**Sputum microbiome temporal variability and dysbiosis in chronic obstructive pulmonary disease exacerbations: an analysis of the COPDMAP study**

Zhang Wang, Richa Singh, Bruce E. Miller, Ruth Tal-Singer, Stephanie Van Horn, Lynn Tomsho, Alexander Mackay, James P. Allinson, Adam J. Webb, Anthony J. Brookes, Leena M. George, Bethan Barker, Umme Kolsum, Louise E Donnelly, Kylie Belchamber, Peter J. Barnes, Dave Singh, Christopher E. Brightling, Gavin C. Donaldson, Jadwiga A. Wedzicha, James R. Brown on behalf of COPDMAP

**SUPPLEMENTARY MATERIAL**

**Material and Methods**

**Study subjects and sample collection**

COPDMAP was conducted in accordance with the Declaration of Helsinki [1](#_ENREF_1) and Good Clinical Practice [2](#_ENREF_2), and was approved by the Imperial College London, University of Leicester and University of Manchester Research Ethics Committee. All participants provided written informed consent. Subjects with a physician diagnosis of COPD were recruited from three clinical centres at Imperial College London, University of Leicester and University Hospital of South Manchester, and through local advertising to enter studies investigating biomarkers in COPD as previously described [3](#_ENREF_3)[4](#_ENREF_4).The Imperial samples were collected at Royal Free Hospital of University College London at the time of the study. Subjects with asthma, or significant respiratory disease other than COPD, or the inability to produce sputum after sputum induction were excluded from the study. Sputum samples from COPD subjects were collected at multiple longitudinal visits including both baseline (defined as no evidence of symptom-defined exacerbations in the preceding four weeks and the subsequent two weeks post-clinic visit) and exacerbations (defined according to Anthonisen criteria [5](#_ENREF_5) and/or healthcare utilization [6](#_ENREF_6)). All exacerbation sputum samples were collected prior to the institution of any exacerbation treatment. Demographic, baseline and longitudinal clinical data were recorded for samples. A number of 15 subjects were missing any demographic or clinical data and were excluded for biostatistical analysis.

**16S rRNA sequencing**

For quality control purposes, all DNA extractions, sequencing and data analyses occurred in a single, centralized lab at the GSK R&D facility in Collegeville, Pennsylvania, USA. Frozen sputum samples homogenized in sterile 1x sputasol (0.1% DTT) was thawed completely on bench. Bacterial genomic DNA was extracted from sputum samples using the Qiagen DNA Mini kit (Qiagen, CA, USA) as per manufacture protocol. The V4 hypervariable region of the 16S rRNA gene was PCR amplified using specific primers (515F: 5’ GTGCCAGCMGCCGCGGTAA3’, 806R: 5’GGACTACHVGGGTWTCTAAT3’), including Illumina sequencing adapters [7](#_ENREF_7). The reverse amplification primer contained a 12 bp error-correcting Golay barcode sequence allowing for pooling of multiple samples in the same flowcell [8](#_ENREF_8). Negative controls for extraction (no sputum material) and PCR amplification (no template, Qiagen Elution Buffer only) were included in each experiment. The extraction negative control for each experiment was subsequently sequenced to identify any potential contaminating bacterial species.

The amplification mix (25 μl) contained 4 μl sputum DNA, 2 μl (0.2 µM) each of forward and reverse primers (Integrated DNA Technologies, Coralville, IA), 12.5 μl of 2x KAPA HiFi HotStart Ready Mix (KK2602, Kapa biosystems, Boston MA), and 4.5 μl RNase free water. PCR amplification was performed on an ABI 9700 thermocycler using the following cycling protocol: initial denaturation at 95°C for 3 min, followed by 35 cycles of 98°C for 20 sec, 66°C for 15 sec, and 72°C for 15 sec, with a final hold of 72°C for 1 min. Aliquots of reaction mixture (3 µl each) were analysed by 2% agarose gel (2% Egel, Invitrogen) with samples containing a band of approximately 385 bp considered ‘PCR positive’. Samples with no visible amplified product were considered ‘PCR negative’. Unincorporated nucleotides and remaining primers were removed using Agencourt AMPure XP-PCR clean up (A63882, Beckman Coulter, Pasadena, CA), according to the manufacturer’s protocol. The DNA concentration of the eluted product was quantified using the KAPA Library Quantification Kit for Ilumina platform (KK4835, Kapa biosystems, Boston MA). PCR products were normalized to 10 nM and quantified again using the KAPA Library Quantification kit and pooled into equimolar 4 nM pools.

The amplified PCR products were sequenced in five runs on an Illumina MiSeq sequencer (Illumina, San Diego, CA). Following cluster formation on the MiSeq instrument, the amplicons were sequenced using primers complimentary to the V4 region and designed for paired-ends sequencing. A third sequencing primer was used for reading the barcodes. To check for proper cluster density and sample normalization, a MiSeq single-end 26 bp+12 bp index sequencing run was performed using the MiSeq instrument. The pool was mixed with a PhiX library (Illumina, San Diego CA) at a ratio of 1:9 in order to increase the entropy of the library. A final MiSeq 2x 150 bp+12 bp index sequencing run was performed on the pooled samples.

Although negative reagent controls were performed for all DNA isolation, extraction and PCR amplification step, we performed further analyses to ensure that potential contamination risks were minimized. We compared our results against the 92 contaminant genera detected in sequenced negative ‘blank’ controls by Salter et al. [9](#_ENREF_9). We failed to detect 42 out of the 92 contaminant genera in our dataset (Table S6). Of the remaining genera that were found in our data, none had an average relative abundance greater than 0.002, or had a relative abundance greater than 0.1 in any particular sample, except for *Pseudomonas* and *Streptococcus* which are known lung pathogens (Table S6).

**16S rRNA sequence analysis and OTU classification**

First, all reads mapping to PhiX reference sequence (GenBank: NC\_001422.1) using bowtie v1.0.1 [10](#_ENREF_10) were removed from the analysis. Remaining paired-end reads were merged using pear v0.9.5-64 [11](#_ENREF_11), discarding all reads containing ambiguous bases (option ‘-u 0’). A paired-end read was discarded if one of the following conditions was met: overlap < 10 bp, assembly length < 50 bp or p-value of alignment > 0.01. Sequencing reads were processed using QIIME pipeline version 1.9 [12](#_ENREF_12). Eukaryotic, mitochondria and chloroplast sequences were filtered by BLASTN against the SILVA database [13](#_ENREF_13). Chimeric reads were identified using UCHIME using both *de novo* and reference based methods with default parameters [14](#_ENREF_14). A total of 68,643,967 reads were generated, and 55,786,582 reads were retained after filtering processes. The remaining reads were subject to a 97% identity cutoff close reference OTU picking using the UCLUST method [15](#_ENREF_15) against the August 2013 edition of the Greengenes 16S rRNA database (v13\_8) [16](#_ENREF_16). OTU clustering was performed on each run separately and the resulting OTU tables were merged afterwards. OTUs that contain a single read (singleton OTUs) were excluded to remove potential sequencing artefacts. All 716 samples were rarefied to 46,056 reads which is the minimum number of aligned reads across all samples. The rarefied OTU table was used for assessing alpha, beta diversity and for subsequent statistical analyses. Alpha and beta diversity was also calculated from functional prediction of microbial gene families and pathways using the software PICRUSt [17](#_ENREF_17).

**Statistical analysis**

PLS-DA between the sputum microbiome and exacerbation phenotypes

Exacerbation phenotypes were defined using slightly modified microbiological and clinical criteria as established previously [4](#_ENREF_4)[18](#_ENREF_18). In particular, the total bacterial load was estimated by the qPCR copy number of *Haemophilus influenzae* normalized by the proportion of the species in the sputum microbiome, and a bacterial exacerbation was defined as a total bacterial load >= 107 cells[4](#_ENREF_4)[18](#_ENREF_18). A virus exacerbation was defined as a positive sputum viral detection by PCR. An eosinophilic exacerbation was defined as the presence of more than 3% non-squamous cells in sputum. Samples with multiple criteria satisfied were classified as the corresponding combination of phenotypes. The remaining samples associated with limited changes in the inflammatory profile were classified as ‘pauci-inflammatory’. The phenotype could not be defined for 146 exacerbation samples due to missing data. A total of 25 clinical variables were included in the analysis. PLS-DA was performed using SIMCA-P (Umetrics, Stockholm, Sweden) [19](#_ENREF_19) as previously described [18](#_ENREF_18). For subjects with multiple exacerbations, only the initial exacerbation sample was included in the analysis to meet the independence assumption for PLS-DA.

Dysbiosis and temporal variability of the sputum microbiome

A total of 64 subjects that had at least two baseline and one exacerbation samples were included in the dysbiosis analysis. We measured the dysbiosis of exacerbations relative to baseline samples of the same subject using PC1 of weighted UniFrac distance. Assuming a normal distribution of the baseline PC1s within each subject, the mean and standard deviation of the baseline PC1s were calculated. And a Z-score was calculated for each exacerbation as:

where PC1eis the exacerbation PC1, and μband σb are the mean and standard deviation of all baseline PC1s of the same subject, respectively. An absolute Z-score greater than 2 was used as cutoff for dysbiosis, which is equivalent to a probability of 0.05 in observing the exacerbation PC1 from the subject under the distribution of its baseline PC1s.

A total of 126 subjects that had at least three samples were included in the analysis of the microbial temporal variability. We adopted the metrics from Flores et al. [20](#_ENREF_20) to assess temporal variability of microbial alpha and beta diversity. For the variability of alpha diversity, we calculated the coefficient of variation (CV) as standard deviation normalized by mean for the Shannon of all samples within each subject. For the variability of beta diversity, we calculated the median of the pairwise weighted UniFrac distances of all samples within each subject. Higher values of these measurements represent a more variable microbial community. Subjects were then divided into quartiles based on the CV of Shannon and the median of pairwise weighted UniFrac distances where the first quartile was defined as ‘low’, the second and third quartiles as ‘medium’ and the fourth quartile as ‘high’ for each measurement of temporal variability.

