Impact of Prophylactic and 'Recue Pack' Antibiotics on Airway Microbiome in Chronic Lung Disease

(Data Supplement)

1. DNA extraction

In a microbiological safety cabinet in a Containment Level 3 laboratory, 1 millilitre of each sample was allowed to thaw at room temperature, then centrifuged at 13,000xg for 10 min; the supernatant was discarded and the pellets were heated at 95°C for 30 min, then they were disrupted with silica beads (FastPrep® lysis matrix B) on Fast-Prep®-24 Instrument (MP Biomedicals™, Fisher Scientific, UK) speed 6.5 m/s for 45 sec. DNA was then extracted on the automated LIAISON® Ixt extraction platform using DiaSorin® Arrow DNA extraction kit.

2. Multiplex qPCR for respiratory pathogens

This method was previously described by Garcha et al. [1]. The master-mix was prepared using Platinum® quantitative PCR Supermix-UDG (Thermo-Fisher Scientific, UK) and additional magnesium chloride at final concentration of 3 mM. The thermo-cycles of 95°C for 3 min followed by 40 cycles of 95°C for 10 sec and 60°C for 45 sec were carried out on Qiagen Rotor– gene® 6000 real-time PCR machine (Corbett Research UK, Cambridgeshire, UK). An internal amplification control, Spud A: 101 bp of the PhyB gene of *Solanum tuberosum*, was used at final concentration of 0.04 pM to test for PCR inhibition.[2]

3. *P. aeruginosa* and total bacterial load Diplex qPCR

The total bacterial load (copies/μL) and load of *P. aeruginosa* (CFU/mL) were determined in sputum samples using another Taq-Man® assay diplex qPCR targeting a 466 bp fragment of 16S rRNA gene (V3-V4 hypervariable regions) using the universal bacterial primers Bact340F and Bact806R and a 65 bp fragment of the *regA* gene in *P. aeruginosa*.[3] The IAC was also employed in this method to test for PCR inhibition. The master-mix was prepared using Platinum® quantitative PCR Supermix-

Rofael SAD, *et al*. *BMJ Open Resp Res* 2023; 10:e001335. doi: 10.1136/bmjresp-2022-001335

UDG (Thermo-Fisher Scientific, UK). The thermo-cycling conditions were 95°C for 10 min followed by 40 cycles of 95°C for 15 sec and 60°C for 60 sec as carried out on Qiagen Rotor– gene® 6000 real-time PCR machine (Corbett Research UK, Cambridgeshire, UK). The signal of *P. aeruginosa* was captured on the green channel, 16S rRNA on the orange channel and IAC on the red channel. Load of *P. aeruginosa* was calculated in colony forming unit (CFU) per mL of sputum while the load of 16S rRNA was calculated as copies/ μL.

Table S1: The primers and probes for the bacterial targets and the internal amplification control in qPCR for typical respiratory pathogens

Targets	5'–3' Sequence
P. aeruginosa	
Ps Forward Primer	TGCTGGTGGCACAGGACAT
Ps Reverse Primer	TTGTTGGTGCAGTTCCTCATTG
Ps probe	[FAM] CAGATGCTTTGCCTCAA [TAM]
S. pneumoniae	
Spn Forward primer	AGTCGTTCCAAGGTAACAAGTCT
Spn Reverse primer	ACCAACTCGACCACCTCTTT
Spn Probe	ROX-TACATGTAGGAAACTATTTTCCTCACAAA-BHQ2
H. influenzae	
Hi Forward primer	CCGGGTGCGGTAGAATTTAATAA
Hi Reverse primer	CTGATTTTTCAGTGCTGTCTTTGC
Hi Probe	6FAM-ACAGCCACAACGGTAAAGTGTTCTACGT-DB
M. catarrhalis	
Mc Forward primer	GTGAGTGCCGCTTTTACAACC
Mc Reverse primer	TGTATCGCCTGCCAAGACAA
Mc Probe	6JOE-TGCTTTTGCAGCTGTTAGCCAGCCTAA-BHQ1
16S rRNA qPCR	
q16S rRNA Forward Primer (Bact340F)	TCCTACGGGAGGCAGCAGT
q16S rRNA Reverse Primer (Bact806R)	GGACTACCAGGGTATCTAATCTT
q16S rRNA probe	[ROX] CGTATTACCGCGGCTGCTGGCAC [BHQ2]
IAC	
Spud Forward primer	AACTTGGCTTTAATGGACCTCCA
Spud Reverse primer	ACATTCATCCTTACATGGCACCA
Spud Probe	Cy5-TGCACAAGCTATGGAACACCACGT-BBQ
SpudA	AACTTGGCTTTAATGGACCTCCAATTTTGAGTGTG CACAAGCTATGGAACACCACGTAAGACATAAAAC GGCCACATATGGTGCCATGTAAGGATGAATGT

