Supplementary material

The airway microbiota and exacerbations of COPD

Elise Orvedal Leiten[1], Rune Nielsen[1,2], Harald Gotten Wiker[1,3], Per Sigvald Bakke[1], Einar Marius Hjellestad Martinsen[1], Christine Drengenes[1,2], Solveig Tangedal[1,2], Gunnar Reksten Husebø[1,2], Tomas Mikal Lind Eagan[1,2]

Affiliations:

[1]Department of Clinical Science, University of Bergen
[2]Department of Thoracic Medicine, Haukeland University Hospital, Bergen
[3]Department of Microbiology, Haukeland University Hospital, Bergen

Methodological considerations

Collection of the microbial samples.

In the current study, we explore the lower airways microbiome by two different sampling methods, sterile brushes and protected broncheoalveolar lavage (BAL). Both methods have advantages and weaknesses, and therefore may enhance each other when both are employed.

When a bronchoscope is inserted through the upper airways, invariably the tip of the bronchoscope will come into contact with the upper airways' microbiome. We employed no suction prior to placement with the bronchoscope below the vocal cords, yet even though there is no 100% avoidance of contaminating the working channel from the microbiome that may have attached itself on the tip of the bronchoscope during the passage through the upper airways.

However, the protected specimen brushes will be truly sterile if handled correctly, since they have a wax sealed tip which can by released once the brush is visually in safe distance from the bronchoscope tip, and then be applied by visual inspection in the bronchi. The brush shall then be retracted to safe distance within their plastic sheet covering by visual guidance before pulling out, and the brush tip must be cut by sterile scissors into a sterile tube, as was done in the current study, and rarely reported upon in detail by others. The main problem with the brushes is that the area sampled is distinct but small, and the biomass per one brush tends to be very small. To mitigate this, we used three brushes in each patient at each sampling site, and each brush was brushed ten times back and forth without excessive force, at each sampling site.

Sampling BAL offers the advantage of obtaining microbial biomass from a much larger geographical area of a bronchial tree, and also, unrelated to the current analyses, the potential for fluid for biomarker analyses and differential cell counts.

However, BAL have contamination issues, especially if one merely installs the fluid through the working channel, likely contaminated by previous suctioning, and not usually fully sterile by regular scope washing procedures. We used a sterile (Combicath) inner catheter with a sealed tip to install the fluid. Even though, our BAL procedure can still not be 100% sterile by design, since when the fluid is instilled, some minute amounts of fluid will touch the tip of the bronchoscope, and thus possibly be contaminated. Another potential issue with BAL is yield. Although a controlled similar amount of fluid may be installed, yield will vary between patients. Especially in patients with a large degree of emphysema, much more fluid is "lost" to the periphery, and not returned. How to correct for this is uncertain, as the effects on the concentration of microbial mass is uncertain. Like other researchers before us, we choose not to attempt any adjustment for yield. This potential problem is presumably absent with the use of brushes, however there is invariably variation of force and length of each brushing, for each patient and each bronchoscopist, illustrating that no sampling method can be 100% standardized.

Control samples

Each bronchoscopy day, we opened a new, sealed, sterile 500mL Phosphate Buffered Saline (PBS) bottle. From this bottle, PBS was drawn with sterile syringes, used for BAL sampling, the fluid used for Oral wash sampling, and also the fluid in which the sterile brushes were immersed in sterile tubes. Also, we collected small amounts of this fluid in sterile tubes and froze them. These latter samples were our negative control samples. Thus, the negative

controls are PBS fluid - which all other samples will have been in contact with - but clean from the sterile bottle - the very same bottle all sample fluids were drawn from. For every single procedure in each examined participant, we thus have a negative control sample available for microbial sequencing. Sequencing the negative controls enabled us to assess sequences which must be considered contaminants, since they either must have been acquired from the PBS bottle, or from the laboratory handling. These negative control samples were extremely useful for the Decontam algorithm, allowing the use of the prevalence-based method. We have previously examined in detail the possible contamination in the MicroCOPD study, showing the usefulness of these negative control samples [1].

Further details on data curation

All samples in our prediction of exacerbation analysis belong to the larger MicroCOPD study, collected with an identical protocol, and thus likely to see the same sampling and laboratory related contaminants, in addition to other study-wide phenomena such as index bleeding. The samples were analysed on one MiSeq instrument, however over different 30 runs due to the large number of samples (>2500). Since the quality scores will differ for the different runs, and the Divisive Amplicon Denoising Algorithm version 2 (DADA2) algorithm uses the quality scores for the denoising, DADA2 requires that each run is imported and denoised separately before merging. In addition to specifying the minimum phred score per base, the DADA2 relies on visual inspection of the quality plot of the forward and reverse reads, and then deciding where to cut the forward and reverse reads per run. Invariably, this is subjective to the eye of the beholder, and somewhat arbitrary over 30 runs.

Thus, we chose to pre-trim the sequences by the software tool Trimmomatic (v0.39) [2], using the following parameters:

-phred33, and algorithm: HEADCROP:17:300:21:300 LEADING:20 TRAILING:25 SLIDINGWINDOW:4:20 MINLEN:220.

The HEADCROP:17:300:21:300 ensures the primers which were 17 bp forward and 21 bp reverse were removed. Since quality was regularly lower for the very first as well as the tail end of the reads, LEADING and TRAILING were set to 20 and 25 respectively. SLIDING WINDOW represents the possibility not present in DADA2, effectively picking out low quality reads within the middle range of the reads. However, the main advantage to using Trimmomatic in our study was to ensure all runs were treated exactly the same regarding quality requirements.

After this trimming, each of the 30 runs were imported into QIIME 2 (v2018.8) and denoised with DADA2, using the command:

qiime dada2 denoise-paired --i-demultiplexed-seqs /Volumes/.../RUNX_workingfiles/RUNX_importdemux/RUNXimport-trimmed-demux.qza
--output-dir /Volumes/LaCie/RUN6_workingfiles/RUNX_denoisedpaired
--p-trunc-len-f 0 --p-trunc-len-r 0 --p-trim-left-f 0 --p-trim-left-r 0
[already trimmed by trimmomatic]
--p-chimera-method consensus

And after that the files from the 30 runss were merged, before another round of chimera removal using VSEARCH.

