

Online supplement

A HIGH-RISK AIRWAY MYCOBIOME IS ASSOCIATED WITH FREQUENT EXACERBATION AND MORTALITY IN COPD

Pei Yee Tiew¹⁻², Alison J. Dicker³, Holly R Keir³, Mau Ern Poh⁴, Sze Lei Pang⁵, Micheál Mac Aogáin¹,
Branden Chua Qi Yu¹, Jiunn Liang Tan⁴, Huiying Xu⁶, Mariko Siyue Koh², Augustine Tee⁷, John
Arputhan Abisheganaden⁶, Fook Tim Chew⁵, Bruce E. Miller⁸, Ruth Tal-Singer⁸, James D. Chalmers³,
Sanjay H. Chotirmall¹

SUPPLEMENTARY MATERIALS AND METHODS

Ethics Statement: This study was approved by the Institutional Review Boards (IRBs) of all participating hospitals and institutions. Written informed consent was obtained from all participants. Reference numbers pertaining to ethical approvals at each site were as follows: CIRB 2016/2549, CIRB: 2016/2715, CIRB 2017/2933, CIRB 2017/2109 (all mutually recognized by DSRB, Singapore), UMMC 2018725-6524 (Malaysia). Non-diseased (healthy) control recruitment was approved by Nanyang Technological University (NTU) under IRB-2017-12-010 (Singapore). Patients in the UK were recruited under approval granted by the East of Scotland Research Ethics Committee (13/ES/0030).

COPD specimen collection and processing (sputum and serum): *Sputum:* Spontaneously expectorated or induced 'representative' sputum was obtained from a deep cough with the assistance of a chest physiotherapist (where appropriate) and collected in sterile containers and transported (on ice) for evaluation [1]. An equal volume of Sputasol (Oxoid Limited, UK) was added to each sample and shaken for 15 minutes at 37°C. Sputasol-homogenised samples mixed with two volumes of RNeasy (Life Technologies, New Zealand) were stored at -80°C for subsequent DNA extraction and mycobionome analysis as described [2]. *Serum and Plasma:* Where appropriate, venous blood draws were performed and specimens collected in serum or EDTA vacutainer tubes (BD biosciences), centrifuged at 1300g for 10 minutes at 18°C to separate serum, where appropriate, which was used for subsequent immunological studies. All specimens from clinical sites were transported promptly, appropriately and processed centrally in Singapore to ensure consistency and standardization of all assessments. All sputum was extracted using the same DNA extraction kit (Zymo reagents) and stored at -80°C. To ensure quality control of materials transported from sites outside Singapore, specimens were temperature controlled and their integrity checked on arrival to Singapore before experimental use. All experiments were performed at a single site: Nanyang Technological University, Singapore.

Non-diseased (healthy) sputum collection: Spontaneously expectorated sputum was obtained through directed coughing using the Huff cough manoeuvre by direct supervision and with clear instruction from trained medical personnel [3]. Specimens were collected in sterile containers and transported (on ice) to the laboratory for further processing and evaluation.

Sputum DNA extraction: Sputum DNA was extracted using the Zymo DNA isolation kit with an appropriate blank control (Zymo Research, Irvine, CA, USA) as per manufacturer's instructions. Briefly, sputum was centrifuged at 13,000g for 10 minutes and resultant pellets resuspended with elution buffer. Next, the homogenised samples were incubated at 55°C for 10 minutes with storage buffer and proteinase K. Subsequently, samples containing Genomic binding buffer was transferred to the Zymo-spin column and centrifuged for 1 minute at 12,000g. The columns were washed and final DNA eluted into 50uL of elution buffer. All DNA was quantified using a Qubit. fluorometer 2.0 double-stranded DNA (dsDNA) assay (Invitrogen, Carlsbad, CA, USA).

PCR amplification, library preparation, mycobiome sequencing and data processing: The Fungal Internal Transcribed Spacer (ITS)-2 region was selected for amplification using the following primers: Fseq: ATGCCTGTTTGAGCGTC; Rseq: CCTACCTGATTTGAGGTC [4]. Primers were selected based on published work illustrating that these were most optimal for mycobiome sequencing approaches in respiratory specimens [4]. Forward and reverse adaptors were added to the ITS-2 primer pairs (Integrated DNA technologies, IDT, Coralville, Iowa, USA) as per Illumina fungal metagenomics protocols. PCR conditions and cycles were optimised as follows: 10 ul KAPA HiFi HotStart ReadyMix polymerase (Kapa Biosystems, Salt River, Cape Town, South Africa) was added to 4 uL (2uM) of primers (2 ul each forward and reverse) (IDT, Coralville, Iowa, USA) and 6uL of sputum DNA (total volume: 20 uL). Initial denaturation was performed at 95°C for 3 minutes followed by 35 cycles at 95°C for 15 seconds, 50°C for 15 seconds and 72°C for 30 seconds with a final extension step at 72°C for 7 minutes using a Verti Thermal cycler (Invitrogen, Carlsbad, CA, USA). After amplification, the PCR products were visualized using a 1.5% agarose gel to ensure successful amplification and an

