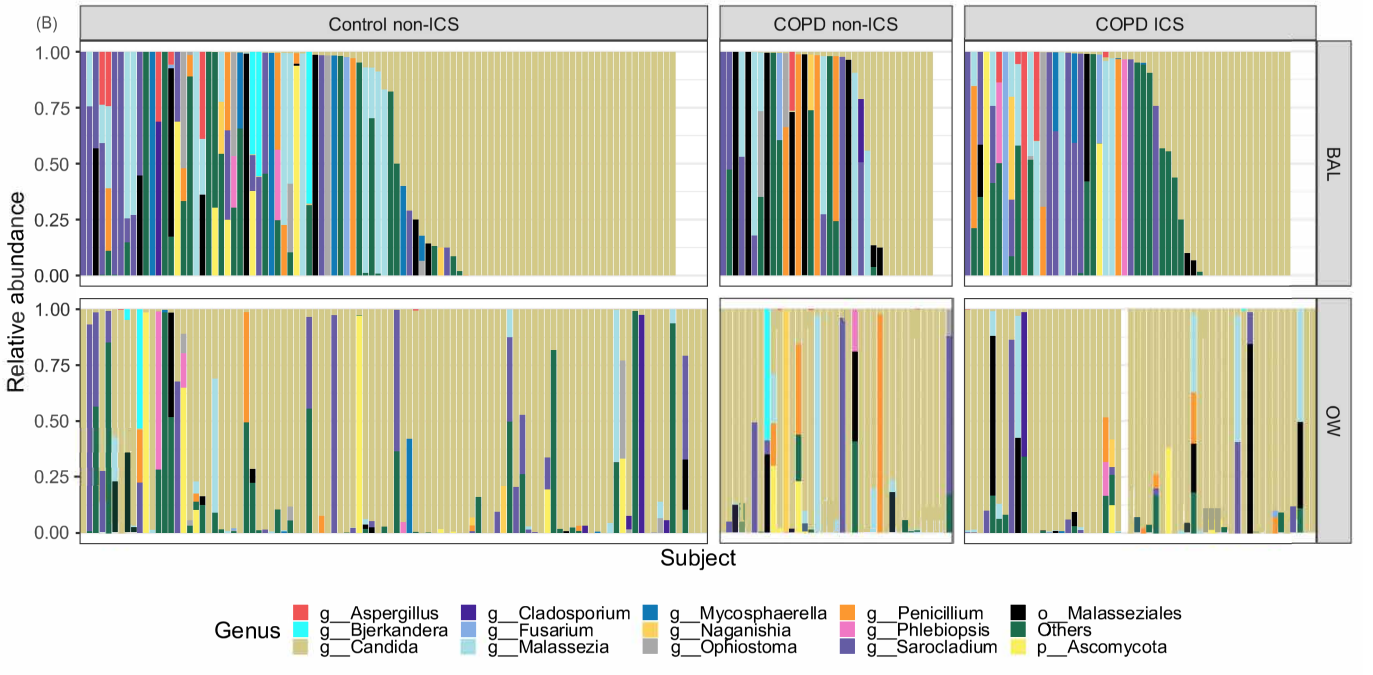
Fig 2. Rank abundance plots using most abundant fungi in (A) oral wash and (B) bronchoalveolar lavage



Plots display the nine most abundant taxa in each group. Remaining, low abundance taxa are merged in the "Others" category. Not all sequences could be assigned taxonomy at the genus level and are therefore displayed as o__*Malasseziales*, p__*Ascomycota*, and o__*Capnodiales*.