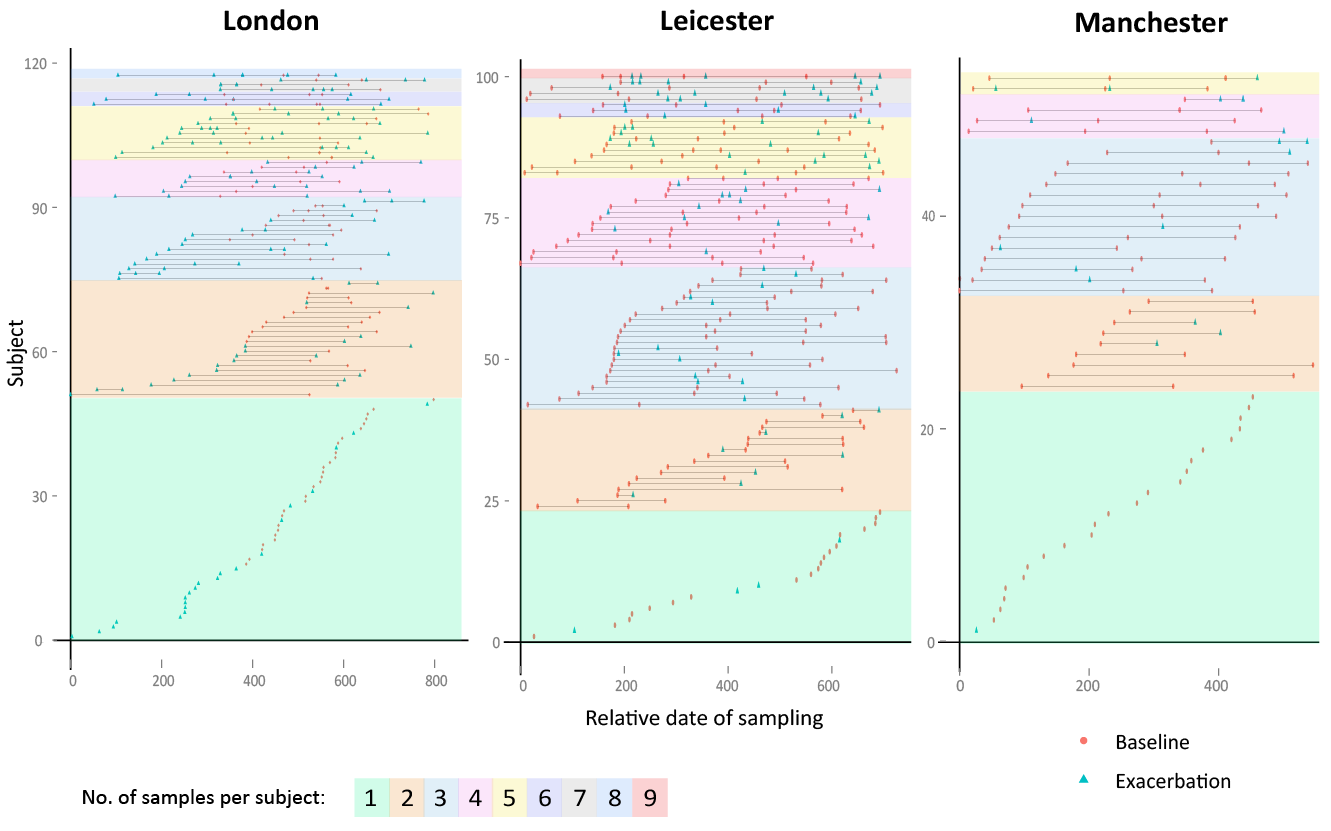
Clinical predictors of the sputum microbiome diversity and temporal variability

Multivariate models were constructed to assess the significant association between patient demographic and clinical variables to alpha and beta diversity and their temporal variability. Both the microbiome and clinical datasets were pre-processed as described previously [18](#_ENREF_18). To identify clinical predictors of temporal variability of alpha and beta diversity, a general linear model (GLM) was constructed between demographic/baseline clinical variables and each measurement respectively, for all subjects and subjects within each centre. A set of 14 demographic and baseline clinical variables were included for each subject. A total of 113 subjects with complete measurements of these data were included in the analysis. The 113 subjects were not significantly different from the remaining subjects in terms of study centre distribution, major demographic and baseline clinical variables (age, FEV1, FVC, CAT score, etc) and microbiome profiles. As each subject was associated a single measurement, we were able to meet the independence assumption. The model was optimized in terms of Akaike information criterion (AIC) through backward elimination of non-significant effects in a stepwise algorithm using the “step” function in the R stats package [21](#_ENREF_21).

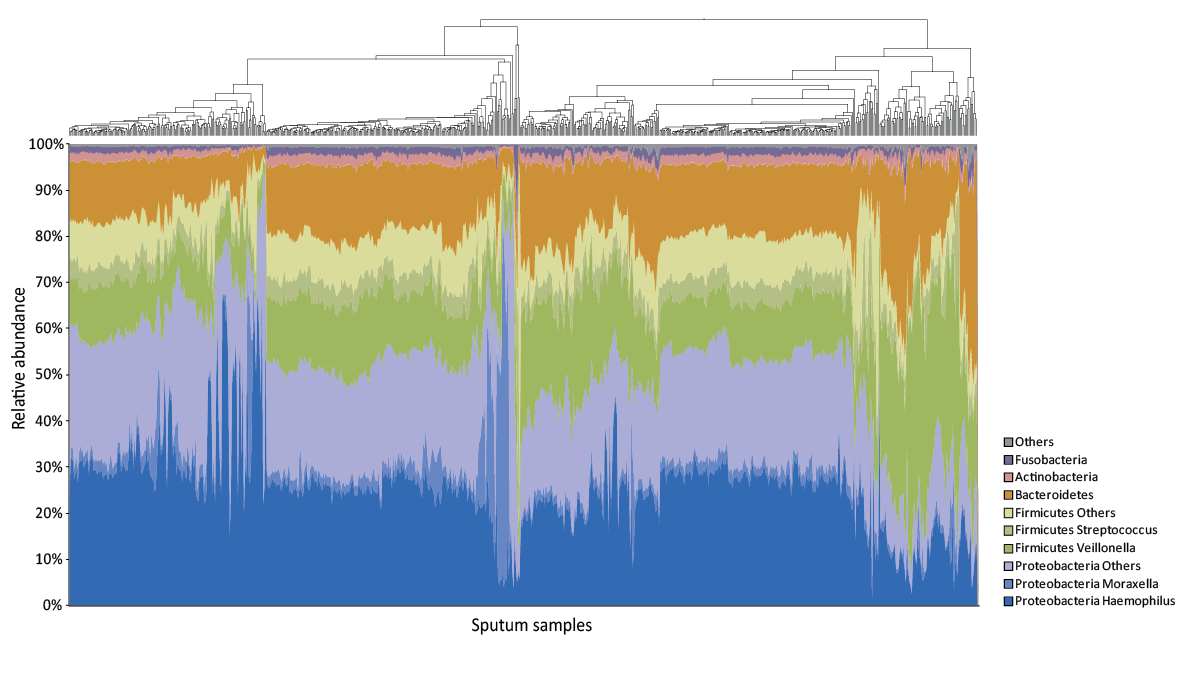
To identify clinical predictors of microbial alpha diversity, a general linear mixed model (GLMM) was constructed between clinical variables and Shannon for all samples as well as samples within each centre. Subject ID of each sample was used as the random factor to adjust for dependency of repeated measurements on the same subject. A set of 22 demographic, baseline and longitudinal clinical variables were included for all samples. A total of 391 samples with complete measurements of these data were included in the analysis. The 391 samples were not significantly different from the remaining samples in terms of study centre distribution, major clinical variables (FEV1, FVC, CAT score, etc) and microbiome profiles. The model was optimized in terms of Akaike information criterion (AIC) through backward elimination of non-significant effects in a stepwise algorithm using the “step” function in the R lmerTest package [22](#_ENREF_22). A GLMM was also constructed between clinical variables and Shannon for the predicted metagenome from microbial taxa.

To identify clinical predictors of beta diversity, we carried out a canonical correspondence analysis (CCA) using the R Vegan package [23](#_ENREF_23). To meet the independence assumption, only the initial baseline samples from each subject were included in the analysis. The same set of 22 demographic, baseline and longitudinal clinical variables were included in the analysis. A total of 129 initial baseline samples with complete measurements of these data were included in the analysis. CCA was performed on clinical variables and the relative abundance of taxa at each of the phylum, genus and OTU level, for all samples as well as samples within each centre. At each level, taxa present in at least 10% of samples were included in the analysis. The model was optimized in terms of Akaike information criterion (AIC) in a stepwise algorithm using the “step” function in the R stats package [21](#_ENREF_21). The statistical significance of each clinical variable was obtained by permutation test. CCA analysis was also performed between clinical variables and L2, L3 functional categories for the predicted metagenome from microbial taxa.