4. 16S rRNA gene sequencing

A sequence library was created by amplification of V3-V4 regions of the bacterial 16S *rrna* gene through conventional PCR on the extracted metagenomic DNA using 341 forward primer (CCTACGGGNGGCWGCAG) and 805 reverse primer (GACTACHVGGGTATCTAATCC).[4, 5] Each sample was assigned a unique pair combination of standard Illumina® dual indexed primers (with adaptors attached: P5 and P7 in the forward and reverse primers respectively). The PCR master-mix per reaction was composed of; 0·2 μM for each of the forward and reverse primers, 10·8 μL Mol Taq 16S basic Master-mix (Molzym, VH Bio Limited, UK). The amount of DNA template added was adjusted such that the final DNA input per reaction was 300 ng. The thermo-cycling conditions were 95°C for 5 min, followed by 30 cycles of 95°C for 30 sec, 55°C for 40 sec and 72°C for 1 min, in addition to a final extension phase at 72°C for 10 min. The PCR products (577bp amplicons) were cleaned up using Agencourt AMPure XP beads (Beckman Coulter, UK) with a binding buffer of 2·5 M sodium chloride and 20 g% PEG-8000, 80% ethanol and EB Buffer® (Qiagen, UK) to remove nonspecific amplicons <200bp and primer dimers. DNA in the cleaned products was then quantified using Qubit™ dsDNA HS kit and Qubit® 2·0 Fluorometer (Thermo Fisher Scientific, UK). The samples were pooled in an equimolar ratio at 5 nM into one library. The library was checked on bioanalyzer. Sequencing was performed using Illumina MiSeq Platform using costume sequencing primers for read 1: GCGAGTCAGTCAGCGGACTACHVGGGTATCTAATCC, read 2: GCGAGTCAGTCAGCGGACTACHVGGGTATCTAATCC and index i7: GGATTAGATACCCBDGTAGTCCGCTGACTGACTCGC, MiSeq® Reagent Kit v2 (500 cycles) (cat no. MS-102-2003) and PhiX control V3 KIT (cat no. FC-110-3001) as internal control for the sequencing run (Illumina Cambridge, Ltd, UK). The extraction negative control and a no-template PCR control (water) were run throughout the amplification and sequencing process as negative controls to allow for the evaluation of potential contamination.

A mock community composed of equal proportions of DNA concentrations from *Streptococcus pneumoniae* ATCC 49619, *Haemophilus influenzae* ATCC 8468, *Streptococcus pyogenes* ATCC 19615, *Escherichia coli* ATCC 27853, *Klebsiella*

pneumoniae NCTC 13438, *Staphylococcus aureus* ATCC 29213, *Acinetobacter baumannii* clinical isolate, *Pseudomonas aeruginosa* clinical isolate, *Moraxella catarrhalis* clinical isolate was run as a positive control.

5. Bioinformatic and Statistical Analyses

In bioinformatic analysis we adopted the workflow established by Microbiome helper.[6] Briefly, the paired end reads were stitched together using PEAR.[7] The low-quality reads with quality score <30 over 10% of its bases and length less than 400 bp were filtered out using FASTX-toolkit (v.0.0.14).[8] The reads were then screened for possible chimeras that may have resulted from PCR using VSEARCH (v1.11.1). [9] The subsequent steps were through QIIME pipeline v1.9.1 [10] where the sequences were clustered based on 97% similarity into Operational Taxonomic Units (OTU) and taxonomic classification was assigned to OTUs using open reference OTU picking against Greengenes database version 13_8. The OTU table was then rarefied per sample to 5000. Alpha and beta diversity indices were calculated on the rarefied OTU table using QIIME. PERMANOVA and ANOSIM were used to test the associations between the metadata and the microbiome ß-diversity. PERMANOVA is a non-parametric multivariate analysis test that partitions a distance matrix among sources of variation in order to describe the strength and significance that a categorical covariate has in determining variation of distances; while, ANOSIM is a non-parametric hypothesis test that assesses whether two or more groups of samples are significantly different based on a categorical covariate in the metadata. [11]

The appropriate statistical significance tests were calculated using SPSS v. 23 or QIIME wrapper scripts. STAMP (v2.1.3) [12] was used to visualize the results and explore the OTUs showing significant differences across the two study groups using White's non-parametric t-test.[13] Whenever applicable the p-values were corrected using Benjamini-Hochberd False Discovery Rate (FDR) method for multiple comparisons.