Fig 3. Most abundant fungal taxonomic assignments at (A) phylum level and (B) genus level





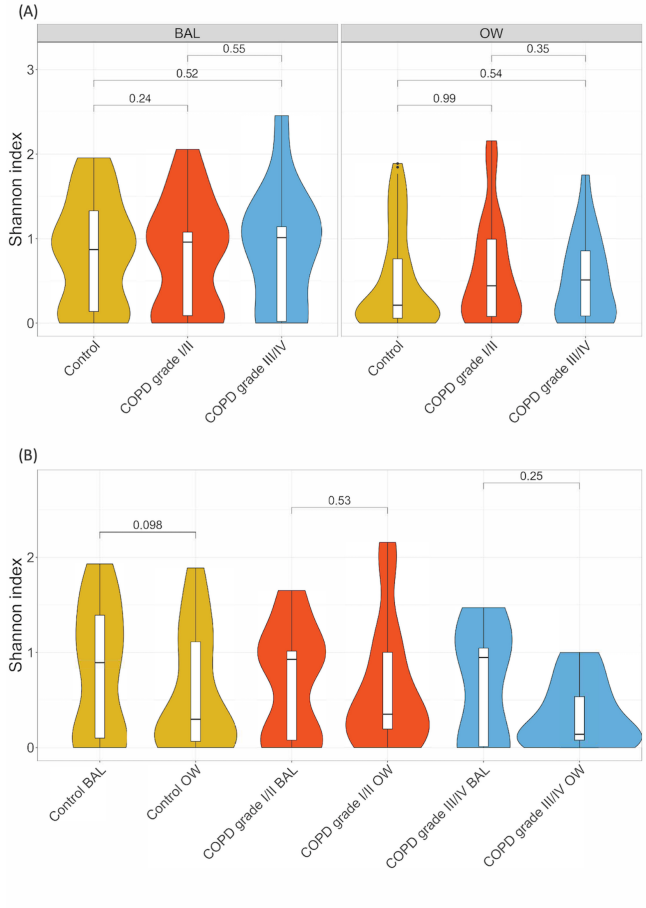
ICS, inhaled steroids; BAL, bronchoalveolar lavage; OW, oral wash. Taxa are sorted on *Ascomycota* in bronchoalveolar lavage samples in Fig 3A, and *Candida* in bronchoalveolar lavage in Fig 3B. Each column represents a sample, and columns from BAL and OW corresponds to each other. That means, a BAL column and the corresponding OW column below show samples from the same participant. Not all sequences could be assigned taxonomy at the phylum or genus level and are therefore displayed as *k__Fungi*, *p__Ascomycota* or *o__Malasseziales*.

The taxonomic composition of the OW and BAL mycobiomes are displayed on group level in the rank abundance plots (Figs 2A and 2B). Both controls and participants with COPD were dominated by *Candida*, particularly in the OW samples, reaching nearly 80% of total mean relative abundance. The relative abundances of *Malassezia* and *Sarocladium* were high in the control and COPD groups. We also plotted percentage of reads belonging to either *Basidiomycota* or *Ascomycota* (S1 Fig). There seemed to be a tendency towards higher proportions of *Basidiomycota* in the COPD group compared to the control group both in the OW and BAL plot. No consistent differences were found between study groups in terms of differential abundance/distribution, and results varied between available statistical tests (S2 Table).

Taxonomy is displayed for each individual participant in Figs 3A and 3B, enabling us to compare OW and BAL samples from this particular participant. We observed intra-individual differences between the sample types for the plotted taxonomic levels (phylum and genus). This was elaborated with Yue-Clayton testing for each OW/BAL/negative control sample pair (S2 Fig). The Yue-Clayton measure is 0 with perfect similarity and 1 with perfect dissimilarity. The average Yue-Clayton measure from OW and BAL samples was 0.63, and 121 out of 180 sample pairs had a Yue-Clayton measure above 0.2. ANCOM and ALDEx2 (S2 Table) found *Candida* to differ in abundance between OW and BAL, also stratified by smoking status and ICS usage. The taxonomy of regular ICS users and non-ICS users was not easily distinguishable (Figs 3A and 3B), and no consistent differences were seen in differential abundance/distribution testing (S3 Table).