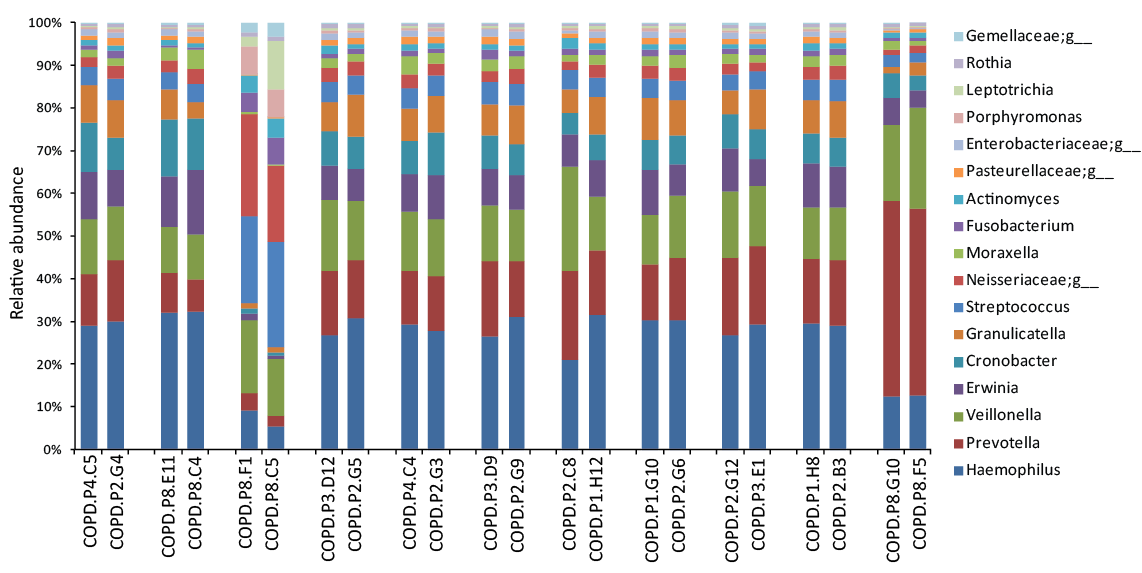
**Supplementary Results**

****

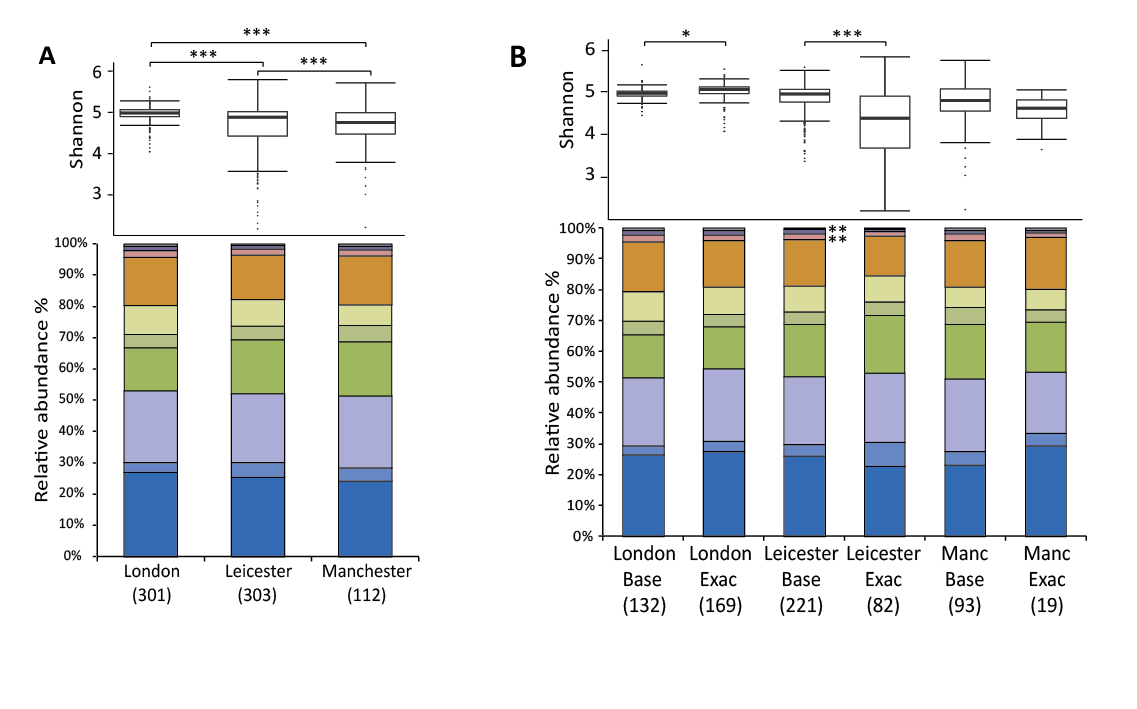
**Figure S1**. Time points of sample collection from subjects at London, Leicester and Manchester. Each set of connected dots represents samples collected from the same subject at different visits. The X axis represents days of sample collection date relative to the earliest collection date of all samples at each centre. Subjects were firstly grouped by the number of samples and then ordered by the initial sample collection date.

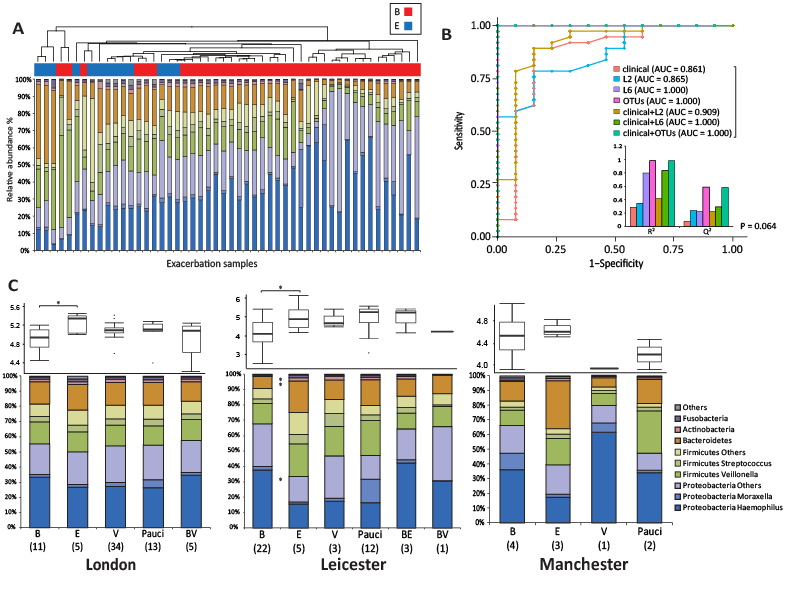


**Figure S2.** Overview of the sputum microbiome across all 716 samples. Each column represents one sample. Y-axis represents relative abundance of major phyla and genera. Samples were clustered by UPGMA clustering based on the weighted UniFrac distances.

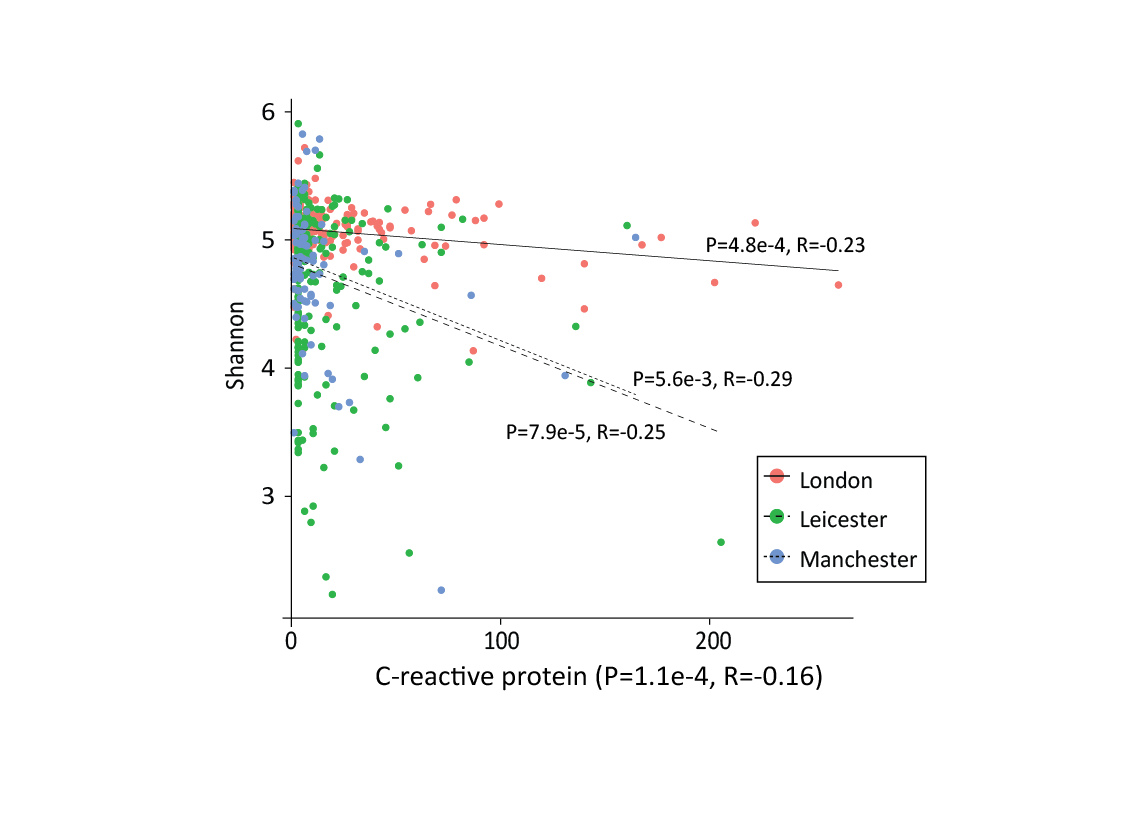


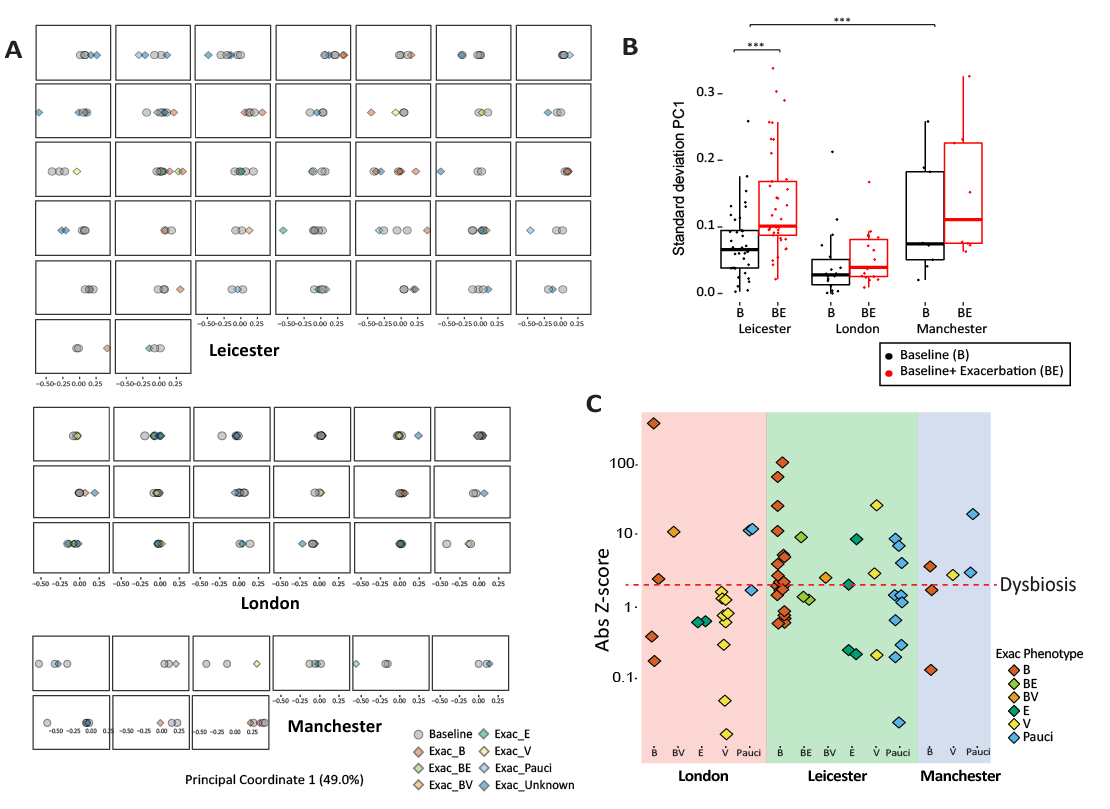
**Figure S3.** Highly similar microbial profiles between duplicate samples collected from the same individual at the same date. Duplicate samples were grouped by subject. Genus level microbial composition was shown for each sample. Genera with average relative abundance greater than 0.005 were included.