appropriately sized band. PCR products were then purified using Ampure XP beads (Beckman Coulter Genomics, part #A63881) with a modified two-step clean up: first with an 0.6 ratio volume of Ampure XP beads to PCR product, followed by an 0.8 ratio where supernatant was collected in-between and a final elution performed in 20 uL elution buffer. Second stage index PCR and clean up were performed to include the index sequences and flow-cell attachment regions. Agarose gel electrophoresis (1.5%) was performed to ensure successful purification and the absence of primer dimers prior to pooling of the library. Final libraries were denatured and quantified using the Kappa library quantification kit (Kapa Biosystems, Salt River, Cape Town, South Africa) and pooled libraries of 8.5pM with 20% phiX loaded onto the Illumina MiSeq platform (2x 101bp reads) (Illumina, San Diego, CA, USA) in parallel to extraction blanks and respective negative controls for the various PCR steps. Mycobiome sequencing was then performed on an Illumina MiSeq platform (2x 101bp reads) (Illumina, San Diego, CA, USA). ITS analysis was performed using Illumina Miseq platforms as previously described [4]. Briefly sequences were mapped to the UNITE database (v7.2) described by Kõljalg *et al.* using the Illumina ITS Metagenomics pipeline (version 1.0.1; Basespace labs, San Diego, CA, USA) with high-performance implementation of the Ribosomal Database project (RDP) classifier [5, 6]. All sequence data has been uploaded to the National Center for Biotechnology Information (NCBI) Sequence read archives (SRA) under project accession PRJNA609892. <https://dataview.ncbi.nlm.nih.gov/object/PRJNA609892?reviewer=fgrp8fvt98lcf28lk7icfmqu8m>

Negative controls and extraction blanks: A total of fourteen negative control samples (i.e. PBS) and extraction blanks (i.e. Zymo reagents) were sequenced and included with each respective sequencing run. The average DNA concentration (in nanograms) and number of assigned reads (at genus level) compared between the sputum samples and blank controls are illustrated in Figure E5. Average sequence read count per ul for extraction blanks, sequencing blanks and the sputum samples were 342, 9 and 42,556 respectively (Figure E5). Contamination was identified using the R “Decontam” package with a prevalence threshold of 0.5 [7]. All fungal genera identified as contaminants (including those present in the sequencing blanks, as illustrated in Table E1) were removed from downstream analysis.

Immuno-dot-blot assays to measure specific-IgE responses against fungi: Using plasma from a subset of n=42 COPD patients as previously described, specific IgE (sIgE) immuno-dot blot assays against the following fungi were performed: *Aspergillus*, *Curvularia*, *Penicillium*, *Cladosporium*, *Schizophyllum*, *Fusarium* and *Trametes* using methodology as previously described [8]. Briefly, fungi were isolated from Singapore outdoor air using the Spin-Air agar impactor (IUL S.A., Barcelona, Spain). The air was impacted onto various nutrient media, single colonies selected and re-streaked onto fresh plates. This process was repeated for three generations until single fungal isolates were obtained. Final fungal identity of isolated fungi was confirmed by Sanger sequencing. Immuno-dot blots were then performed as follows: crude fungal proteins (from isolated fungi) were blotted in duplicate onto nitrocellulose membranes with Bovine Serum Albumin (BSA) as protein control and PBS as negative control. The membranes were blocked with 1X PBS 0.1% Tween-20 for 40 minutes followed by exposure to patient serum overnight (16 hours). Washing steps were then performed in triplicate using 1X PBS 0.05% Tween-20. The washed membrane was then incubated with anti-human IgE antibodies (1:1000) conjugated to alkaline phosphatase (Sigma Aldrich, USA) for 2 hours. Nitroblue tetrazolium (NBT)/5-bromo-4-chloro-3'-indolyphosphate (BCIP) solution (Thermo Fisher Scientific) was next added to detect alkaline phosphatase activity for 10 minutes and data analyzed using Syngene imaging software [8]. Inter- and intra-assay reproducibility was 90% and 95% respectively and specific Ig-E binding calculated and subsequently presented as optical density (OD) intensity.

Statistical analysis: Data was analyzed using R (version 3.6.1, R Foundation for Statistical Computing, Vienna, Austria.) and python (version 3.8). Shapiro-Wilk tests were to assess data for normality distribution.