Handling of contamination using information from the control samples

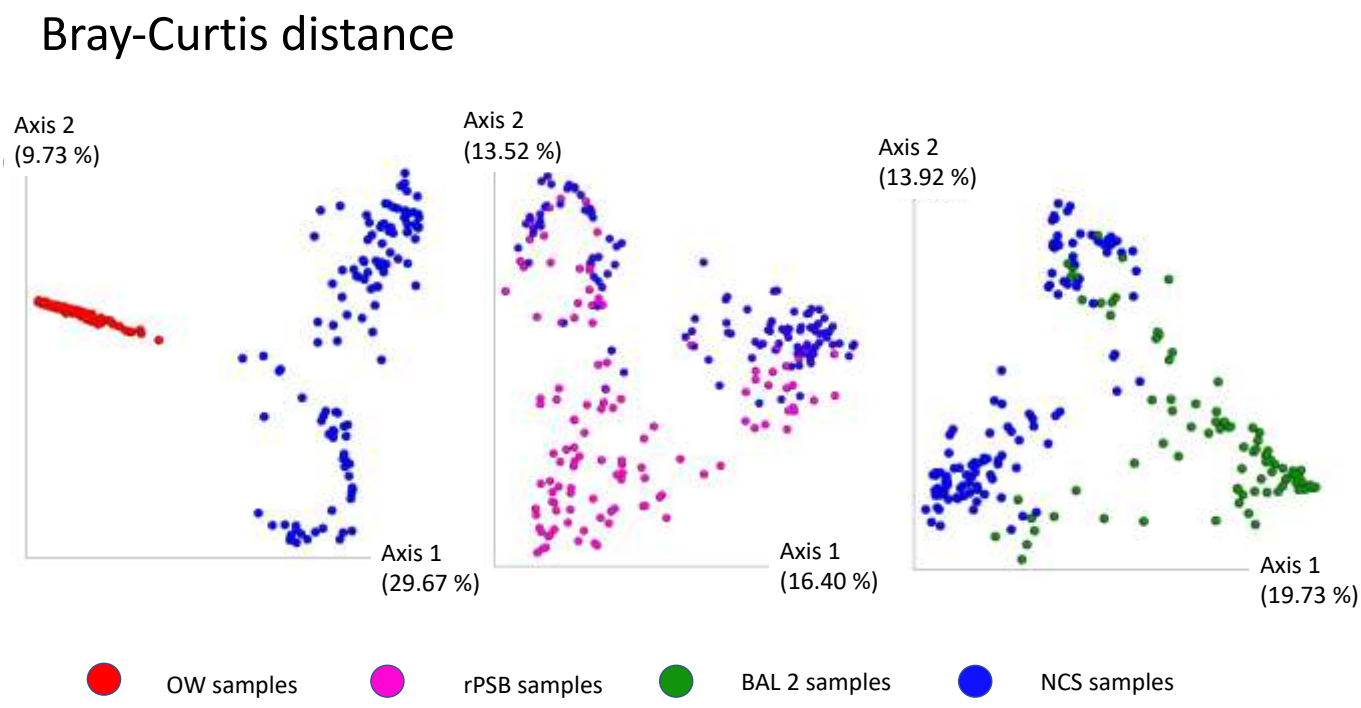
In order to do effective contamination removal, taking into account all samples in the relevant runs, we kept all samples as long as possible in the up-stream processing. After the

steps described above, the ASV table was exported for use in R (v3.4.1) with the Decontam algorithm (v1.1.2) [3].

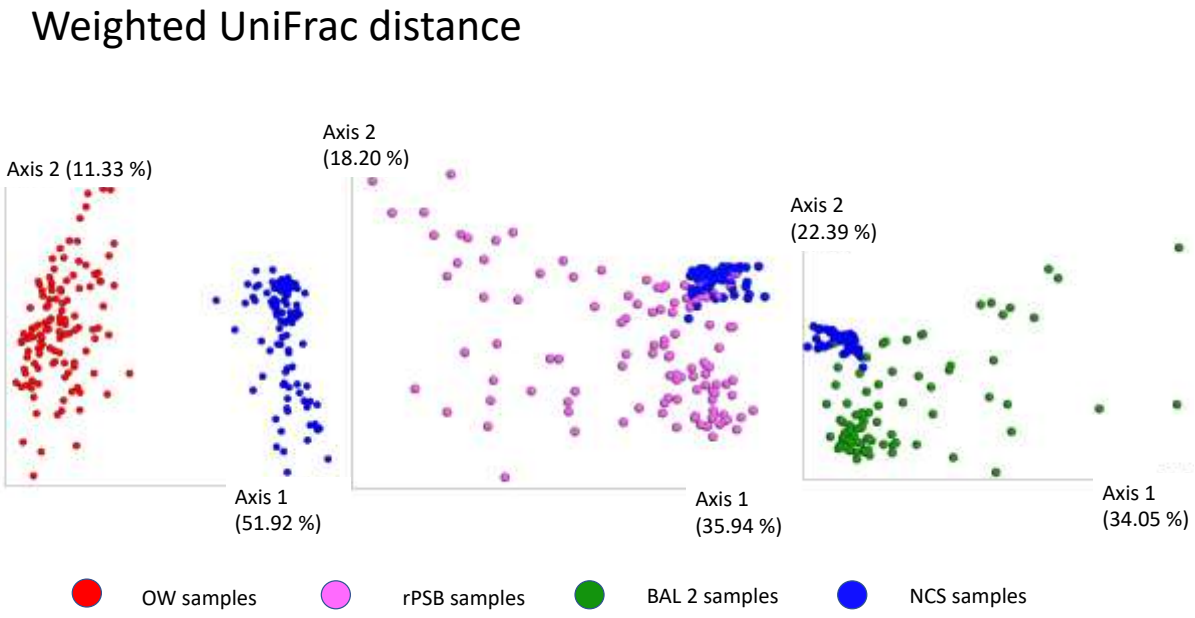
We chose the prevalence method which is adviced for low-biomass samples such as ours, and where the contaminants are identified by increased prevalence in negative controls relative to the true biological samples. A key parameter is the threshold parameter, which is the probability threshold below which a contaminant should be rejected in favor of a non-contaminant. After running the algorithm, frequency distribution plots can be run to assess how well the contaminants and non-contaminants are distinguished between each other. We ran Decontam with thresholds 0.3, 0.4 and 0.5, and found that a threshold of 0.3 struck the best balance between distinguishing likely contaminants, and retaining likeli non-contaminant reads. With threshold = 0.3, we identified 280 ASVs deemed likely contaminants.