**Figure S4.** Alpha diversity (Shannon) and composition of major phyla and genera in samples A) at London, Leicester and Manchester, and B) baseline and exacerbations within each centre.

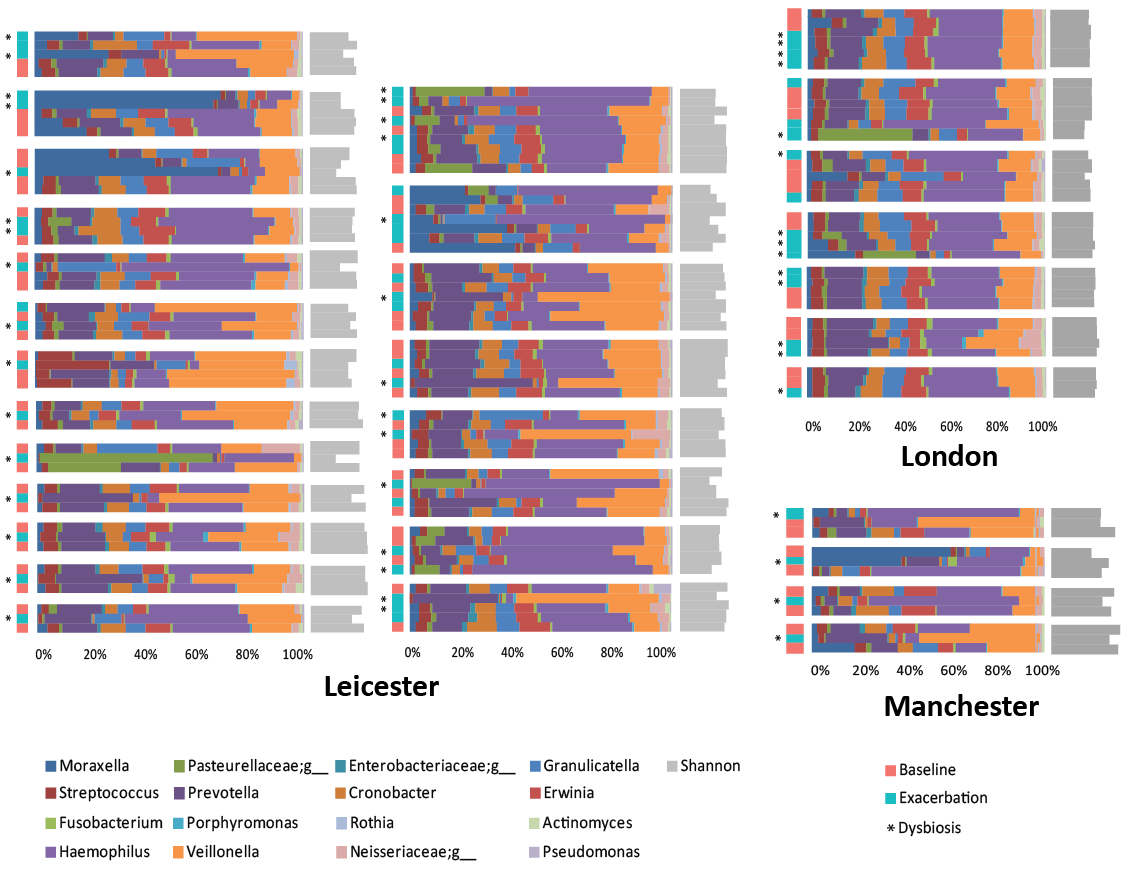


**Figure S5.** Microbiome distinguished bacterial and eosinophilic exacerbations. A) Unweighted pair group method with arithmetic mean clustering showing distinct clustering of samples with bacterial and eosinophilic exacerbations. B) PLS-DA classification of bacterial and eosinophilic exacerbations using clinical, microbiome and their combined variables at phylum (L2), genus (L6) and OTU levels. The models were evaluated in terms of area under Receiver Operating Characteristic curve (AUC), R2 and Q2 scores. C) Alpha diversity (Shannon) and composition of major phyla and genera in exacerbation samples of different exacerbation phenotypes at London, Leicester and Manchester. The number of samples is indicated in the parenthesis under each subgroup in the bar chart. Error bars are within 1.5 interquartile range of the upper and lower quartiles. B: bacterial; V: viral; E: eosinophilic; BE: bacterial and eosinophilic; BV: bacterial and viral; and Pauci: pauci-inflammatory. \*\*\* adj. *P*<0.001; \*\* adj. *P*<0.01; \* adj. *P*<0.05.

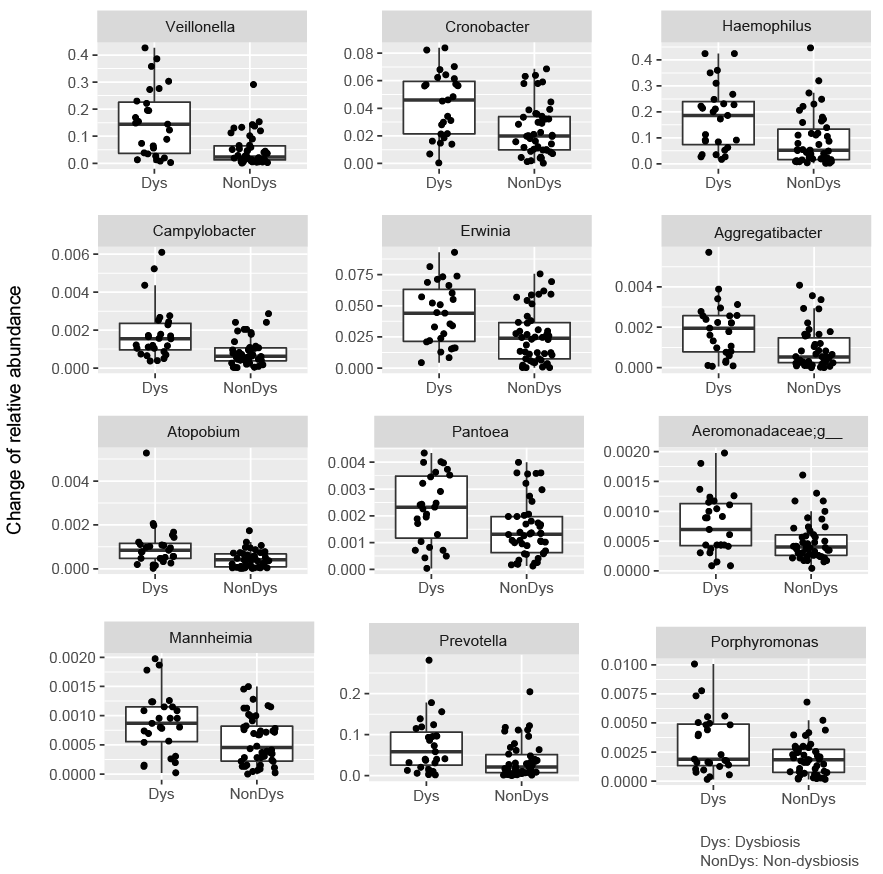
**Figure S6.** Significant negative correlation between CRP and alpha diversity (Shannon). Each dot represents a sample coloured by centre.



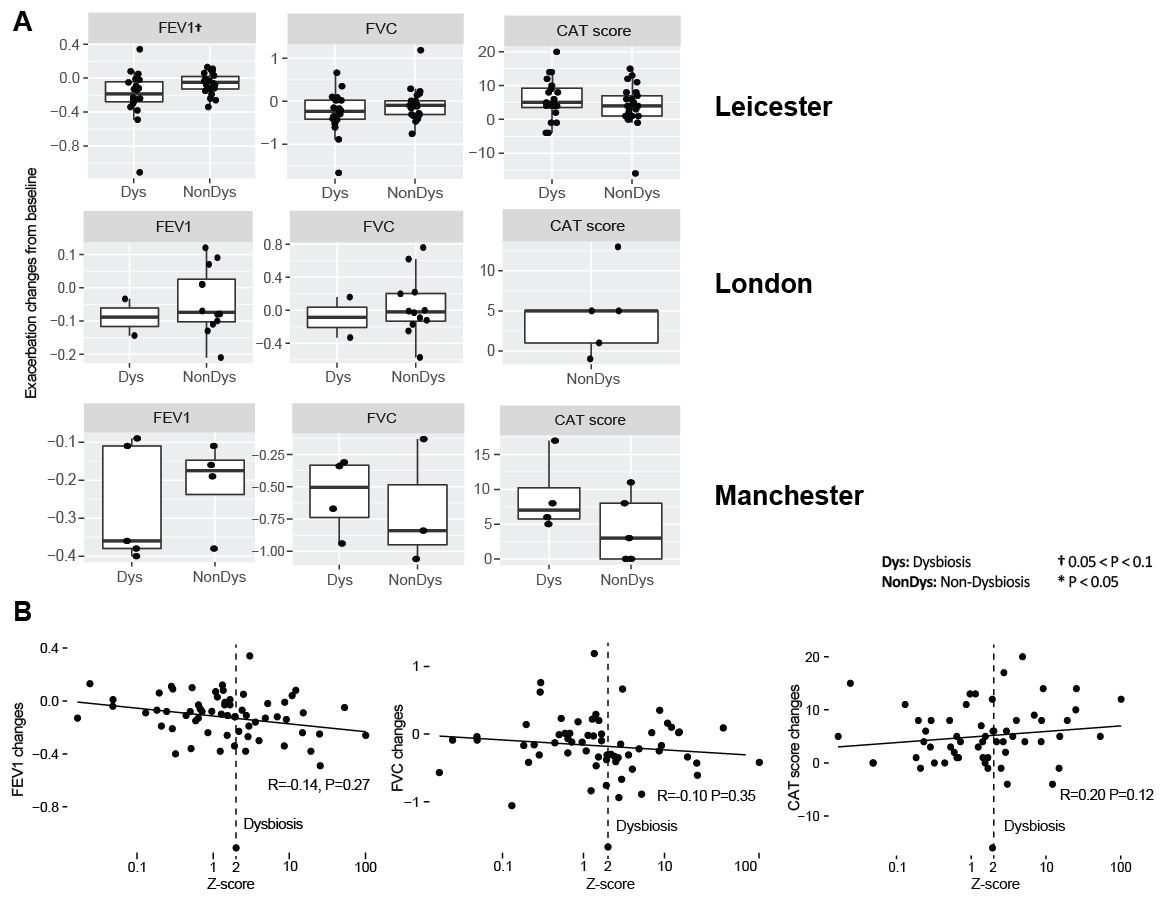
**Figure S7.** Dysbiosis of the sputum microbiome. A) Scatter plot of the first Principal Coordinate (PC1) of all samples within each subject at each centre. Only subjects with at least two baseline and one exacerbation samples were included. Exacerbation samples were coloured by different exacerbation phenotypes. Most deviations of PC1 occurred at exacerbations. B) Box-whisker plots showing significantly increased standard deviation of PC1 combining baseline and exacerbation samples within each subject compared to baseline samples only. C). Scatter plot of absolute Z-score for exacerbations at each centre. Exacerbations were coloured by different exacerbation phenotypes. Error bars are within 1.5 interquartile range of the upper and lower quartiles. B: bacterial; V: viral; E: eosinophilic; BE: bacterial and eosinophilic; BV: bacterial and viral; and Pauci: pauci-inflammatory. \*\*\* adj. *P*<0.001; \*\* adj. *P*<0.01; \* adj. *P*<0.05.