Statistical analysis: Data was analyzed using R (version 3.6.1, R Foundation for Statistical Computing, Vienna, Austria.) and python (version 3.8). Shapiro-Wilk tests were to assess data for normality distribution. Data are presented as medians with interquartile range (non-normally distributed data) or mean with standard deviation (normally distributed data). For between group comparisons, Wilcoxon's, Mann-Whitney-U and/or Kruskal-Wallis tests with Benjamin-Hochberg correction for false discovery

was performed as appropriate. Categorical data was analyzed using Chi-squared or Fisher-exact tests as appropriate. Contamination was identified with R “Decontam” package with a prevalence threshold of 0.5 [7]. The fungal genera identified as contaminants were removed from further downstream analysis (Table E1). Shannon and Simpson diversity indexes were calculated with the “*diversity*” function of the R package “vegan”. Bray-Curtis dissimilarity index was generated with the “*vegdist*” function of the R package “vegan”. Beta diversity was visualized using principal coordinate analysis (PCoA) plots with Bray-Curtis dissimilarity. The distance to centroids were calculated using the “*betadisper*” function of the R package “vegan”. Between group differences were obtained with permutation-based testing of multivariate homogeneity of group dispersion (PERMDISP2). For multivariate analysis adjusted for age, gender, body mass index, smoking status (as pack year exposure), post bronchodilator FEV₁, inhaled corticosteroid use and country of origin, permutational multivariate analysis of variance using distance matrices (PERMANOVA) was computed with 10,000 permutations using the “*adonis*” function of the R package “vegan”. Linear discriminant analysis effect size (LEfSe) was performed using the webtool: <http://huttenhower.sph.harvard.edu/galaxy/> and a cut-off Linear Discriminant Analysis (LDA) score of >3.5. Graphs were generated using the R package “ggplot2” and significance level defined as $p < 0.05$.

Cluster analysis: As *Candida* was the predominant fungal genera within the COPD mycobiome ($\geq 65\%$ relative abundance), it was excluded from clustering analyses to allow for the identification, if relevant, of other differentiating fungi in relation to COPD outcomes. Unsupervised hierarchical clustering with Ward’s minimum variance was performed with the “*hclust*” function of the R “Cluster” package and visualized with *heatmap.2* using the R “gplot” package. Optimal clusters were assessed with the R “Nbclust” package. Cluster stability was assessed using Jaccard similarities index with bootstrapping over 100 iterations. The mean Jaccard similarities index for clusters 1 and 2 respectively were 0.91 and 0.92 suggesting that the identified clusters were highly stable (Jaccard index ≥ 0.75 suggests valid and stable clusters [9]).

Co-occurrence network analysis: Co-occurrence network analyses was generated using methodologies described by *Fraust et al*, with slight modification [10]. Pearson's correlation, Spearman's correlation, Bray-Curtis index, Generalized Boosted Linear Models (GBLMs) and mutual information were all ensembled with a weighted Sime's test and visualized using Cytoscape (version 3.7.2) [11]. Fungal genera present in the airway mycobiome in at least 5% of the total population at an abundance of >1% were included in co-occurrence analyses.

Table E1: Summary of fungal genera identified as contaminants using the “Decontam” R package. P: true probability of a genera being present in an examined sample. Cut off <0.5 in prevalence was identified as a contaminant.

Genus	Probability
<i>Dichomitus</i>	0.028
<i>Trichosporon</i>	0.057
<i>Dekkera</i>	0.085
<i>Hortaea</i>	0.127
<i>Puccinia</i>	0.127
<i>Polyporus</i>	0.141
<i>Pichia</i>	0.258
<i>Tinctoporellus</i>	0.276
<i>Wickerhamia</i>	0.327
<i>Hyphodontia</i>	0.456

Table E2: Summary of network analyses based on fungal genus comparing very frequent (vFE) with non-frequent (non-FE) COPD exacerbators. The number of interactions (edge counts), critical (stress centrality) and influential (betweenness centrality) measures are illustrated for between group comparisons. The top taxa with the highest number of interactions, which are most influential and critical within the network are highlighted.

Top taxa	Non- FE			
	Microbes	No. of interactions (Edges)	Critical (Stress Centrality)	Influential (Betweenness Centrality)
	<i>Candida</i>	28	79	0.131846
	<i>Saccharomyces</i>	24	51	0.068446
	<i>Alternaria</i>	22	36	0.048775
	<i>Penicillium</i>	22	29	0.03046
	<i>Cryptococcus</i>	20	28	0.044562
	<i>Malassezia</i>	18	19	0.019183
	<i>Curvularia</i>	18	14	0.011748
	<i>Aspergillus</i>	18	14	0.013139
Top taxa	vFE			
	Microbes	No. of interactions (Edges)	Critical (Stress Centrality)	Influential (Betweenness Centrality)
	<i>Alternaria</i>	28	11	0.006207
	<i>Aspergillus</i>	28	11	0.006207
	<i>Cryptococcus</i>	28	11	0.006207
	<i>Curvularia</i>	28	11	0.006207
	<i>Lodderomyces</i>	28	11	0.006207
	<i>Malassezia</i>	28	11	0.006207
	<i>Penicillium</i>	28	11	0.006207
	<i>Saccharomyces</i>	28	11	0.006207
	<i>Debaryomyces</i>	26	8	0.00464
	<i>Cladosporium</i>	25	4	0.002198
	<i>Candida</i>	26	2	0.000881
	<i>Hanseniaspora</i>	26	2	0.000881
	<i>Mycosphaerella</i>	26	2	0.000881
	<i>Schizophyllum</i>	26	2	0.000881
	<i>Wickerhamomyces</i>	19	1	0.000423