To see how the negative controls differed from the biological samples, we provide principal coordinate analysis (PCoA) plots with Bray-Curtis, weighted and unweighted UniFrac distances, showing the clustering of samples before the Decontam step (Supplemetary figures 1-3).

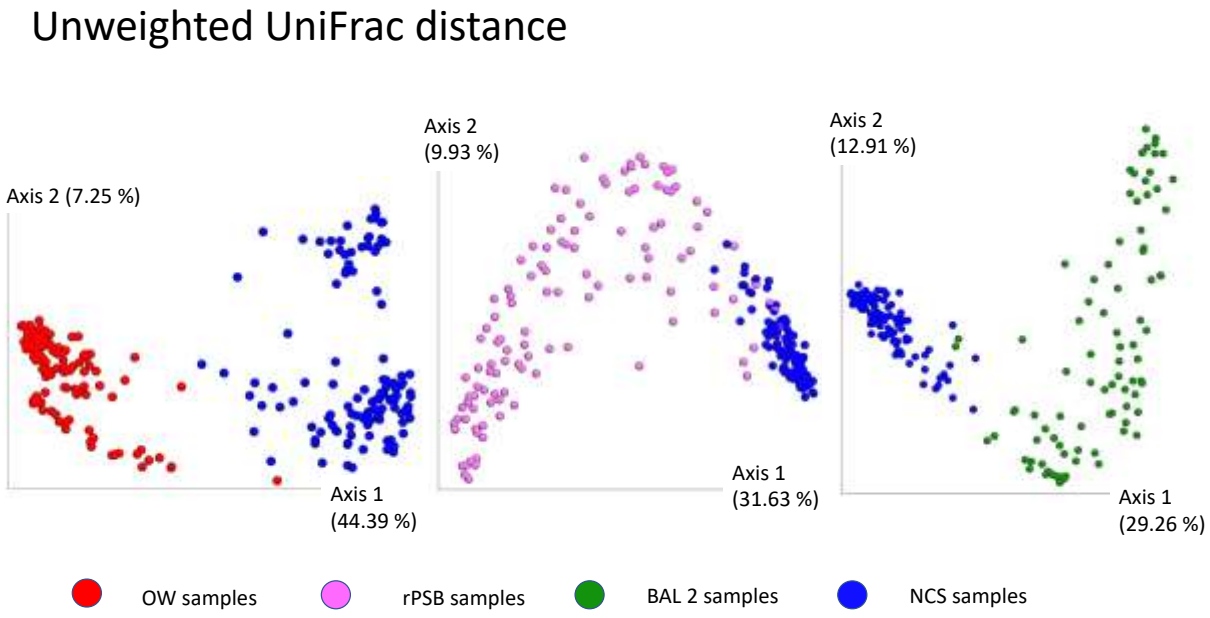
Supplementary figure 1: Bray-Curtis distances between airway samples and negative control samples. Samples in this figure are from the participants in this study, before the Decontam step and rarefied to a sampling depth of 1000 before diversity analyses. OW: Oral wash, rPSB: Right protected specimen brushes, BAL 2: Second fraction of bronchoalveolar lavage, NCS: Negative control samples.



Supplementary figure 2: Weighted UniFrac distances between airway samples and negative control samples. Samples in this figure are from the participants in this study, before the Decontam step and rarefied to a sampling depth of 1000 before diversity analyses. OW: Oral wash, rPSB: Right protected specimen brushes, BAL 2: Second fraction of bronchoalveolar lavage, NCS: Negative control samples.



Supplementary figure 3: Unweighted UniFrac distances between airway samples and negative control samples. Samples in this figure are from the participants in this study, before the Decontam step and rarefied to a sampling depth of 1000 before diversity analyses. OW: Oral wash, rPSB: Right protected specimen brushes, BAL 2: Second fraction of bronchoalveolar lavage, NCS: Negative control samples.



When collapsed to the genus level (not all could be assigned taxonomy at the genus level), the number of separate taxa was 82. These 82 taxa are listed below. Taxa are ordered according to their relative abundance among identified contaminants across all samples, in increasing order:

g__Neisseriaceae_[G-1]
g__Veillonellaceae_[G-1]
p__Bacteroidetes;__;__;__;__
g__Clostridiales_[F-3][G-1]
g__Lactococcus
g__Caulobacter
g__Bdellovibrio
g__Mitsuokella
f__Comamonadaceae;__
p__Firmicutes;__;__;__;__
g__Defluvibacter
g__Enhydrobacter
g__Comamonas
g__Dialister
c__Alphaproteobacteria;__;__;__
g__Bosea
g__Lactobacillus

g__Alloprevotella
p__Proteobacteria;__;__;__;__
f__Rhizobiaceae;__
o__Burkholderiales;__;__
g__Campylobacter
g__Achromobacter
g__Ochrobactrum
g__Paracoccus
g__Corynebacterium
g__Bifidobacterium
g__Moraxella
g__Microbacterium
g__Bacillus
f__Neisseriaceae;__
g__Saccharibacteria_(TM7)_[G-6]
g__Bacteroidales_[G-2]
g__Butyrivibrio
g__Arthrospira
g__Lachnospiraceae_[G-2]
g__Acinetobacter
g__Megasphaera
g__Cupriavidus
g__Micrococcus
g__Leptothrix
g__Peptostreptococcaceae_[XI][G-9]
g__Lysinibacillus
g__Cutibacterium
g__Gemella
g__Leptotrichia
c__Betaproteobacteria;__;__;__
g__Kingella
g__Roseomonas
g__Pedobacter
c__Gammaproteobacteria;__;__;__
g__Sphingomonas
g__Actinomyces
g__Stenotrophomonas
g__Mesorhizobium
g__Capnocytophaga
g__Selenomonas
o__Rhizobiales;__;__
f__Bradyrhizobiaceae;__
f__Enterobacteriaceae;__
g__Fusobacterium
k__Bacteria;__;__;__;__;__
g__Brevundimonas
g__Staphylococcus
g__Porphyromonas
c__Actinobacteria;__;__;__
g__Burkholderia
g__Agrobacterium
g__Haemophilus
f__Burkholderiaceae;__
g__Bergeyella
g__Segetibacter
g__Prevotella
g__Veillonella

g__Streptococcus
g__Rothia
g__Pseudomonas
g__Neisseria
g__Delftia
c__Negativicutes;__;__;__
g__Klebsiella
g__Ralstonia