Diversity

Fig 4. Alpha diversity plots and comparisons between (A) study groups and (B) sample types



BAL, bronchoalveolar lavage; OW: oral wash. Alpha diversity was evaluated using Shannon index. Statistical differences in alpha diversity were tested using Kruskal-Wallis for unpaired variables (between study groups), and Wilcoxon signed-rank test for paired analyses (between sample types). Number of samples in each group was as follows: Figure 4A (unpaired): Control BAL: 40, control OW: 76, COPD grade I/II BAL: 29, COPD grade I/II OW: 50, COPD grade III/IV OW: 23, COPD grade III/IV BAL: 12. Figure 4B (paired): Control BAL: 31, control OW: 31, COPD grade I/II BAL: 24, COPD grade I/II OW: 24, COPD grade III/IV OW: 9, COPD grade III/IV BAL: 9.

We found no significant differences in alpha diversities between the different study groups or ICS usage in BAL samples, nor between BAL and OW samples (Figs 4A and 4B, S3 Table). Beta-diversity results resembled those of alpha diversity (S3 Fig). However, principal coordinates analysis (PCoA) plots before and after symmetric Procrustes transformation (S4 Fig), indicated that there were significant differences in the composition between OW and BAL samples from the same individual. The Procrustes transformation yields a sum of squared distances (M^2) that specifies how similar sample pairs are. Generally, a M^2 above 0.3 is interpreted as unsimilar. OW and BAL samples clustered differently, and M^2 were 0.953 and 0.8958 for Bray-Curtis and Jaccard, respectively. However, this was statistically significant only for Jaccard ($p = 0.003$).

Discussion

We have reported the oral and pulmonary mycobiome in a large bronchoscopy study, the first of its kind on non-immunocompromised patients and with a large healthy control population. The mycobiomes were dominated by *Candida*, and there were more *Candida* in OW compared to BAL for both study groups. Observed differences in taxonomic composition were not consistent between three different differential abundance/distribution tests. There was no difference in diversity between study groups. No apparent effects were seen on the mycobiomes from ICS usage.

We observed a high *Candida* load in OW samples from controls, in good agreement with previously published studies (34-36). None of these studies included more than 20 individuals. Thus, our data from 100 controls adds valuable data to this field. The healthy lung mycobiome is reported to be highly variable between individuals (4, 6, 7, 37-39). Some of the most abundant taxa are *Candida* (6)*, Davidiellaceae* (37), *Cladosporium* (37-39), *Saccharomyces* (4, 6), *Penicillium* (4), *Debaryomyces* (38), *Aspergillus* (7, 37), *Eremothecium* (39), *Systenostrema* (39), and *Malasseziales* (7). The listed studies include few participants, ranging from 10 (7) to 47 (6). BAL samples from controls in the current study were dominated by *Candida*, followed by *Malassezia* and *Sarocladium*. That

Candida, one of the most well-known fungal pathogens (40), resides in the lungs of healthy individuals is clinically interesting. It has been shown that colonising *Candida* in the gut could become invasive due to certain triggers (41), and similar mechanisms are not unlikely to happen in the lungs. Primer bias might explain some of the observed differences between our study and previous studies (42), and our chosen primer set has shown improved coverage of *Candida* compared to the ITS1 – ITS2 primer set used by two (37, 38) of the listed papers above (11). Furthermore, *Malassezia* are common skin commensals, and despite protected sampling and the bioinformatic contamination removal (Decontam), we cannot exclude contamination *per se*. Also, some reports indicate that some extraction protocols and primers might be less suited to *Malassezia* (36, 43). Different DNA extraction methods and primers could thus explain the observed differences in *Malassezia* proportions between our study and others.

OW samples differed from the BAL samples for all measures including taxonomy, Yue-Clayton measures, and beta diversity. Cui et al. reported that OW and BAL overlapped in PCoA plots from healthy individuals, while induced sputum (IS) samples clustered separately (4). In agreement with our result, they also found more *Candida* in OW and IS, compared to BAL. They discovered that 39 fungal species were disproportionately more abundant in the BAL and 203 species in the IS, as compared with the OW. We could not replicate this latter finding, but differences could be explained by the different methodologies applied.