Table E3: Demographics of survivors (n=51) and non-survivors (n=15) within the acute exacerbation of COPD (AECOPD) cohort. Data is presented as the number of patients (n) (with percentage; %) or median (and interquartile range; IQR) as appropriate. BMI: body mass index, CAT: COPD assessment test, FEV₁: forced expiratory volume in the 1st second, FVC: forced vital capacity, IQR: interquartile range, NA: not applicable, GOLD: global initiative for COPD, SAMA: short acting muscarinic antagonist, SABA: short-acting beta agonist, LAMA: long-acting muscarinic antagonist, LABA: long-acting beta agonist, ICS: inhaled corticosteroid.

Patient Characteristics	Survivor	Non-survivor	p-value
N	51	15	
Age (years), Median (IQR)	67 (63-73)	72 (69-78)	0.02
BMI (Kg/m ²), Median (IQR)	26.0 (22.0-30.8)	22.3 (17.8-26.0)	0.03
Gender (Male), n (%)	42 (82.4)	11 (73.3)	ns
Current smoker, n (%)	36 (70.6)	8 (53.3)	ns
Ex-smoker, n (%)	15 (29.4)	7 (46.7)	
Smoking pack years, Median (IQR)	40.0 (30.0-67.0)	50.0 (40.0-65.0)	ns
CAT Score, Median (IQR)	22.5 (18.0-27.0)	24.0 (19.0-27.0)	ns
FEV ₁ (% predicted), Median (IQR)	51.9 (41.5-72.6)	50.0 (38.6-69.9)	ns
FEV ₁ /FVC (% predicted), Median (IQR)	51.0 (42.5-59.5)	50.0 (41.5-59.0)	ns
No. of Exacerbations/year, Median (IQR)	3 (2-4)	2 (2-4)	ns
GOLD Group, n (%)			ns
A	2 (3.9)	1 (6.7)	
B	7 (13.7)	1 (6.7)	
C	9 (17.7)	1 (6.7)	
D	33 (64.7)	12 (80.0)	
GOLD Grade, n (%)			ns
1	10 (19.6)	2 (13.3)	
2	23 (45.1)	6 (40.0)	
3	15 (29.4)	5 (33.4)	
4	3 (5.9)	2 (13.3)	
LABA monotherapy, n (%)	1 (2.0)	0 (0.0)	ns
LAMA monotherapy, n (%)	3 (5.9)	2 (13.3)	ns
LABA/LAMA, n (%)	9 (17.7)	4 (26.7)	ns
LABA/ICS, n (%)	4 (7.8)	1 (6.7)	ns
LABA/LAMA/ICS, n (%)	32 (62.7)	6 (40.0)	ns
SAMA/SABA, n (%)	2 (3.9)	2 (13.3)	ns
Macrolide, n (%)	6 (11.8)	2 (13.3)	ns

Figure E1: Airway mycobiome profiles are unaffected by GOLD ABCD group in patients with stable COPD (a) Stacked bar charts illustrating the relative abundance of the top 25 fungal genera between patients in GOLD A (n=81), B (n=104), C (n=41) and D (n=111) group, which illustrates no significant change in α -diversity measured by the (b) Shannon and (c) Simpson index and (d) Principle Coordinate Analysis (PCoA) plot between the groups. GOLD: global initiative for COPD. ns: non-significant.