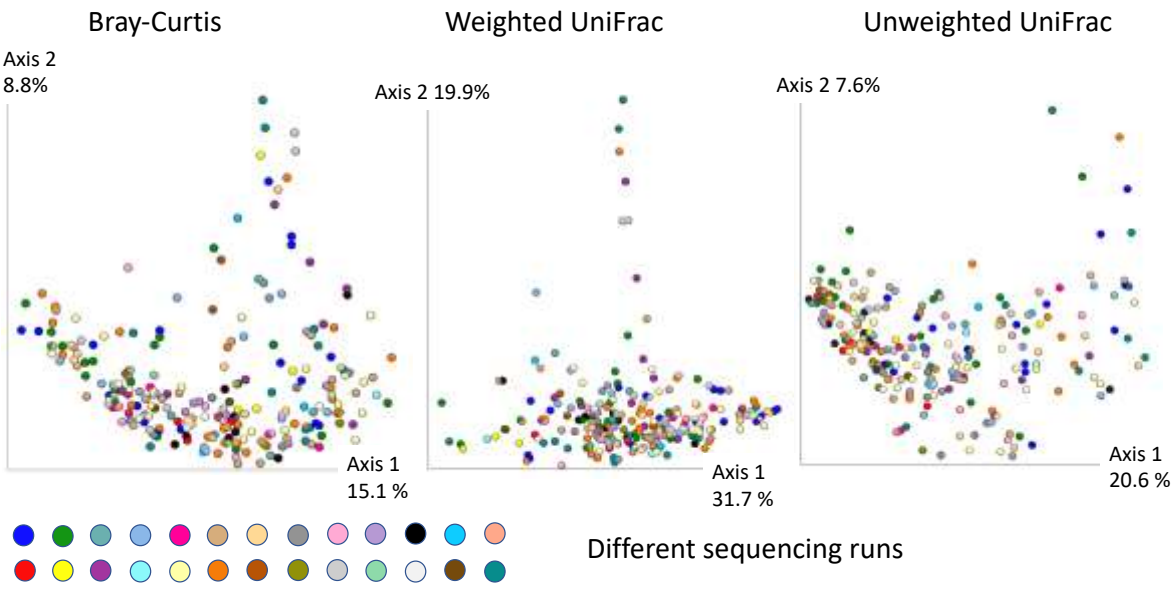
g = genus, c = class, f= family, o = order, p = phylum, k = kingdom

Examining potential batch effects

As mentioned above, data curation and quality criteria were similar for all 30 RUNs.

After merging and curation was finalized, we created a PCoA plot of all samples per patient to see if some samples clustered by sequencing RUN. We were unable to see clustering according to sequencing RUN (Supplementary figure 4).

Supplementary figure 4: PCoA plots of three different beta diversity metrices (Bray-Curtis, weighted and unweighted UniFrac distances) for all samples from the 105 participants with exacerbation follow-up after rarefaction to a sampling depth of 1000, coloured according to sequencing run.



Additional results

Differential abundance testing

Supplementary table 1. Results from 4 different tests (ANCOM [4], gneiss [5], ALDEx2 [6], MicrobiomeDDA [7]) of differences in abundance of taxa, either at the ASV or genus level, between patients who did and did not experience one or more later COPD exacerbations.

Test	**Sample type**	**Level**	**Taxonomic annotation of differentially expressed features:**	**Details**
ANCOM	BAL	ASV	No differentially abundant features	
		Genus	No differentially abundant genera	
	rPSB	ASV	No differentially abundant features	
		Genus	No differentially abundant genera	
gneiss	BAL	ASV	No significant balances	
	rPSB	ASV	No significant balances	
ALDEx2	BAL	ASV	No differentially abundant features	
		Genus	No differentially abundant genera	
	rPSB	ASV	No differentially abundant features	
		Genus	No differentially abundant genera	
MicrobiomeDDA	BAL	ASV	No differentially expressed features	
		Genus	No differentially expressed genera	
	rPSB	ASV	*Capnocytophaga gingivitis* (more abundant, less prevalent and less dispersed in group with exacerbations)	abund.LFC: 1.693, prev.change: -0.054 disp.LFC: -3.301 statistic: 20.035 P.adj: 0.011
			Prevotella pallens (less abundant, more prevalent and more dispersed in group with exacerbations)	abund.LFC: -0.601 prev.change: 0.060 disp.LFC: 28.006 statistic: 15.627 P.adj: 0.041
		Genus	No differentially abundant genera	

Abund.LFC: log2-fold change in fitted mean abundance parameter between exacerbation and no-exacerbation group. Prev.change: linear difference in prevalence between exacerbation and no-exacerbation group. Disp.LFC: log2-fold change in fitted dispersion parameter between exacerbation and no-exacerbation group. Statistic: Value of likelihood ratio test statistic. Padj: p-value adjusted for multiple comparisons according to the FDR/BH method.

Supplementary table 2. Results from 4 different tests (ANCOM [4], gneiss [5], ALDEx2 [6], MicrobiomeDDA [7]) of differences in abundance of taxa, either at the ASV or genus level, between patients who had none or one and patients who experienced two or more later COPD exacerbations.

Test	**Sample type**	**Level**	**Taxonomic annotation of differentially expressed features:**	**Details**
ANCOM	BAL	ASV	No differentially abundant features	
		Genus	No differentially abundant genera	
	rPSB	ASV	No differentially abundant features	
		Genus	No differentially abundant genera	
gneiss	BAL	ASV	No significant balances	
	rPSB	ASV	No significant balances	
ALDEx2	BAL	ASV	No differentially abundant features	
		Genus	No differentially abundant genera	
	rPSB	ASV	No differentially abundant features	
		Genus	No differentially abundant genera	
MicrobiomeDDA	BAL	ASV	*Capnocytophaga leadbetteri* (less abundant, more prevalent and less dispersed in group with two or more exacerbations)	abund.LFC: -2.462 prev.change: 0.052 disp.LFC: -2.155 statistic: 24.360 P.adj: 0.008
			Prevotella oris (less abundant, more prevalent and more dispersed in group with two or more exacerbations)	abund.LFC: -3.424 prev.change: 0.0004 disp.LFC: 1.826 statistic: 19.169 P.adj: 0.045
		Genus	*Moraxella* (less abundant, more prevalent and less dispersed in group with two or more exacerbations)	abund.LFC: 4.076 prev.change: 0.175 disp.LFC: 27.53 statistic: 18.215 P.adj: 0.026
	rPSB	ASV	No differentially expressed features	
		Genus	No differentially expressed genera	

Abund.LFC: log2-fold change in fitted mean abundance parameter between exacerbation and no-exacerbation group. Prev.change: linear difference in prevalence between exacerbation and no-exacerbation group. Disp.LFC: log2-fold change in fitted dispersion parameter between exacerbation and no-exacerbation group. Statistic: Value of likelihood ratio test statistic. Padj: p-value adjusted for multiple comparisons according to the FDR/BH method.