Only three previous studies have explored the lung mycobiome in COPD (4-6). Cui et al. found that the lung mycobiome in HIV-infected individuals with COPD (n=10) was associated with an increased prevalence of *Pneumocystis jirovecii*, as compared to HIV-positive individuals without COPD (n=22) (4). No *Pneumocystis* was observed in our data. However, *Pneumocystis* is known to be associated with HIV, and the *Pneumocystis* genome only includes one copy of the ITS1 locus, which could result in a negative sequencing result (44).

Both Su et al. (5) and Tiew et al. (6) collected sputum samples from COPD patients. When Tiew compared to healthy subjects they found high abundances of *Candida* in both groups, but also found increased alpha diversity in COPD (6). Su investigated samples during exacerbations, and found *Candida*, *Phialosimplex*, *Aspergillus*, *Penicillium*, *Cladosporium,* and *Eutypella* (5). Both studies utilised sputum samples, which hampers direct comparison to our BAL samples. Indeed, IS samples have been shown to cluster separately from BAL samples in PCoA ordinations (4).

Few differences were seen when we compared the mycobiomes from controls and participants with COPD. However, hypothesis testing of microbiome compositional data is an ongoing research area without standardisation. Thus, we chose to perform three different tests with different foundations. ANCOM v2 and ALDEx2 agreed there was no significantly differential abundant taxa between study groups. MicrobiomeDDA tests the difference in the entire distribution, taking abundance, prevalence, and dispersion all into account, and detected significant taxa differences between study groups both in OW and BAL (S2 Table). These conflicting results complicate a general conclusion.

Some studies on inflammatory bowel disease, and cystic fibrosis (CF) have found dysbiosis to be expressed in terms of the *Basidiomycota* to *Ascomycota* ratio (45, 46). Most known fungal pathogens are found in the *Ascomycota* phylum. In our study, medians of the *Basidiomycota* to *Ascomycota* ratios were all 0 from different study groups in OW and BAL separately, in line with the 0.03 median found in a CF study (46). That means, despite a higher *Basidiomycota* fraction in COPD compared to controls in our data (S1 Fig), the majority of samples were dominated by *Ascomycota*.

Some limitations of the current study deserve mentioning. First, a longitudinal study with analyses on interactions between fungi and other kingdoms, and between fungi and host responses, could have provided more insight into the details of the COPD mycobiome.

Secondly, contamination is particularly problematic for mycobiome studies because of airborne particles, and samples from the lower respiratory tract are especially vulnerable due to the low biomass. We countered this by using protected sampling methods, and collecting negative control samples, subject to the same laboratory protocol as the procedural samples, for each procedure. These samples were used for contamination removal through the R package Decontam, and subject to detailed analyses (S2 Fig, S4 Table, and S5 Fig). Third, we did not include positive controls or mock communities in our project. Fourth, ITS primers are biased (42, 43), possibly explaining the low prevalence of *Aspergillus* and difficulties identifying *Yarrowia lypolytica* in our data. Still, ITS is the recommended marker-gene region for fungal studies (47), though no consensus seems to prevail whether ITS1 or ITS2 should be used (43, 48, 49). Fifth, all mycobiome studies suffer from a fungal dual naming system (50), and also suffer from incomplete reference libraries and inconsistencies due to taxonomic reassignments (2). We manually reviewed every sequence assigned only as "*k__Fungi*" to secure the best possible taxonomy. Finally, confounding factors and batch effects could not be ruled out. We included a thorough examination of diversity and differential abundance/distribution testing to look for potential confounding effects from several important clinical parameters (S3 Table). No apparent effects were seen. We observed no statistically significant difference in alpha diversity between sequencing runs (S5 Fig), but there might have been an effect on beta diversity (S6 Fig), particularly between sequencing run 1 and 2 (S5 Table). In terms of differential abundance/distribution, it seemed that *Sarocladium* differed in abundance/distribution between sequencing run 1 and 2 (S7 Table).