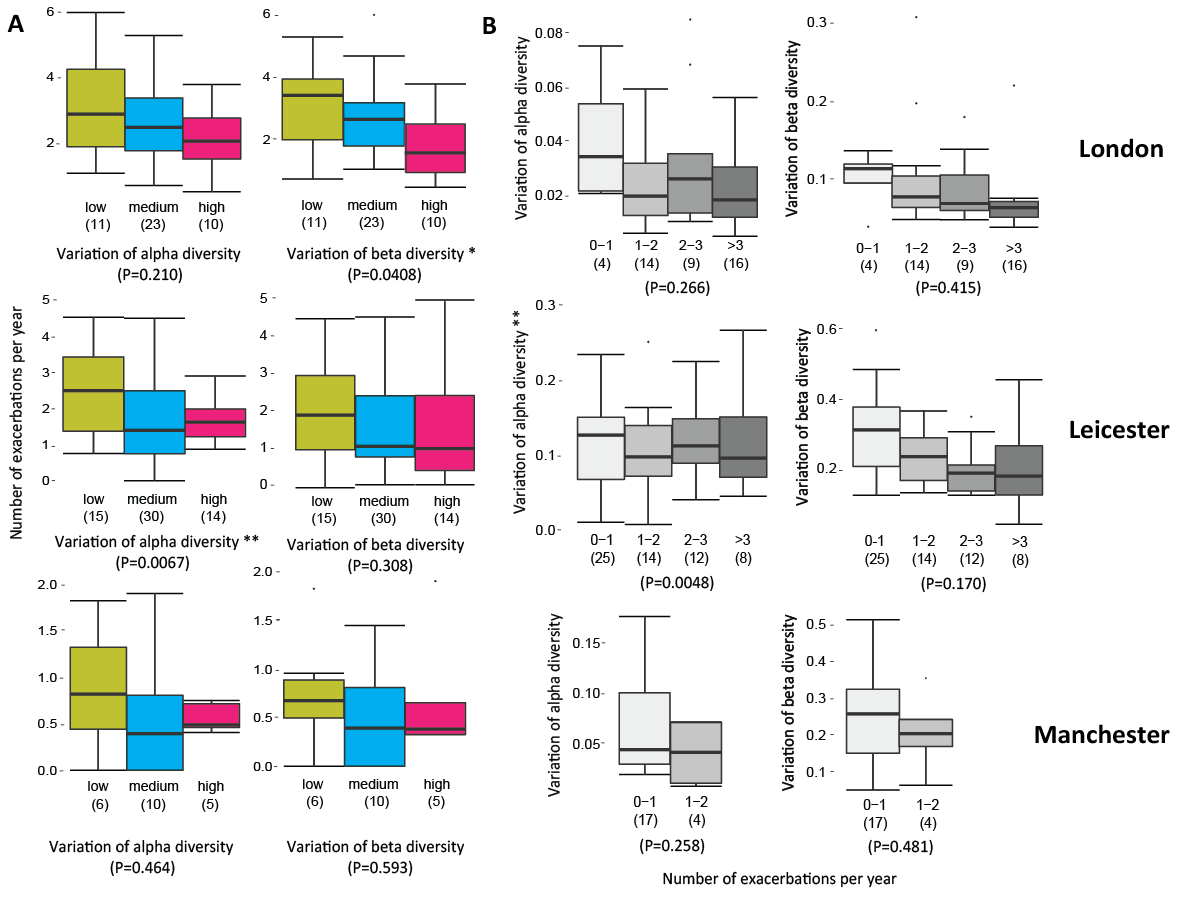
**Figure S8.** Temporal dynamics of the sputum microbiome in subjects with dysbiosis exacerbations at London, Leicester and Manchester. Each horizontal bar represents alpha diversity (Shannon) and genus level microbial composition of one sample. Samples were grouped by subject and ordered by collection dates from bottom to top. Baseline or exacerbation samples are indicated at the left of the horizontal bars. Exacerbations with dysbiosis are highlighted in asterisks.



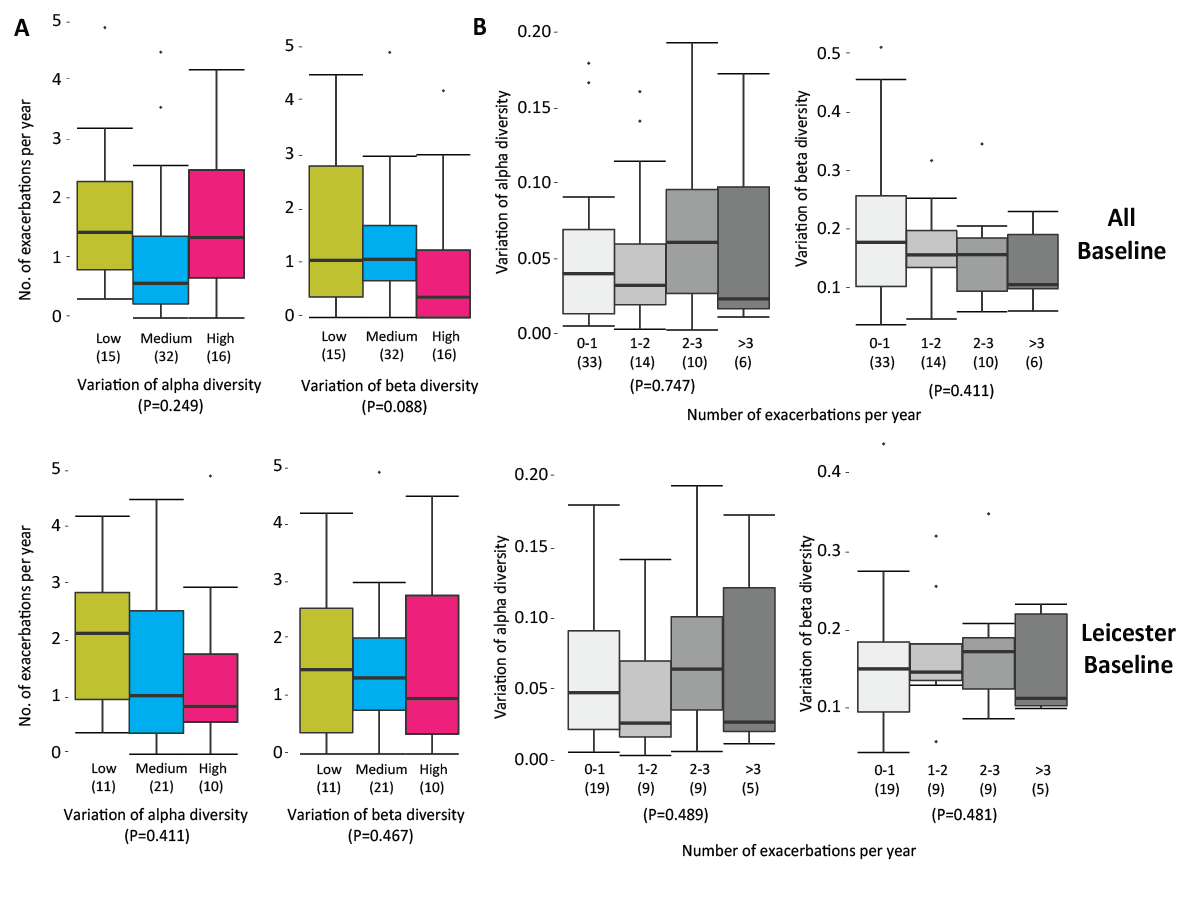
**Figure S9.** Bacterial genera associated with dysbiosis in exacerbations. Box-whisker plots showing the relative abundance changes of bacterial genera relative to the last baseline measurements between exacerbations with or without dysbiosis. The genera were ordered by the FDR-adjusted *P*-value in T-test. Error bars are within 1.5 interquartile range of the upper and lower quartiles.

****

**Figure S10.** A) Box-whisker plots showing changes of FEV1, FVC and CAT score between dysbiosis and non-dysbiosis exacerbationsat London, Leicester and Manchester. No data is available for CAT score change in dysbiosis exacerbations at London due to missing data. B) Correlations between absolute Z-score measuring exacerbation dysbiosis and changes of FEV1, FVC and CAT score from baseline. Error bars are within 1.5 interquartile range of the upper and lower quartiles.