Supplementary table 3. Results from 4 different tests (ANCOM, gneiss, ALDEx2, MicrobiomeDDA) of differences in abundance of taxa, either at the ASV or genus level, between COPD patients who used and did not use inhaled corticosteroids (ICS) at baseline.

Test	**Sample type**	**Level**	**Taxonomic annotation of differentially expressed features:**	**Details**
ANCOM	BAL	ASV	No differentially abundant features	
		Genus	No differentially abundant genera	
	rPSB	ASV	No differentially abundant features	
		Genus	No differentially abundant genera	
gneiss	BAL	ASV	No significant balances	
	rPSB	ASV	No significant balances	
ALDEx2	BAL	ASV	No differentially abundant features	
		Genus	No differentially abundant genera	
	rPSB	ASV	No differentially abundant features	
		Genus	No differentially abundant genera	
MicrobiomeDDA	BAL	ASV	No differentially expressed features	
		Genus	No differentially expressed genera	
	rPSB	ASV	No differentially expressed features	
		Genus	No differentially expressed genera	

Abund.LFC: log2-fold change in fitted mean abundance parameter between ICS usage and no ICS usage group. Prev.change: linear difference in prevalence between ICS usage and no ICS usage group. Disp.LFC: log2-fold change in fitted dispersion parameter between ICS usage and no ICS usage group.
Statistic: Value of likelihood ratio test statistic. Padj: p-value adjusted for multiple comparisons according to the FDR/BH method.

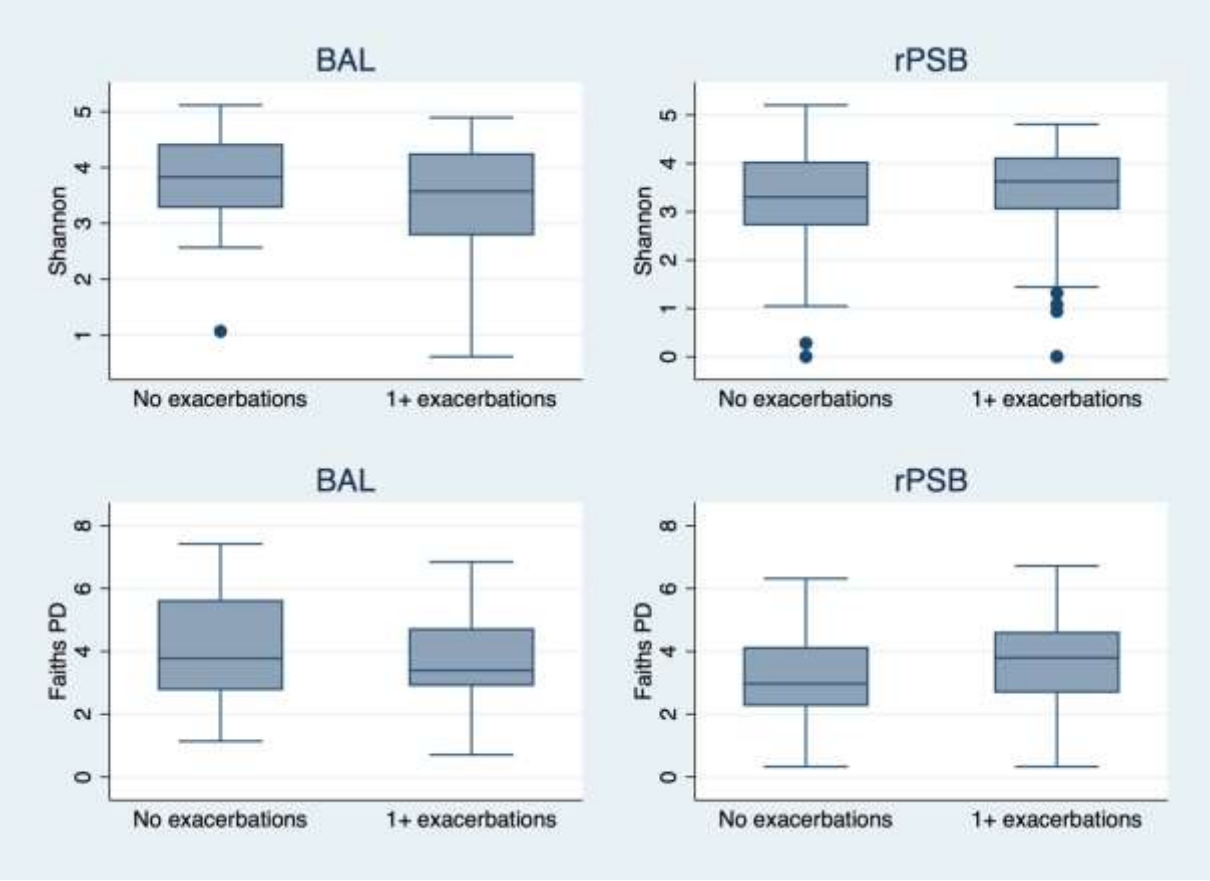
Supplementary table 4. Results from 4 different tests (ANCOM [4], gneiss [5], ALDEx2 [6], MicrobiomeDDA [7]) of differences in abundance of taxa, either at the ASV or genus level, between patients who did and did not experience one or more COPD exacerbations in the 12 months prior to the bronchoscopy.