Studies on the lung mycobiome are still in their infancy, and results from the current study add knowledge to an understudied area. Samples from the mouth differed from pulmonary samples both in controls and participants with COPD, which may indicate the existence of a pulmonary mycobiome. Certain inferences on taxonomic compositions differences between study groups could not be made due to inconsistent results among the

differential abundance/distribution tests used. ICS use could not be seen to significantly affect the lung mycobiome. These findings should be confirmed in other study populations before we can conclude that ICS use has no harmful effect on the lung mycobiome.

Acknowledgements

The MicroCOPD is a large study with many co-workers. The authors wish to give their thanks to Per Sigvald Bakke, Harald G Wiker, Øistein Svanes, Sverre Lehmann, Marit Aardal, Tuyen Hoang, Tharmini Kalananthan, Randi Sandvik, Eli Nordeide, Hildegunn Bakke Fleten, Tove Folkestad, Ane Aamli Gagnat, Solveig Tangedal, Kristina Apalseth, Stine Lillebø, and Lise Østgård Monsen (Haukeland University Hospital and University of Bergen).

References

1. Zhang Z, Reponen T, Hershey GK. Fungal Exposure and Asthma: IgE and Non-IgE-Mediated Mechanisms. Curr Allergy Asthma Rep. 2016;16(12):86.
2. Weaver D, Gago S, Bromley M, Bowyer P. The Human Lung Mycobiome in Chronic Respiratory Disease: Limitations of Methods and Our Current Understanding. Current Fungal Infection Reports. 2019;13(3):109-19.
3. Tipton L, Ghedin E, Morris A. The lung mycobiome in the next-generation sequencing era. Virulence. 2017;8(3):334-41.
4. Cui L, Lucht L, Tipton L, Rogers MB, Fitch A, Kessinger C, et al. Topographic diversity of the respiratory tract mycobiome and alteration in HIV and lung disease. Am J Respir Crit Care Med. 2015;191(8):932-42.
5. 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.
6. Tiew PY, Dicker AJ, Keir HR, Poh ME, Pang SL, Mac Aogáin M, et al. A high-risk airway mycobiome is associated with frequent exacerbation and mortality in COPD. Eur Respir J 2020; in press (https://doi.org/10.1183/13993003.02050-2020).
7. Fraczek MG, Chishimba L, Niven RM, Bromley M, Simpson A, Smyth L, et al. Corticosteroid treatment is associated with increased filamentous fungal burden in allergic fungal disease. J Allergy Clin Immunol. 2018;142(2):407-14.
8. Mac Aogáin M, Chandrasekaran R, Lim AYH, Low TB, Tan GL, Hassan T, et al. Immunological corollary of the pulmonary mycobiome in bronchiectasis: the CAMEB study. Eur Respir J. 2018;52(1).
9. 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.
10. Global Initiative for Chronic Obstructive Lung Disease. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease, 2020. Available from: https://goldcopd.org/.

11. Usyk M, Zolnik CP, Patel H, Levi MH, Burk RD. Novel ITS1 Fungal Primers for Characterization of the Mycobiome. mSphere. 2017;2(6).
12. 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.
13. R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available online at https://www.R-project.org/.
14. Rivers AR, Weber KC, Gardner TG, Liu S, Armstrong SD. ITSxpress: Software to rapidly trim internally transcribed spacer sequences with quality scores for marker gene analysis. F1000Res. 2018;7:1418.
15. 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.
16. Froslev TG, Kjoller R, Bruun HH, Ejrnaes R, Brunbjerg AK, Pietroni C, et al. Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. Nat Commun. 2017;8(1):1188.
17. 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.
18. UNITE Community (2019): UNITE QIIME release for Fungi. Version 18.11.2018. UNITE Community. https://doi.org/10.15156/BIO/786334.
19. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, et al. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome. 2018;6(1):90.
20. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. Journal of machine learning research. 2011;12(Oct):2825–2830.