Figure E1

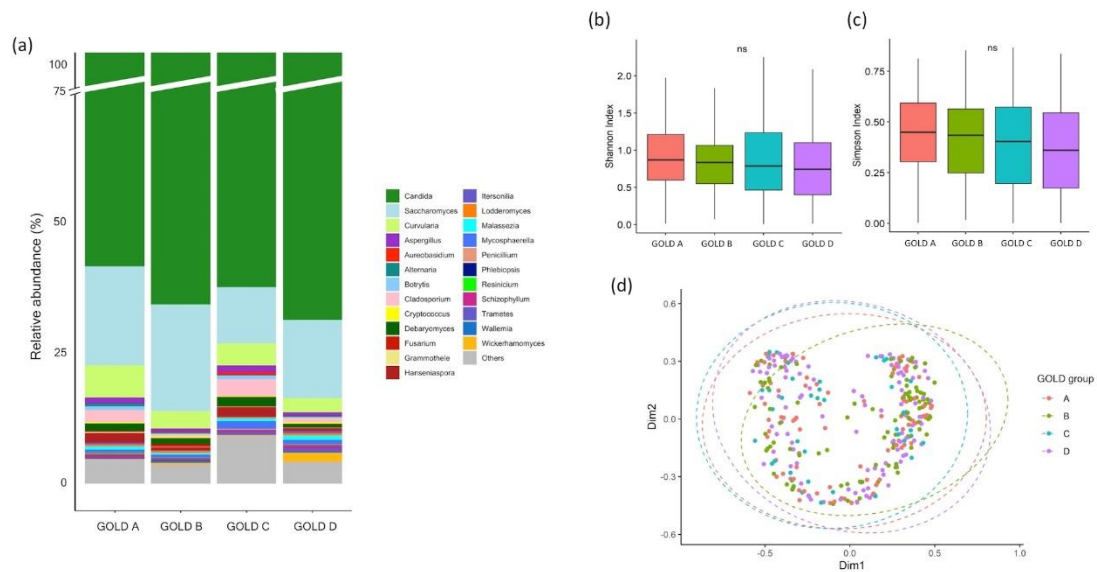


Figure E2: Airway mycobiome profiles are unaffected by GOLD lung function (FEV₁) grade in patients with stable COPD (a) Stacked bar charts illustrating the relative abundance of the top 25 fungal genera between patients in GOLD 1 (n=57), 2 (n=163), 3 (n=92) and 4 (n=25) grouping, which illustrates no significant change in α -diversity measured by the (b) Shannon and (c) Simpson index and (d) Principle Coordinate Analysis (PCoA) plot between the groups. GOLD: global initiative for COPD. ns: non-significant.

Figure E2

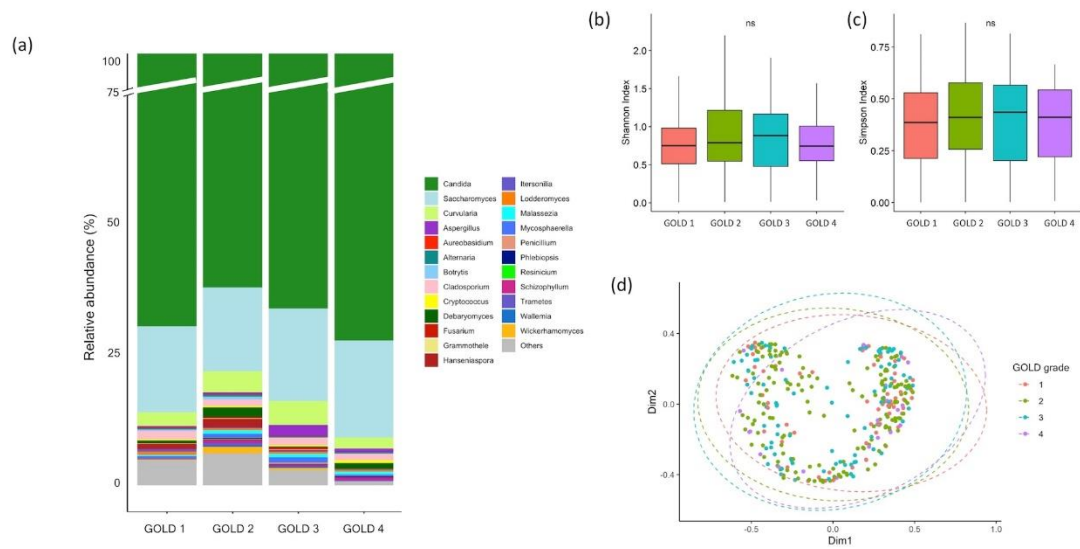


Figure E3: Airway mycobiome profiles are unaffected by treatment with inhaled corticosteroids in patients with stable COPD (a) Stacked bar charts illustrating the relative abundance of the top 25 fungal genera between patients receiving long-term treatment with an inhaled corticosteroid (ICS) (n=166) and those not receiving ICS (n=171) which illustrates no significant change in α -diversity measured by the (b) Shannon and (c) Simpson index and (d) Principle Coordinate Analysis (PCoA) plot between the groups. ns: non-significant.