**Figure S11.** A) Box-whisker plots showing exacerbation frequency of subjects within different quartile groups of temporal variability of alpha and beta diversity at each centre, with the first quartile defined as ‘low’, the second and third quartiles as ‘medium’ and the fourth quartile as ‘high’. B) Box-whisker plots showing temporal variability of alpha and beta diversity in subjects within different classes of exacerbation frequency at each centre. The number of samples is indicated in the parenthesis under each subgroup in the box-whisker plot. Error bars are within 1.5 interquartile range of the upper and lower quartiles. \*\*\* adj. *P*<0.001; \*\* adj. *P*<0.01; \* adj. *P*<0.05.



**Figure S12.** A) Box-whisker plots showing exacerbation frequency of subjects within different quartile groups of temporal variability of alpha and beta diversity for baseline samples only, with the first quartile defined as ‘low’, the second and third quartiles as ‘medium’ and the fourth quartile as ‘high’. B) Box-whisker plots showing temporal variability of alpha and beta diversity for baseline samples only in subjects within different classes of exacerbation frequency. Error bars are within 1.5 interquartile range of the upper and lower quartiles. Each comparison was performed on all samples and Leicester samples. The sample sizes for London and Manchester are too small to generate meaningful conclusion. The number of samples is indicated in the parenthesis under each subgroup in the box-whisker plot. Error bars are within 1.5 interquartile range of the upper and lower quartiles. The ANOVA *P*-value was indicated below the plot for each comparison.

**Table S1.** Major demographic and baseline clinical features of all subjects and subjects at each centre.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Features** | **All subjects (N=281) \*** | **Centre** | | |
| **London (N=128)** | **Leicester (N=100)** | **Manchester (N=53)** |
| Gender † | Male: 187 (70.3%), Female: 79 (29.7%) | Male: 75 (64.7%), Female: 41 (35.3%) | Male: 76 (76.8%), Female: 23 (23.2%) | Male: 36 (70.6%), Female: 15 (29.4%) |
| Age ‡ | 70 (8.1) | 71 (8.6) | 69 (7.6) | 67 (7.4) |
| BMI | 27.2 (5.4) | 26.7 (5.7) | 27.8 (5.0) | 26.9 (5.1) |
| Baseline GOLD status | 1: 30 (11.4%), 2: 132 (50.2%), 3: 78 (29.7%), 4: 23 (8.7%) | 1: 9 (7.8%), 2 62 (53.9%), 3: 33 (28.7%), 4: 11 (9.6%) | 1: 8 (8.1%), 2: 51 (51.5%), 3: 32 (32.3%), 4: 8 (8.1%) | 1: 13 (26.5%), 2: 19 (38.8%), 3: 13 (26.5%), 4: 4 (8.2%) |
| Treatment # | Antibiotics: 38 (15.3%), Steroids: 9 (3.6%), Both: 202 (81.1%) | Antibiotics: 22 (13.9%), Steroids: 1 (0.6%), Both: 135 (85.4%) | Antibiotics: 11 (15.1%), Steroids: 8 (11.0%), Both: 54 (74.0%) | Antibiotics: 5 (27.8%), Steroids: 0 (0.0%), Both: 13 (72.2%) |
| Number of cigarette packs per year 1 | 47 (30) | 45 (34) | 47 (28) | 49 (32) |
| Number of exacerbation per year 1 | 1.1 (1.6) | 1.6 (1.8) | 1 (1.7) | 0 (0.7) |
| Baseline FEV1 ‡ | 1.5 (0.6) | 1.3 (0.5) | 1.4 (0.6) | 1.7 (0.6) |
| Baseline FEV1% | 56.3 (18.9) | 54.6 (17.3) | 54.5 (17.4) | 63.7 (23.5) |
| Baseline FEV1 predicted | 2.6 (0.5) | 2.5 (0.5) | 2.7 (0.6) | 2.7 (0.5) |
| Baseline FVC | 2.9 (1.0) | 2.8 (1.0) | 2.7 (0.8) | 3.6 (0.8) |
| Baseline FEV1/FVC ratio | 0.5 (0.1) | 0.5 (0.1) | 0.5 (0.1) | 0.5 (0.1) |
| CAT score | 18.7 (7.3) | 16.7 (7.5) | 20.0 (6.4) | 19.5 (8.1) |
| CES-D score 1 | 10 (13) | 10 (12) | 10 (13) | 13 (16) |
| SGRQ total score | 47.4 (18.2) | 45.3 (15.3) | 48.7 (18.7) | 48.8 (22.3) |

† Categorical data present as number (proportion).

‡ Continuous data present as mean (SD) unless stated below.

1 Median (IQR).

\* 15 subjects were missing any demographic or clinical data.

# The numbers represent exacerbation events, thus include subjects with more than one exacerbation.

**Table S2.** Major longitudinal clinical features at baseline and exacerbations of all samples and samples at each centre.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Features** | **All** | | | **London** | | | **Leicester** | | | **Manchester** | | |
| **All (N=716)** | **Base**  **(N=446)** | **Exac (N=270)** | **All (N=301)** | **Base (N=132)** | **Exac (N=169)** | **All (N=303)** | **Base (N=221)** | **Exac (N=82)** | **All (N=112)** | **Base (N=93)** | **Exac (N=19)** |
| FEV1 | 1.4 (0.5) | 1.5 (0.5) | 1.2 (0.5) | 1.2 (0.5) | 1.3 (0.5) | 1.2 (0.5) | 1.4 (0.5) | 1.5 (0.5) | 1.2 (0.4) | 1.6 (0.6) | 1.7 (0.6) | 1.0 (0.3) |
| FVC | 2.8 (0.9) | 3.0 (0.8) | 2.5 (0.9) | 2.7 (1.0) | 3.0 (1.0) | 2.6 (1.0) | 2.7 (0.7) | 2.7 (0.7) | 2.5 (0.6) | 3.3 (0.9) | 3.4 (0.8) | 2.1 (0.9) |
| FEV1/FVC ratio | 0.5 (0.2) | 0.5 (0.2) | 0.5 (0.2) | 0.5 (0.2) | 0.5 (0.1) | 0.5 (0.2) | 0.5 (0.1) | 0.5 (0.1) | 0.5 (0.1) | 0.5 (0.4) | 0.5 (0.4) | 0.4 (0.2) |
| CAT score | 21.1 (7.4) | 19.6 (7.1) | 24.2 (7.0) | 19.9 (7.8) | 17.3 (6.9) | 22.7 (7.8) | 21.9 (6.8) | 20.5 (6.7) | 25.7 (5.5) | 21.1 (8.0) | 20.0 (8.0) | 26.4 (6.6) |
| C-reactive protein (CRP) 1 | 5.0 (11.0) | 3.0 (5.0) | 10.0 (27.0) | 6.0 (16.0) | 4.0 (6.0) | 9.0 (25.0) | 5.0 (10.0) | 3.0 (5.0) | 10.0 (31.0) | 5.0 (8.0) | 4.0 (5.0) | 13.0 (28.0) |
| Blood neutrophil count (X109 cells/L) | 5.5 (2.3) | 4.9 (1.7) | 6.2 (2.7) | 5.9 (2.5) | 5.5 (1.9) | 6.2 (2.8) | 5.3 (2.1) | 4.9 (1.7) | 6.3 (2.6) | 4.8 (1.7) | 4.5 (1.4) | 6.3 (2.2) |
| Blood lymphocyte count (X109 cells/L) | 1.8 (0.7) | 1.8 (0.6) | 1.8 (0.7) | 1.8 (0.7) | 2.0 (0.7) | 1.7 (0.7) | 1.9 (0.7) | 1.8 (0.6) | 2.0 (0.7) | 1.7 (0.6) | 1.8 (0.6) | 1.6 (0.8) |
| Blood monocyte count (X109 cells/L) | 0.7 (0.3) | 0.6 (0.2) | 0.7 (0.3) | 0.8 (0.3) | 0.8 (0.2) | 0.8 (0.3) | 0.5 (0.2) | 0.5 (0.1) | 0.6 (0.2) | 0.6 (0.2) | 0.6 (0.2) | 0.7 (0.3) |
| Blood eosinophil count (X109 cells/L) 1 | 0.2 (0.2) | 0.2 (0.2) | 0.2 (0.2) | 0.2 (0.2) | 0.2 (0.2) | 0.2 (0.2) | 0.2 (0.2) | 0.2 (0.2) | 0.2 (0.2) | 0.2 (0.2) | 0.2 (0.2) | 0.2 (0.2) |
| Blood basophil count (X109 cells/L) | 0.0 (0.0) | 0.1 (0.0) | 0.0 (0.0) | 0.0 (0.0) | 0.0 (0.0) | 0.0 (0.0) | 0.1 (0.0) | 0.1 (0.0) | 0.1 (0.0) | 0.0 (0.0) | 0.0 (0.0) | 0.0 (0.0) |
| Sputum neutrophil count % 1 | 78.8 (33.8) | 75.1 (34.0) | 84.2 (28.5) | 73.7 (34.0) | 73.0 (29.0) | 73.7 (36.0) | 77.4 (34.2) | 73.0 (34.2) | 89.5 (22.2) | 88.2 (13.5) | 88.0 (17.5) | 89.2 (7.8) |
| Sputum lymphocyte count % 1 | 0.0 (0.5) | 0.0 (0.3) | 0.2 (1.0) | 0.0 (2.0) | 0.0 (0.0) | 2.0 (3.0) | 0.2 (0.4) | 0.2 (0.5) | 0.0 (0.2) | 0.0 (0.1) | 0.0 (0.2) | 0.0 (0.0) |
| Sputum eosinophil count % 1 | 0.8 (2.0) | 0.8 (2.2) | 0.5 (2.0) | 0.0 (2.0) | 0.0 (1.0) | 1.0 (2.0) | 0.8 (2.2) | 0.8 (2.8) | 0.5 (1.5) | 1.1 (3.0) | 1.2 (3.0) | 1.0 (8.2) |
| Sputum macrophage count % 1 | 13.0 (21.2) | 14.5 (23.2) | 8.5 (19.0) | 20.0 (27.0) | 22.0 (22.0) | 20.0 (29.0) | 10.8 (19.2) | 14.2 (23.2) | 5.2 (9.5) | 7.0 (10.8) | 8.0 (12.9) | 5.5 (9.5) |
| Sputum epithelial cell count % 1 | 3.2 (8.0) | 4.0 (9.8) | 2.0 (4.8) | 0.0 (1.0) | 0.0 (0.0) | 0.0 (2.0) | 4.8 (11.8) | 6.0 (12.2) | 3.5 (7.8) | 1.9 (3.2) | 1.5 (5.2) | 3.0 (2.5) |

Data present as mean (SD) unless stated below.