21. Zhang Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA sequences. J Comput Biol. 2000;7(1-2):203-14.
22. UNITE Community (2019): UNITE QIIME release for Fungi 2. Version 18.11.2018. UNITE Community. https://doi.org/10.15156/BIO/786349.
23. UNITE Community (2019): UNITE QIIME release for eukaryotes. Version 18.11.2018. UNITE Community. https://doi.org/10.15156/BIO/786335.
24. UNITE Community (2019): UNITE QIIME release for eukaryotes 2. Version 18.11.2018. UNITE Community. https://doi.org/10.15156/BIO/786350.
25. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10(1):421.
26. StataCorp. 2017. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC.
27. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. 2019. vegan: Community Ecology Package. Available from: https://CRAN.R-project.org/package=vegan.
28. Yue JC, Clayton MK. A Similarity Measure Based on Species Proportions. Communications in Statistics - Theory and Methods. 2005;34(11):2123-31.
29. Chen J, King E, Deek R, Wei Z, Yu Y, Grill D, et al. An omnibus test for differential distribution analysis of microbiome sequencing data. Bioinformatics. 2018;34(4):643-51.
30. Mandal S, Van Treuren W, White RA, Eggesbø M, Knight R, Peddada SD. Analysis of composition of microbiomes: a novel method for studying microbial composition. Microb Ecol Health Dis. 2015;26:27663.
31. Fernandes AD, Macklaim JM, Linn TG, Reid G, Gloor GB. ANOVA-Like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq. PLOS ONE. 2013;8(7):e67019.
32. Fernandes AD, Reid JN, Macklaim JM, McMurrough TA, Edgell DR, Gloor GB. Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq,

16S rRNA gene sequencing and selective growth experiments by compositional data analysis. Microbiome. 2014;2:15.
33. Gloor GB, Macklaim JM, Fernandes AD. Displaying Variation in Large Datasets: Plotting a Visual Summary of Effect Sizes. Journal of Computational and Graphical Statistics. 2016;25(3):971-9.
34. Ghannoum MA, Jurevic RJ, Mukherjee PK, Cui F, Sikaroodi M, Naqvi A, et al. Characterization of the oral fungal microbiome (mycobiome) in healthy individuals. PLoS Pathog. 2010;6(1):e1000713.
35. Mukherjee PK, Chandra J, Retuerto M, Sikaroodi M, Brown RE, Jurevic R, et al. Oral mycobiome analysis of HIV-infected patients: identification of Pichia as an antagonist of opportunistic fungi. PLoS Pathog. 2014;10(3):e1003996.
36. Dupuy AK, David MS, Li L, Heider TN, Peterson JD, Montano EA, et al. Redefining the human oral mycobiome with improved practices in amplicon-based taxonomy: discovery of Malassezia as a prominent commensal. PLoS One. 2014;9(3):e90899.
37. Charlson ES, Diamond JM, Bittinger K, Fitzgerald AS, Yadav A, Haas AR, et al. Lung-enriched organisms and aberrant bacterial and fungal respiratory microbiota after lung transplant. Am J Respir Crit Care Med. 2012;186(6):536-45.
38. Bittinger K, Charlson ES, Loy E, Shirley DJ, Haas AR, Laughlin A, et al. Improved characterization of medically relevant fungi in the human respiratory tract using next-generation sequencing. Genome Biol. 2014;15(10):487.
39. van Woerden HC, Gregory C, Brown R, Marchesi JR, Hoogendoorn B, Matthews IP. Differences in fungi present in induced sputum samples from asthma patients and non-atopic controls: a community based case control study. BMC Infect Dis. 2013;13:69.
40. Nguyen LD, Viscogliosi E, Delhaes L. The lung mycobiome: an emerging field of the human respiratory microbiome. Front Microbiol. 2015;6:89.
41. Lagunes L, Rello J. Invasive candidiasis: from mycobiome to infection, therapy, and prevention. Eur J Clin Microbiol Infect Dis. 2016;35(8):1221-6.