Figure E3

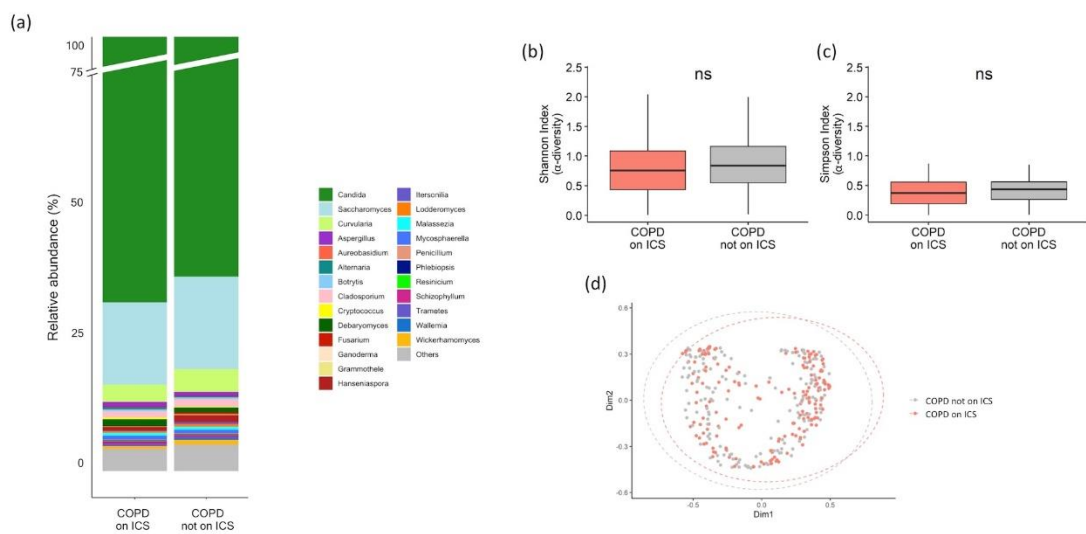


Figure E4: No significant different in the systemic immune response to control fungi allergens between clusters. Systemic specific-IgE (sIgE) binding (expressed as \log_{10} OD intensity) to (a) *Cladosporium* (b) *Fusarium* (c) *Schizophyllum* and (d) *Trametes* illustrated as bar plots for comparison between clusters. ns: non-significant.

Figure E4

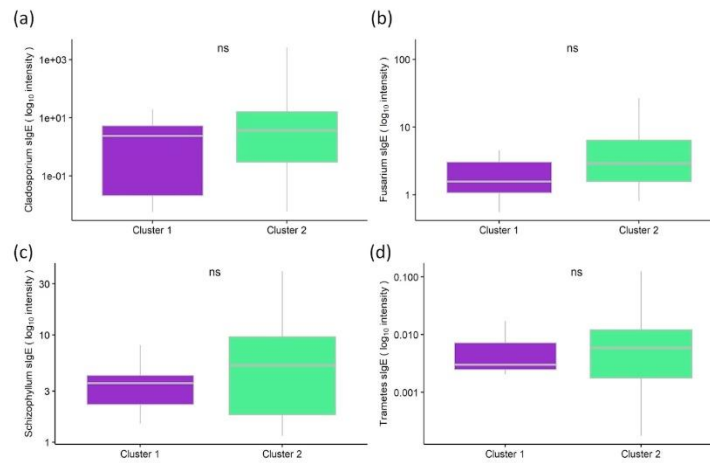


Figure E5: Box and whisker plots illustrating (a) DNA concentration (in nanograms, ng), (b) the number of assigned reads (at genus level), (c) volume of sample added to sequencing pool (in microlitre, μ l) and (d) the number of assigned reads per microliter (μ l), compared between sputum samples and negative controls: DNA extraction blanks (n=4) and sequencing blanks (n=10). Error bars represent the standard deviation (SD). (e) Stacked bar charts illustrating the taxonomy assignment for negative control samples: DNA extraction blanks (E1-E4) and sequencing blanks (S1-10).