Test	**Sample type**	**Level**	**Taxonomic annotation of differentially expressed features:**	**Details**
ANCOM	BAL	ASV	No differentially abundant features	
		Genus	No differentially abundant genera	
	rPSB	ASV	No differentially abundant features	
		Genus	No differentially abundant genera	
gneiss	BAL	ASV	No significant balances	
	rPSB	ASV	No significant balances	
ALDEx2	BAL	ASV	No differentially abundant features	
		Genus	No differentially abundant genera	
	rPSB	ASV	No differentially abundant features	
		Genus	No differentially abundant genera	
MicrobiomeDDA	BAL	ASV	*Actinomyces graevenitzii* (less abundant, more prevalent and less dispersed in patients having experienced exacerbations)	abund.LFC: -1.513 prev.change: 0.028 disp.LFC: -0.618 statistic: 23.660 P.adj: 0.011
		Genus	No differentially expressed genera	
	rPSB	ASV	No differentially expressed features	
		Genus	No differentially expressed genera	

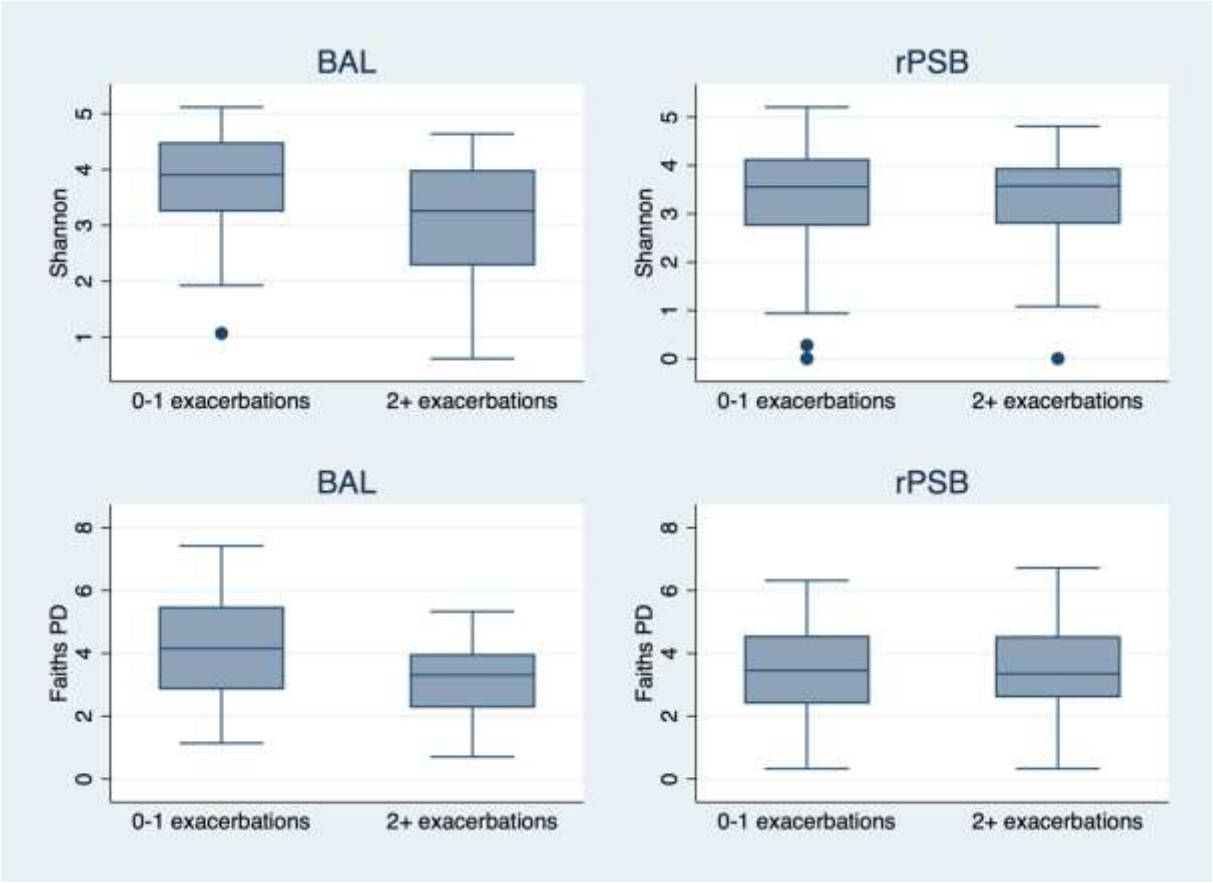
Abund.LFC: log2-fold change in fitted mean abundance parameter between exacerbation and no-exacerbation group. Prev.change: linear difference in prevalence between exacerbation and no-exacerbation group. Disp.LFC: log2-fold change in fitted dispersion parameter between exacerbation and no-exacerbation group. Statistic: Value of likelihood ratio test statistic. Padj: p-value adjusted for multiple comparisons according to the FDR/BH method.

Diversity analyses

Supplementary figure 5: Box plots of Faith phylogenetic alpha diversity (Faith PD) and Shannon Index alpha diversity in patients without compared with one or more later exacerbations in 60 bronchoalveolar lavage (BAL2) and 73 protected specimen brush samples (rPSB). Differences in diversity were tested with Kruskal-Wallis. There were no statistically significant associations.



Supplementary figure 6: Box plots of Faith phylogenetic alpha diversity (Faith PD) and Shannon Index alpha diversity in patients zero or one compared with two or more later exacerbations in 60 bronchoalveolar lavage (BAL2) and 73 protected specimen brush samples (rPSB). Differences in diversity were tested with Kruskal-Wallis. There were no statistically significant associations.



Supplementary table 5. Permutational multivariate analysis of variance (PERMANOVA [8]) of the beta-diversity (weighted UniFrac) by exacerbation category (zero versus one or more) without and with adjustment for age, sex, FEV_1 and use of inhaled steroids, in bronchoalveolar lavage samples and right protected specimen brush samples. Analysed with the vegan package in R.