42. Bellemain E, Carlsen T, Brochmann C, Coissac E, Taberlet P, Kauserud H. ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. BMC Microbiol. 2010;10:189.
43. Hoggard M, Vesty A, Wong G, Montgomery JM, Fourie C, Douglas RG, et al. Characterizing the Human Mycobiota: A Comparison of Small Subunit rRNA, ITS1, ITS2, and Large Subunit rRNA Genomic Targets. Front Microbiol. 2018;9:2208.
44. Delhaes L, Monchy S, Fréalle E, Hubans C, Salleron J, Leroy S, et al. The airway microbiota in cystic fibrosis: a complex fungal and bacterial community--implications for therapeutic management. PLoS One. 2012;7(4):e36313.
45. Sokol H, Leducq V, Aschard H, Pham HP, Jegou S, Landman C, et al. Fungal microbiota dysbiosis in IBD. Gut. 2017;66(6):1039-48.
46. Soret P, Vandenborght LE, Francis F, Coron N, Enaud R, Avalos M, et al. Respiratory mycobiome and suggestion of inter-kingdom network during acute pulmonary exacerbation in cystic fibrosis. Sci Rep. 2020;10(1):3589.
47. Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, et al. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc Natl Acad Sci U S A. 2012;109(16):6241-6.
48. McTaggart LR, Copeland JK, Surendra A, Wang PW, Husain S, Coburn B, et al. Mycobiome Sequencing and Analysis Applied to Fungal Community Profiling of the Lower Respiratory Tract During Fungal Pathogenesis. Front Microbiol. 2019;10:512.
49. Nilsson RH, Anslan S, Bahram M, Wurzbacher C, Baldrian P, Tedersoo L. Mycobiome diversity: high-throughput sequencing and identification of fungi. Nat Rev Microbiol. 2019;17(2):95-109.
50. Hawksworth DL, Crous PW, Redhead SA, Reynolds DR, Samson RA, Seifert KA, et al. The amsterdam declaration on fungal nomenclature. IMA Fungus. 2011;2(1):105-12.

Supporting information

S1 Fig. Percentage of reads belonging to Ascomycota/Basidiomycota in (A) oral wash and (B) bronchoalveolar lavage

S1 File. Bioinformatic processing

S1 Table. Presumed fungal contaminants identified by Decontam in the MicroCOPD study
ASV: amplicon sequence variant. The R package "Decontam" identified the ASV IDs above as contaminants. ASVs presumed to be contaminants were removed prior to analyses.

S2 Fig. Yue-Clayton measures from (A) controls and (B) participants with COPD
YC: Yue-Clayton measure. OW: oral wash, NCS: negative control sample, BAL: bronchoalveolar lavage. A Yue-Clayton measure of 0 means identical sample pairs, while a Yue-Clayton measure of 1 means unidentical sample pairs.

S2 Table. Differential abundance/distribution testing on fungi in the MicroCOPD study using ANCOM v2, MicrobiomeDDA, and ALDEx2
ANCOM v2: the second version of analysis of composition of microbiomes, MicrobiomeDDA: Microbiome Differential Distribution Analysis omnibus test, ALDEx2: the second version of ANOVA-Like Differential Expression, OW: oral wash, BAL: bronchoalveolar lavage. The most conservative value in ANCOM v2 has been used in the analyses (i.e. 0.9). Significance level=0.05. Never- and ex-smokers were merged into non-smokers. The ALDEx2 approach works poorly if there are only a small number of taxa (less than about 50), so some groups were not analysed.

S3 Fig. Principal coordinates analysis plots by (A) study group and (B) inhaled steroids use
Differences in beta diversity were tested with permuted analysis of variance (PERMANOVA) adjusted for sex, age, and percentage of predicted FEV_1 (permutations = 10000). No significant differences were seen in spread/dispersion (permutations = 1000).

S3 Table. Taxonomy and diversity comparisons of selected clinical variables in the MicroCOPD study divided by sample type and study group
PERMANOVA: permuted analysis of variance, OW: oral wash, BAL: bronchoalveolar lavage, AN: ANCOM v2, M: MicrobiomeDDA, AL: ALDEx2, sign: significant, FEV_1: forced expiratory volume in 1 second. Analyses on FEV_1 were omitted for each study group separately due to a majority of controls having above 80% of predicted, and a majority of participants with COPD having below 80% of predicted. Diversity analyses on smoking habits in BAL samples from controls were omitted due to a lack of current smokers. Analyses on smoking habits were done by comparing current vs non-current smokers.