1 Median (IQR).

**Table S3.** The relative abundances of major phyla and genera (average relative abundance > 1%) in the sputum microbiome. The phylum and genus level taxa were separated by the dotted line.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Taxa** | **All** | | | **London** | | | **Leicester** | | | **Manchester** | | |
| **All (N=716)** | **Base**  **(N=446)** | **Exac (N=270)** | **All (N=301)** | **Base (N=132)** | **Exac (N=169)** | **All (N=303)** | **Base (N=221)** | **Exac (N=82)** | **All (N=112)** | **Base (N=93)** | **Exac (N=19)** |
| *Proteobacteria* | 52.3 | 51.5 | 53.7 | 53.0 | 51.4 | 54.2 | 52.1 | 51.8 | 52.8 | 51.3 | 50.9 | 53.3 |
| *Firmicutes* | 28.7 | 29.0 | 28.1 | 27.2 | 28.0 | 26.5 | 30.0 | 29.4 | 31.8 | 29.3 | 29.8 | 26.8 |
| *Bacteroidetes* | 15.0 | 15.3 | 14.5 | 15.5 | 16.1 | 15.0 | 14.4 | 14.9 | 12.9 | 15.5 | 15.2 | 16.8 |
| *Actinobacteria* | 1.9 | 2.0 | 1.8 | 2.1 | 2.2 | 2.0 | 1.7 | 1.9 | 1.3 | 1.9 | 2.0 | 1.4 |
| *Fusobacteria* | 1.4 | 1.5 | 1.3 | 1.6 | 1.6 | 1.6 | 1.3 | 1.4 | 0.9 | 1.3 | 1.4 | 1.0 |
| *Haemophilus* | 25.8 | 25.6 | 26.1 | 27.0 | 26.5 | 27.3 | 25.2 | 26.1 | 22.7 | 24.0 | 23.0 | 29.3 |
| *Veillonella* | 15.8 | 16.1 | 15.3 | 13.7 | 14.0 | 13.5 | 17.3 | 16.7 | 18.9 | 17.4 | 17.7 | 16.3 |
| *Prevotella* | 13.2 | 13.5 | 12.6 | 13.6 | 14.3 | 13.1 | 12.6 | 13.2 | 11.2 | 13.4 | 13.1 | 15.0 |
| *Erwinia* | 7.0 | 7.3 | 6.6 | 7.8 | 8.2 | 7.5 | 6.1 | 6.6 | 4.8 | 7.4 | 7.6 | 6.3 |
| *Granulicatella* | 6.8 | 6.8 | 6.9 | 7.5 | 8.0 | 7.1 | 6.9 | 6.9 | 6.9 | 4.8 | 4.7 | 5.4 |
| *Cronobacter* | 6.4 | 6.5 | 6.3 | 6.5 | 5.8 | 7.0 | 6.0 | 6.4 | 4.8 | 7.3 | 7.4 | 6.3 |
| *Streptococcus* | 4.4 | 4.5 | 4.3 | 4.3 | 4.4 | 4.3 | 4.3 | 4.2 | 4.5 | 5.1 | 5.4 | 3.7 |
| *Moraxella* | 4.0 | 3.6 | 4.8 | 3.1 | 2.7 | 3.4 | 4.8 | 3.7 | 7.8 | 4.4 | 4.5 | 4.1 |
| *Actinomyces* | 1.0 | 1.1 | 0.9 | 1.1 | 1.2 | 1.1 | 0.9 | 1.0 | 0.6 | 1.1 | 1.2 | 0.7 |

**Table S4.** List of clinical variables significantly associated with microbial alpha diversity and phylum level beta diversity among samples. P-values are indicated for variables in the model. Significant variables are highlighted in asterisks.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Microbial diversity** | **Alpha diversity (Shannon)** | | | | **Beta diversity (Phylum level abundance)** | | | | | | |
| **All** | **London** | **Leicester** | **Manchester** | **All** | **London** | | **Leicester** | | **Manchester** | |
| C-reactive protein (CRP) | 0.02\* | 0.005\* | 0.04\* | 0.02\* | 0.04\* | | 0.18 | | 0.17 | | 0.01\* |
| FEV1/FVC ratio | 0.003\* | 0.10 | 0.03\* | 0.003\* | 0.04\* | | 0.51 | | 0.26 | | 0.26 |
| Age | 0.03\* | 0.12 | 0.40 | 0.97 | 0.02\* | | 0.30 | | 0.26 | | 0.01\* |
| Number of exacerbations per year | 0.26 | 0.16 | 0.04\* | 0.97 | 0.11 | | 0.86 | | 0.01\* | | 0.48 |
| White blood cell | 0.28 | 0.02\* | 0.24 | 0.01\* | 0.47 | | 0.90 | | 0.40 | | 0.27 |
| Blood neutrophil count | 0.25 | 0.03\* | 0.22 | 0.01\* | 0.59 | | 0.97 | | 0.35 | | 0.40 |
| Blood lymphocyte count | 0.67 | 0.01\* | 0.85 | 0.01\* | 0.66 | | 0.18 | | 0.71 | | 0.27 |
| Blood basophil count | 0.63 | 0.01\* | 0.55 | 0.29 | 0.81 | | 0.20 | | 0.53 | | 0.11‡ |
| FEV1 | 0.27 | 0.49 | 0.64 | 0.04\* | 0.89 | | 0.36 | | 0.96 | | 0.17 |
| FVC | 0.75 | 0.98 | 0.09 | 0.01\* | 0.95 | | 0.36 | | 0.90 | | 0.48 |
| CAT score | 0.16 | 0.16 | 0.02\* | 0.24 | 0.58 | | 0.18 | | 0.86 | | 0.53 |
| BMI | 0.90 | 0.82 | 0.19 | 0.28 | 0.93 | | 0.08 | | 0.07‡ | | 0.96 |
| Visit type | <1E-7\* | 0.001\* | <1E-7\* | 0.18 | NAa | | NA | | NA | | NA |
| Centre | <1E-7\* | NA | NA | NA | 0.001\* | | NA | | NA | | NA |

‡ Variables not statistically significant but present in the model.

a Only initial baseline samples were used.

**Table S5.** List of clinical variables significantly associated with alpha and beta diversity of the inferred metagenomic profiles of the sputum microbiome by PICRUSt [17](#_ENREF_17) among samples. P-values are indicated for variables in the model. Significant variables are highlighted in asterisks.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Inferred functional profile** | **Alpha diversity (Shannon)** | | | | **Beta diversity (L1)** | | | | **Beta diversity (L2)** | | | |
| **All** | **London** | **Leicester** | **Manchester** | **All** | **London** | **Leicester** | **Manchester** | **All** | **London** | **Leicester** | **Manchester** |
| CRP | 0.05\* | <1E-7\* | 0.55 | 0.01\* | 0.03\* | 0.02\* | 0.20 | 0.09‡ | 0.04\* | 0.14‡ | 0.29 | 0.31 |
| Basophil count | 0.63 | 0.06 | 0.01\* | 0.33 | 0.53 | 0.15‡ | 0.55 | 0.03\* | 0.58 | 0.18 | 0.74 | 0.02\* |
| FEV1 | 0.002\* | 0.06 | 0.02\* | 0.78 | 0.41 | 0.18 | 0.99 | 0.11‡ | 0.23 | 0.45 | 0.96 | 0.10‡ |
| Age | 0.81 | 0.77 | 0.89 | 0.16 | 0.11‡ | 0.29 | 0.65 | 0.11‡ | 0.14 | 0.74 | 0.73 | 0.11‡ |
| FEV1/FVC ratio | 0.002\* | 0.87 | 0.005\* | 0.49 | 0.45 | 0.59 | 0.56 | 0.84 | 0.21 | 0.58 | 0.46 | 0.76 |
| FVC | 0.005\* | 0.27 | 0.03\* | 0.84 | 0.95 | 0.24 | 0.64 | 0.91 | 0.88 | 0.24 | 0.76 | 0.84 |
| Lymphocyte count | 0.32 | 0.17 | 0.99 | 0.17 | 0.23 | 0.02\* | 0.96 | 0.40 | 0.59 | 0.05\* | 0.91 | 0.38 |
| Years smoked | 0.36 | 0.21 | 0.80 | 0.96 | 0.86 | 0.07‡ | 0.64 | 0.71 | 0.84 | 0.01\* | 0.63 | 0.81 |
| While blood cells | 0.92 | 0.01\* | 0.71 | 0.21 | 0.30 | 0.87 | 0.53 | 0.68 | 0.35 | 0.63 | 0.32 | 0.62 |
| CAT score | 0.95 | 0.63 | 0.05\* | 0.93 | 0.47 | 0.85 | 0.67 | 0.87 | 0.49 | 0.58 | 0.66 | 0.78 |
| Number of exacerbations per year | 0.28 | 0.58 | 0.90 | 0.50 | 0.11‡ | 0.30 | 0.18 | 0.39 | 0.13 | 0.67 | 0.25 | 0.39 |
| Visit type | <1E-7\* | 0.01\* | <1E-7\* | 0.32 | NAa | NA | NA | NA | NA | NA | NA | NA |
| Centre | <1E-7\* | NA | NA | NA | 0.39 | NA | NA | NA | 0.15 | NA | NA | NA |

‡ Variables not statistically significant but present in the model.

a Only initial baseline samples were used.