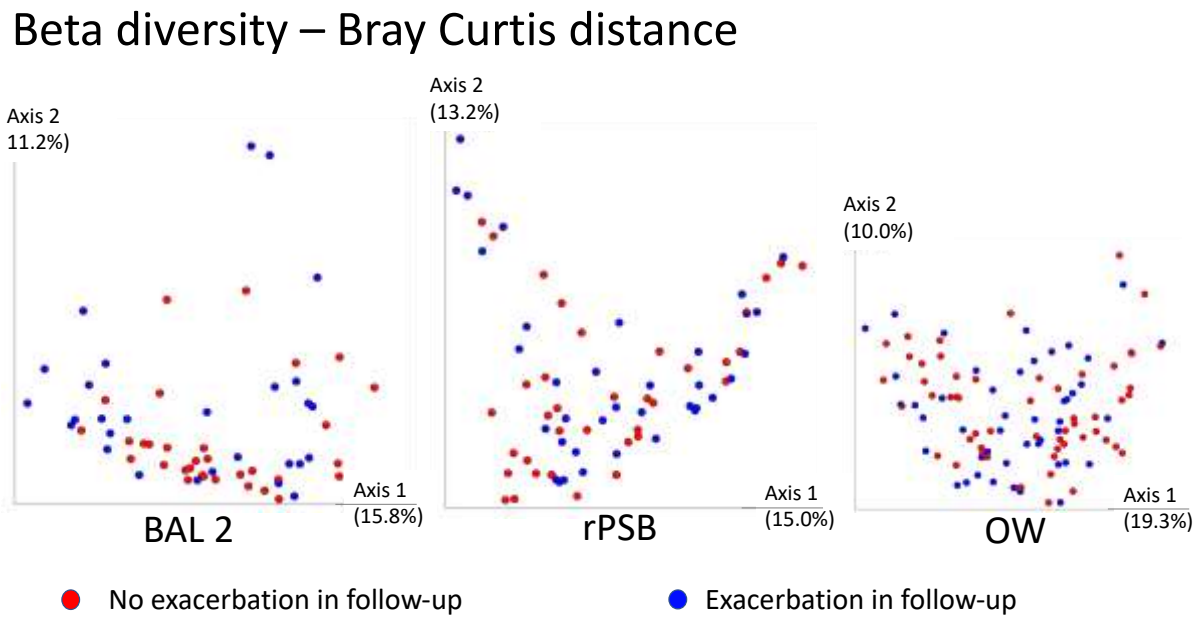
Bronchoalveolar Lavage						
	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Exacerbation category	1	0.04811	0.048107	1.173	0.01982	0.287
Residuals	58	2.37876	0.041013		0.98018	
	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Exacerbation category	1	0.04811	0.048107	1.15092	0.01982	0.324
Age	1	0.01602	0.016016	0.38316	0.00660	0.898
Sex	1	0.04037	0.040369	0.96578	0.01663	0.416
FEV1 in percent predicted	1	0.02836	0.028357	0.67840	0.01168	0.618
Inhaled steroid use	1	0.03686	0.036863	0.88190	0.01519	0.507
Residuals	54	2.25715	0.041799		0.93007	
right Protected Specimen Brush						
	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Exacerbation category	1	0.0518	0.051754	0.88513	0.01231	0.462
Residuals	71	4.1514	0.058470		0.98769	
	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Exacerbation category	1	0.0518 0.051754	0.9192	0.01231	0.414	
Age	1	0.1884	0.188388	3.3458	0.04482	
Sex	1	0.0395	0.039534	0.7021	0.00941	0.589
FEV1 in percent predicted	1	0.0834 0.083391	1.4810	0.01984	0.188	
Inhaled steroid use	1	0.0676	0.067642	1.2014	0.01609	0.294
Residuals	67	3.7724	0.056305		0.89753	

Supplementary table 6. Permutational multivariate analysis of variance (PERMANOVA [8]) of the beta-diversity (weighted UniFrac) by exacerbation category (zero or one versus two or more) without and with adjustment for age, sex, FEV_1 and use of inhaled steroids, in bronchoalveolar lavage samples and right protected specimen brush samples. Analysed with the vegan package in R.

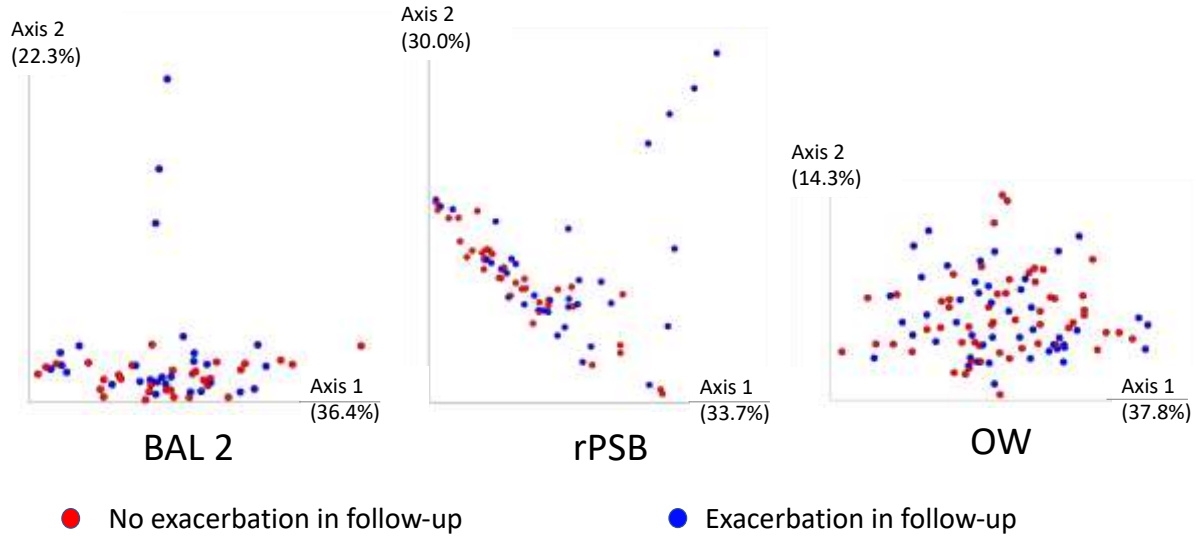
*Bronchoalveolar Lavage**						
	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Exacerbation category	1	0.1078	0.107837	2.64261	0.04358	0.031
Residuals	58	2.3668	0.040807		0.95642	
	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Exacerbation category	1	0.1078	0.107837	2.59727	0.04358	0.031
Age	1	0.0622	0.062161	1.49716	0.02512	0.172
Sex	1	0.0199	0.019898	0.47925	0.00804	0.822
FEV1 in percent predicted	1	0.0223	0.022272	0.53643	0.00900	0.763
Inhaled steroid use	1	0.0204	0.020431	0.49208	0.00826	0.800
Residuals	54	2.2420	0.041519		0.90601	
right Protected Specimen Brush						
	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Exacerbation category	1	0.2154	0.215415	3.83368	0.05123	0.006
Residuals	71	3.9895	0.056190		0.94877	
	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Exacerbation category	1	0.2154	0.215415	3.85941	0.05123	0.007
Age	1	0.0886	0.088620	1.58772	0.02108	0.169
Sex	1	0.0263	0.026253	0.47036	0.00624	0.788
FEV1 in percent predicted	1	0.0733	0.073289	1.31306	0.01743	0.248
Inhaled steroid use	1	0.0617	0.061702	1.10546	0.01467	0.329
Residuals	67	3.7396	0.055815		0.88935	

* For Broncheoalveolar lavage the dispersion was significant, and the results cannot be trusted, the results are shown only for consistency with supplementary table 5.