S4 Fig. Principal coordinates analysis plots by sample type (A) before and (B) after symmetric Procrustes transformation

OW: oral wash, BAL: bronchoalveolar lavage. Arrows are drawn from the OW sample to the BAL sample from the same participant. Non-randomness ("significance") between the two configurations was tested with the protest function including the three first axis from the PCoA and specifying 999 permutations.

S4 Table. Summary of read/sequence counts in the MicroCOPD study

NCS: Negative control sample, OW: oral wash, BAL: bronchoalveolar lavage, DADA2: Divisive Amplicon Denoising Algorithm version 2.

S5 Fig. Plot of Qubit concentrations and comparisons between sample types

BAL: bronchoalveolar lavage, NCS: negative control sample, OW: oral wash. Statistical differences in Qubit concentrations were tested using Wilcoxon signed-rank test as a paired test.

S5 Table. Beta diversity comparisons using Bray-Curtis and Jaccard distances. Comparisons were done (A) merged and (B) pairwise

OW: oral wash, BAL: bronchoalveolar lavage, yrs: years. Differences in beta diversity were tested with permuted analysis of variance (PERMANOVA) adjusted for sample type, study group, sex, and age (permutations = 10000).

S6 Fig. Alpha diversity plots and comparisons between sequencing runs

BAL: bronchoalveolar lavage, OW: oral wash. Alpha diversity was evaluated using Shannon index. Statistical differences in alpha diversity were tested using Kruskal-Wallis.

S6 Table. Permuted multivariate analysis of beta-dispersion using Bray-Curtis and Jaccard distances. Comparisons were done (A) merged and (B) pairwise

OW: oral wash, BAL: bronchoalveolar lavage.

S7 Fig. Principal coordinates analysis plots divided by sequencing run

OW: oral wash, BAL: bronchoalveolar lavage.

S7 Table. Differential abundance/distribution testing on sequencing run using ANCOM v2, MicrobiomeDDA, and ALDEx2

ANCOM v2: the second version of analysis of composition of microbiomes, MicrobiomeDDA: Microbiome Differential Distribution Analysis omnibus test, ALDEx2: the second version of ANOVA-Like Differential Expression, OW: oral wash, BAL: bronchoalveolar lavage. The ALDEx2 approach works poorly if there are only a small number of features (less than about 50). The most conservative value in ANCOM v2 has been used in the analyses (i.e. 0.9).

Errata for
The mycobiome in COPD: Descriptive and longitudinal analysis, and participation in research bronchoscopy studies

Einar Marius Hjellestad Martinsen

Thesis for the degree philosophiae doctor (PhD)

at the University of Bergen

_____ _____

(date and sign. of candidate) (date and sign. of faculty)

Errata

Page 54 Incorrect number: “That means, 62 participants underwent two bronchoscopies (re-bronchoscopy), and 11 participants had three bronchoscopies” corrected to “That means, 62 participants underwent two bronchoscopies (re-bronchoscopy), and 12 participants had three bronchoscopies”.

Page 85 Incomplete and incorrect statement: “One benefit from our literature review is that we included keywords both as subject headings and as text words in PubMed or explosion searches in Embase, thereby broadening our search. It may take a few months before subject headings are added to an article, which confirms the importance of including text words and explosion searches” corrected to “One benefit from our literature review is that we included keywords both as subject headings and as text words in PubMed or explosion searches and multipurpose terms in Embase, thereby broadening our search. It may take a few months before subject headings are added to an article, which confirms the importance of including text words and multipurpose terms”.

Page 132 Incorrect statement: “But the high *Candida* dominance could create a visual illusion because of the less coloured BAL plot” corrected to “But the high *Candida* dominance could create a visual illusion because of the less coloured OW plot”.

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

uib.no

ISBN: 9788230844557 (print)
9788230862599 (PDF)