Figure E5

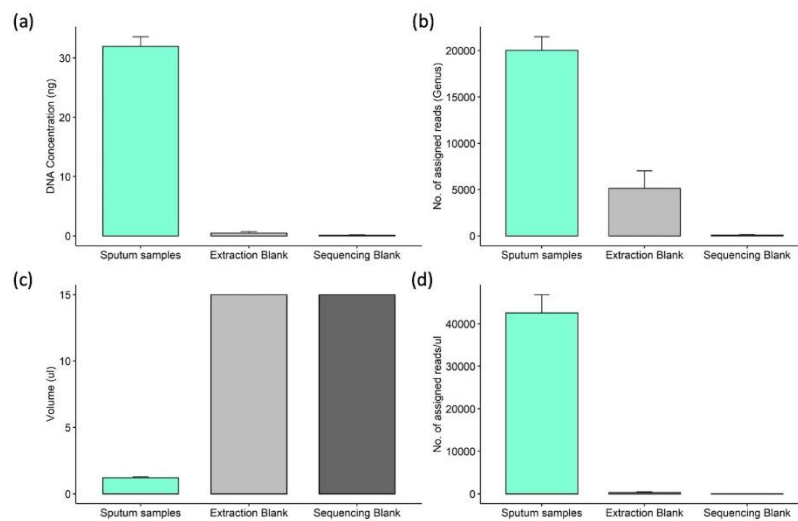


Figure E5

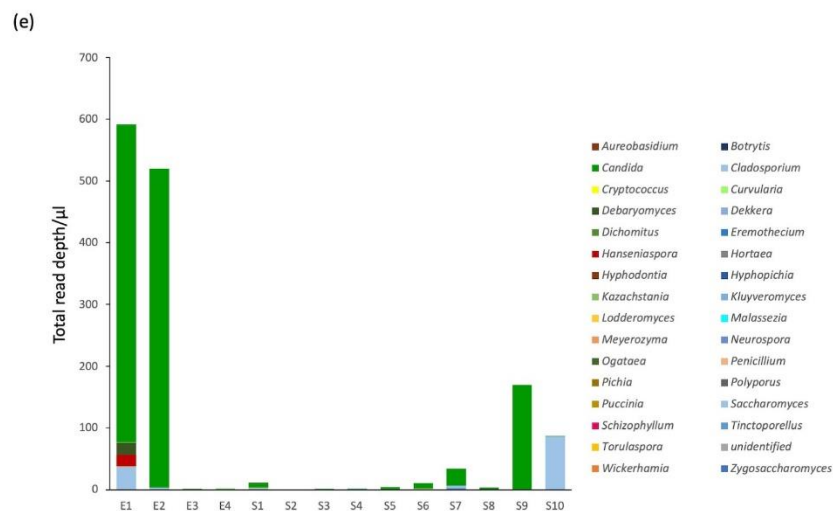


Figure E6: Violin plots illustrating the (log10) relative abundance of individual fungal genera between (a) COPD participants from Singapore and Kuala Lumpur (SG/KL) (n=175) and Dundee (DD) (n=162), (b) Non-frequent (non-FE) (n=245) and very frequent (vFE) (n=92) COPD exacerbators and (c) longitudinal analysis in n=34 COPD participants at baseline (B) (pre-exacerbation), during exacerbation (E) and following treatment (post-exacerbation) (PE).

Figure E6

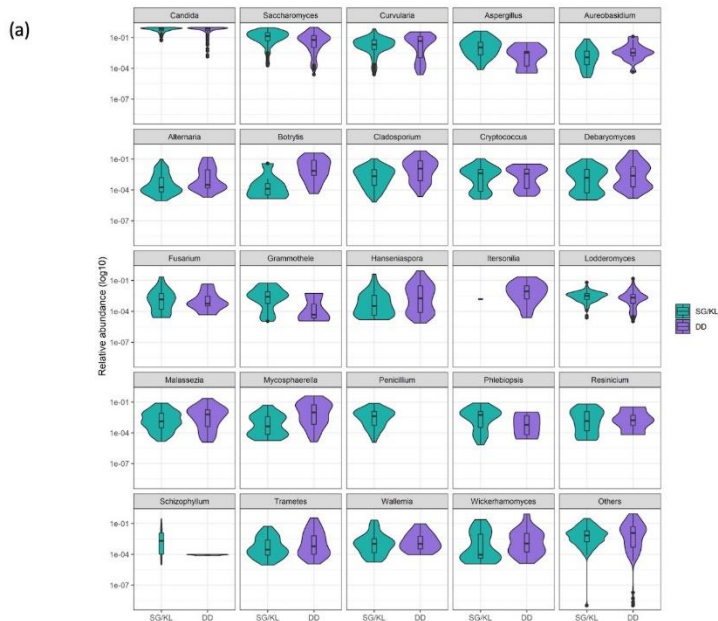


Figure E6

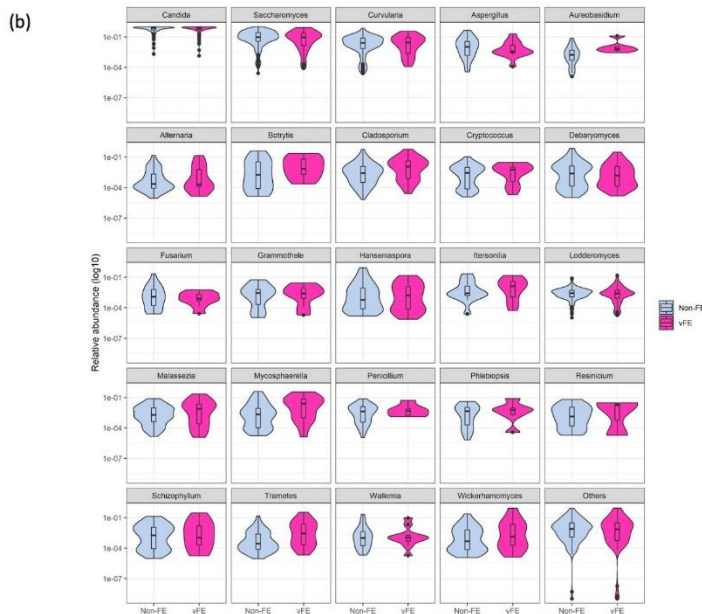


Figure E6

(c)