**Table S6.** Occurrence and average relative abundance of contaminate genera detected in sequenced negative ‘blank’ controls by Salter et al. [9](#_ENREF_9) in COPDMAP dataset. The first column (occurrence rel abundance > 0) was calculated as the fraction of samples in which each genus has abundance greater than 0. The second column (occurrence rel abundance > 0.1) was calculated as the fraction of samples in which each genus has abundance greater than 0.1. And the third column is the average relative abundance of each genus across all samples.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Occurrence (rel abundance > 0)** | **Occurrence (rel abundance > 0.1)** | **Average rel abundance** |
| ***Alphaproteobacteria*** | | | |
| *Afipia* | 0 | 0 | 0 |
| *Aquabacterium* | 0 | 0 | 0 |
| *Asticcacaulis* | 0.002793296 | 0 | 6.06E-08 |
| *Aurantimonas* | 0.025139665 | 0 | 6.97E-07 |
| *Beijerinckia* | 0 | 0 | 0 |
| *Bosea* | 0.001396648 | 0 | 6.06E-08 |
| *Bradyhizobium* | 0 | 0 | 0 |
| *Brevundimonas* | 0.044692737 | 0 | 1.97E-05 |
| *Caulobacter* | 0.001396648 | 0 | 3.03E-08 |
| *Craurococcus* | 0 | 0 | 0 |
| *Devosia* | 0.011173184 | 0 | 4.25E-07 |
| *Hoeflea* | 0 | 0 | 0 |
| *Mesorhizobium* | 0 | 0 | 0 |
| *Methylobacterium* | 0.118715084 | 0 | 4.52E-06 |
| *Novosphingobioum* | 0 | 0 | 0 |
| *Ochrobactrum* | 0.698324022 | 0 | 4.58E-05 |
| *Paracoccus* | 0.086592179 | 0 | 2.30E-06 |
| *Pedomicrobiom* | 0 | 0 | 0 |
| *Phyllobacterium* | 0.009776536 | 0 | 3.34E-07 |
| *Rhizobium* | 0.005586592 | 0 | 1.21E-07 |
| *Roseomonas* | 0 | 0 | 0 |
| *Sphingobium* | 0.036312849 | 0 | 6.22E-06 |
| *Sphingomonas* | 0.160614525 | 0 | 3.02E-05 |
| *Sphingopyxis* | 0.018156425 | 0 | 8.49E-07 |
| ***Betaproteobacteria*** | | | |
| *Acidovorax* | 0.019553073 | 0 | 7.28E-07 |
| *Azoarcus* | 0 | 0 | 0 |
| *Azospira* | 0 | 0 | 0 |
| *Burkholderia* | 0 | 0 | 0 |
| *Comamonas* | 0.008379888 | 0 | 1.82E-07 |
| *Cupriavidus* | 0.001396648 | 0 | 9.10E-08 |
| *Curvibacter* | 0 | 0 | 0 |
| *Delftia* | 0.005586592 | 0 | 1.52E-07 |
| *Duganella* | 0 | 0 | 0 |
| *Herbaspirillum* | 0 | 0 | 0 |
| *Janthinobacterium* | 0.002793296 | 0 | 1.52E-07 |
| *Kingella* | 0.995810056 | 0 | 0.000529353 |
| *Leptothrix* | 0 | 0 | 0 |
| *Limnobacter* | 0 | 0 | 0 |
| *Massilia* | 0 | 0 | 0 |
| *Methylophilus* | 0 | 0 | 0 |
| *Methyloversatilis* | 0 | 0 | 0 |
| *Oxalobacter* | 0.025139665 | 0 | 1.27E-06 |
| *Pelomonas* | 0 | 0 | 0 |
| *Polaromonas* | 0 | 0 | 0 |
| *Ralstonia* | 0.005586592 | 0 | 1.21E-07 |
| *Schlegelella* | 0 | 0 | 0 |
| *Sulfuritalea* | 0 | 0 | 0 |
| *Undibacterium* | 0 | 0 | 0 |
| *Variovorax* | 0.906424581 | 0 | 0.000159964 |
| ***Gammaproteobacteria*** | | | |
| *Acinetobacter* | 1 | 0 | 0.001896829 |
| *Enhydrobacter* | 0.083798883 | 0 | 3.49E-06 |
| *Enterobacter* | 0.997206704 | 0 | 0.000509339 |
| *Escherichia* | 0.048882682 | 0 | 1.09E-06 |
| *Nevskia* | 0.001396648 | 0 | 6.06E-08 |
| *Pseudomonas* | 0.995810056 | 0.004189944 | 0.004687729 |
| *Pseudoxanthomonas* | 0.019553073 | 0 | 8.79E-07 |
| *Psychobacter* | 0 | 0 | 0 |
| *Stenotrophomonas* | 0.780726257 | 0 | 0.000139556 |
| *Xanthomonas* | 0 | 0 | 0 |
| ***Actinobacteria*** | | | |
| *Aeromicrobium* | 0 | 0 | 0 |
| *Arthrobacter* | 0.005586592 | 0 | 2.73E-07 |
| *Beutenbergia* | 0 | 0 | 0 |
| *Brevibacterium* | 0.027932961 | 0 | 1.27E-06 |
| *Corynebacterium* | 1 | 0 | 0.001012521 |
| *Curtobacterium* | 0 | 0 | 0 |
| *Dietzia* | 0.002793296 | 0 | 2.12E-07 |
| *Geodermatophilus* | 0 | 0 | 0 |
| *Janibacter* | 0.018156425 | 0 | 4.25E-07 |
| *Kocuria* | 0 | 0 | 0 |
| *Microbacterium* | 0.159217877 | 0 | 6.22E-06 |
| *Micrococcus* | 0.054469274 | 0 | 2.03E-06 |
| *Microlunatus* | 0.002793296 | 0 | 9.10E-08 |
| *Patulibacter* | 0 | 0 | 0 |
| *Propionibacterum* | 0 | 0 | 0 |
| *Rhodococcus* | 0.909217877 | 0 | 0.000154081 |
| *Tsukamurella* | 0 | 0 | 0 |
| ***Firmicutes*** | | | |
| *Abiotrophia* | 0 | 0 | 0 |
| *Bacillus* | 0.048882682 | 0 | 2.03E-06 |
| *Brevibacillus* | 0 | 0 | 0 |
| *Brochothrix* | 0.002793296 | 0 | 9.10E-08 |
| *Facklamia* | 0.002793296 | 0 | 9.10E-08 |
| *Paenibacillus* | 0.060055866 | 0 | 2.30E-06 |
| *Streptococcus* | 1 | 0.032122905 | 0.044335024 |
| ***Bacteroidetes*** | | | |
| *Chryseobacterium* | 0.205307263 | 0 | 1.11E-05 |
| *Dyadobacter* | 0.002793296 | 0 | 6.06E-08 |
| *Flavobacterium* | 0.026536313 | 0 | 1.46E-06 |
| *Hydrotalea* | 0 | 0 | 0 |
| *Niatella* | 0 | 0 | 0 |
| *Olivibacter* | 0 | 0 | 0 |
| *Pedobacter* | 0.009776536 | 0 | 4.25E-07 |
| *Wautersiella* | 0 | 0 | 0 |
| ***Deinococcus-Thermus*** | | | |
| *Deinococcus* | 0.124301676 | 0 | 5.34E-06 |

**References**

1. General Assembly of the World Medical A. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *The Journal of the American College of Dentists* 2014;81(3):14-8.

2. International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human u. ICH harmonized tripartite guideline: Guideline for Good Clinical Practice. *Journal of postgraduate medicine* 2001;47(1):45-50.

3. Bafadhel M, McKenna S, Terry S, et al. Blood eosinophils to direct corticosteroid treatment of exacerbations of chronic obstructive pulmonary disease: a randomized placebo-controlled trial. *American journal of respiratory and critical care medicine* 2012;186(1):48-55. doi: 10.1164/rccm.201108-1553OC

4. Bafadhel M, McKenna S, Terry S, et al. Acute exacerbations of chronic obstructive pulmonary disease: identification of biologic clusters and their biomarkers. *American journal of respiratory and critical care medicine* 2011;184(6):662-71. doi: 10.1164/rccm.201104-0597OC

5. Anthonisen NR, Manfreda J, Warren CP, et al. Antibiotic therapy in exacerbations of chronic obstructive pulmonary disease. *Annals of internal medicine* 1987;106(2):196-204.

6. Rodriguez-Roisin R. Toward a consensus definition for COPD exacerbations. *Chest* 2000;117(5 Suppl 2):398S-401S.

7. Caporaso JG, Lauber CL, Walters WA, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences of the United States of America* 2011;108 Suppl 1:4516-22. doi: 10.1073/pnas.1000080107

8. Caporaso JG, Lauber CL, Walters WA, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME journal* 2012;6(8):1621-4. doi: 10.1038/ismej.2012.8

9. Salter SJ, Cox MJ, Turek EM, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC biology* 2014;12:87. doi: 10.1186/s12915-014-0087-z

10. Langmead B, Trapnell C, Pop M, et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* 2009;10(3):R25. doi: 10.1186/gb-2009-10-3-r25

11. Zhang J, Kobert K, Flouri T, et al. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 2014;30(5):614-20. doi: 10.1093/bioinformatics/btt593

12. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. *Nature methods* 2010;7(5):335-6. doi: 10.1038/nmeth.f.303

13. Quast C, Pruesse E, Yilmaz P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research* 2013;41(Database issue):D590-6. doi: 10.1093/nar/gks1219

14. Edgar RC, Haas BJ, Clemente JC, et al. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 2011;27(16):2194-200. doi: 10.1093/bioinformatics/btr381

15. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010;26(19):2460-1. doi: 10.1093/bioinformatics/btq461

16. McDonald D, Price MN, Goodrich J, et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME journal* 2012;6(3):610-8. doi: 10.1038/ismej.2011.139

17. Langille MG, Zaneveld J, Caporaso JG, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature biotechnology* 2013;31(9):814-21. doi: 10.1038/nbt.2676

18. Wang Z, Bafadhel M, Haldar K, et al. Lung microbiome dynamics in chronic obstructive pulmonary disease exacerbations. *The European respiratory journal* 2016 doi: 10.1183/13993003.01406-2015

19. Eriksson L JE, Kettaneh-Wold N. Multi- and Megavariate Data Analysis, Part 2, Advanced Applications and Method Extensions. *MKS Umetrics AB* 2006

20. Flores GE, Caporaso JG, Henley JB, et al. Temporal variability is a personalized feature of the human microbiome. *Genome biology* 2014;15(12):531. doi: 10.1186/s13059-014-0531-y

21. R Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.;ISBN 3-900051-07-0, URL <http://www.R-project.org/>

22. Kuznetsova A BPaCR. lmerTest: Tests in Linear Mixed Effects Models. 2014

23. Oksanen J BF, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH and Wagner H. vegan: Community Ecology Package. 2013