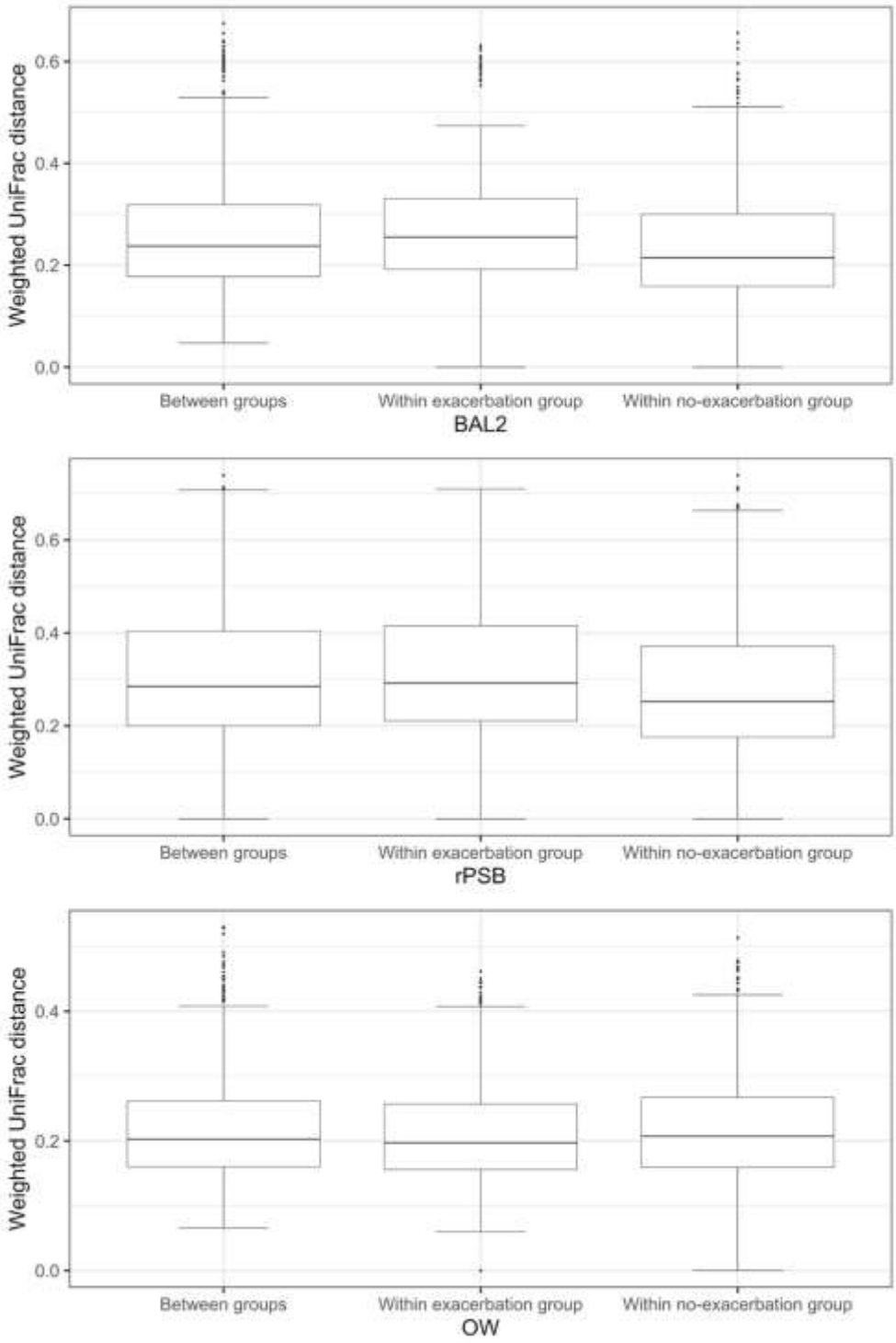
Supplementary figure 7: Principle Coordinate plots of beta diversity, measured by Bray Curtis distance in second fraction of bronchoalveolar lavage (BAL2), protected specimen brush (rPSB) and oral wash (OW) samples. Each sample is coloured according to exacerbation status (zero versus one or more) during follow-up.



Supplementary figure 8: Principle Coordinate plots of beta diversity, measured by weighted UniFrac distance in second fraction of bronchoalveolar lavage (BAL2), protected specimen brush (rPSB) and oral wash (OW) samples. Each sample is coloured according to exacerbation status (zero versus one or more) during follow-up.



Supplementary figure 9: Beta diversity, measured by weighted UniFrac distance. Box plots of distances between groups with and without later exacerbations for each sample type; oral wash (OW), right protected specimen brushes (rPSB) and second fraction of bronchoalveolar lavage (BAL2). Differences between all groups were tested with the ADONIS permutation-based test. There were no statistically significant differences.



References

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

2. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014; 30, 2114-2120.

3. 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, 226.

4. Mandal, S 2015 Analysis of composition of microbiomes: a novel method for studying microbial composition. Microb Ecol Health Dis 2015; 26, 27663.

5. Morton JT, Sanders J, Quinn RA et al. Balance Trees Reveal Microbial Niche Differentiation. mSystems 2017; 2,

6. Fernandes, AD 2014 Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. 1805 21455 0

7. Chen J, King E, Deek R et al. An omnibus test for differential distribution analysis of microbiome sequencing data. Bioinformatics 2018; 34, 643-651.

8. Anderson MJ. A new method for non-parametric multivariate analysis of variance. Austral ecology 2001; 26, 32-46.