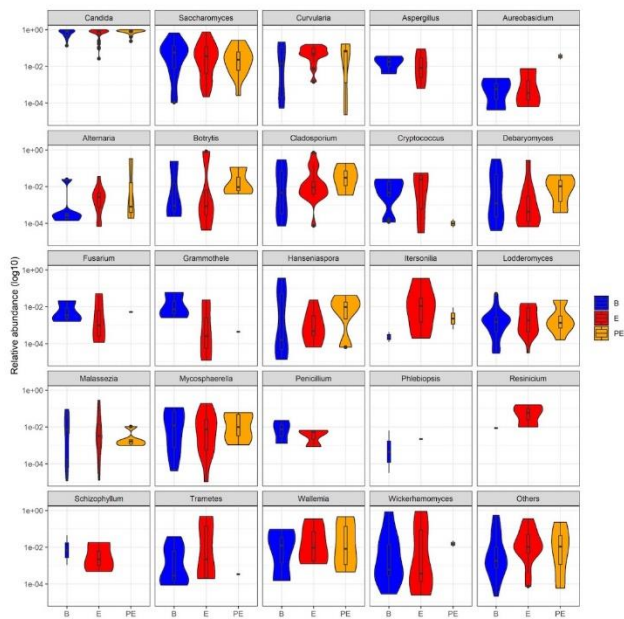
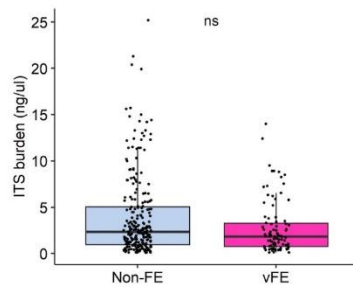


Figure E7: Scattered boxplot illustrating no significant differences in the Internal Transcribed Spacer (ITS) burden between non-frequent (non-FE) and very frequent COPD exacerbators (vFE). ns: non-significant.

Figure E7



References

1. Chotirmall SH, O'Donoghue E, Bennett K, Gunaratnam C, O'Neill SJ, McElvaney NG. Sputum *Candida albicans* presages FEV(1) decline and hospital-treated exacerbations in cystic fibrosis. *Chest* 2010; 138(5): 1186-1195.
2. Coughlan CA, Chotirmall SH, Renwick J, Hassan T, Low TB, Bergsson G, Eshwika A, Bennett K, Dunne K, Greene CM, Gunaratnam C, Kavanagh K, Logan PM, Murphy P, Reeves EP, McElvaney NG. The effect of *Aspergillus fumigatus* infection on vitamin D receptor expression in cystic fibrosis. *American journal of respiratory and critical care medicine* 2012; 186(10): 999-1007.
3. Fink JB. Forced expiratory technique, directed cough, and autogenic drainage. *Respiratory care* 2007; 52(9): 1210-1221; discussion 1221-1213.
4. Ali N, Mac Aogain M, Morales RF, Tiew PY, Chotirmall SH. Optimisation and Benchmarking of Targeted Amplicon Sequencing for Mycobiome Analysis of Respiratory Specimens. *International journal of molecular sciences* 2019; 20(20).
5. Koljalg U, Nilsson RH, Abarenkov K, Tedersoo L, Taylor AF, Bahram M, Bates ST, Bruns TD, Bengtsson-Palme J, Callaghan TM, Douglas B, Drenkhan T, Eberhardt U, Duenas M, Grebenc T, Griffith GW, Hartmann M, Kirk PM, Kohout P, Larsson E, Lindahl BD, Lucking R, Martin MP, Matheny PB, Nguyen NH, Niskanen T, Oja J, Peay KG, Peintner U, Peterson M, Poldmaa K, Saag L, Saar I, Schussler A, Scott JA, Senes C, Smith ME, Suija A, Taylor DL, Telleria MT, Weiss M, Larsson KH. Towards a unified paradigm for sequence-based identification of fungi. *Mol Ecol* 2013; 22(21): 5271-5277.
6. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 2007; 73(16): 5261-5267.
7. 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.
8. Mac Aogain M, Tiew PY, Lim AYH, Low TB, Tan GL, Hassan T, Ong TH, Pang SL, Lee ZY, Gwee XW, Martinus C, Sio YY, Matta SA, Ong TC, Tiong YS, Wong KN, Narayanan S, Bijin Au V, Marlier D, Keir HR, Tee A, Abisheganaden JA, Koh MS, Wang Y, Connolly JE, Chew FT, Chalmers

JD, Chotirmall SH. Distinct 'Immuno-Allertypes' of Disease and High Frequencies of Sensitisation in Non-Cystic-Fibrosis Bronchiectasis. *American journal of respiratory and critical care medicine* 2018.

9. Hennig C. Dissolution point and isolation robustness: Robustness criteria for general cluster analysis methods. *Journal of Multivariate Analysis* 2008; 99(6): 1154-1176.

10. Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, Huttenhower C. Microbial co-occurrence relationships in the human microbiome. *PLoS computational biology* 2012; 8(7): e1002606.

11. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* 2003; 13(11): 2498-2504.