6. Statistical Analysis

The primary outcome was defining the associations between the microbiome parameters (α and ß diversity indices, and differential abundances of OTUs/taxa) and treatment regimes. Since there was no prior work published on which formal power calculations could be based on; therefore, the sample size was established on a pragmatic recruitment rate from clinics in each study within the time frame of the doctoral degree. One sputum sample per patient at baseline that comply with the pre-specified inclusion criteria was selected to be included in the presented cross-sectional study. Missing data was dealt with by removing the related data.

For continuous data, the significance of the observed differences in the results was tested using the parametric t-test as appropriate when the normal distribution and other assumptions are satisfied, otherwise the alternative non-parametric Mann-Whitney test was used. The distribution of each continuous variable had been checked prior to the selection of the most appropriate significance test. Normal distribution was confirmed when Shapiro-Wilk test was insignificant ($p>0.05$), and skewness and kurtosis z-scores were within ±1.96 while equal variance assumption was confirmed when the Levene's test was insignificant ($p>0.05$).

Chi squared tests or Fisher exact test (when the expected frequency is <5 in 25% of the contingency table) were used to test significance in comparisons involving categorical data as appropriate. Other tests are specified in the text. Stratification analysis was used whenever a source of bias is identified using PERMANOVA test.

The statistical analysis was performed using IBM SPSS Statistics, Version 25.0. [14]. Significance was set at $p<0.05$. [15]

7. Resistome analysis

In a pilot study metagenomic sequencing was carried out on 17 sputum samples using MinION system (Oxford Nanopore Technologies, ONT, UK). Aliquots of sputum samples homogenised with Sputasol which had not been heat killed, were treated with saponin 2.5% and HL-SAN endonuclease (Arcticzymes®, Norway) to deplete human DNA as per the published method [16]. Then, metagenomic DNA was extracted on the automated LIAISON® Ixt extraction platform

The extracted DNA was purified using magnetic AMPure XP beads (Beckman Coulter, UK) and washed twice with ethanol 80%. The sequencing library was prepared by multiplexing six samples in addition to the negative control per run using the Rapid Barcoding Kit (Ref: SQK-RLB004) which involved a tagmentation step followed by a non-specific PCR amplification step of 25 cycles of: 15 sec denaturation at 95ºC, 15 sec annealing at 56 ºC and 6 min extension at 65 ºC, and a final 6 min extension step at 65 ºC which enriches and tags all DNA fragments from each sample with unique barcodes. The DNA from differently tagged samples was pooled at equal concentrations at a maximum of 500 ng each; then the pooled library was purified using magnetic AMPure XP beads (Beckman Coulter, UK) and double ethanol 80% washes. The sizes of the DNA fragments were checked on Tape Station Automated Electrophoresis (Agilent, UK) and they were around 3500 bp (a sample report is in Appendix 3.9). The library was adjusted to 10 μL of 200 fmol final DNA concentration (20 nM) prior it to adding the rapid adaptors and other reagents from the kit (as per manufacturer's directions). The sequencing was run for 48 hrs on Oxford Nanopore MinION flow cell (R.9.4.1).

ONT Albacore Sequencing Pipeline Software (version 2.3.4) was used to carry out base-calling, de-multiplex the samples per run. The genomes were assembled using the miniasm/minimap pipeline [17] and a BLAST search against the ResFinder database performed [18] to detect the antimicrobial resistance (AMR) genes. The alignments with accuracy less than 90% have been excluded. The prevalence of AMR genes within the samples was measured in part per million reads (ppm) i.e. the number of sequence reads identified as AMR genes relative to the total number of reads representing the sample.

Results

Recruitment

The respiratory clinic lists were screened for potential participants with a confirmed diagnosis of bronchiectasis (CT) or COPD (spirometry), in the period between February 2017 and May 2018, in total 147 patients were approached; and 90 patients were recruited from three outpatient respiratory clinics at the Royal Free hospital. In the period between February 2017 and July 2019, a total of 163 sputum samples were collected in the context of the longitudinal cohort study from participants with a median of two samples per patient. 84 patients provided sputum samples and completed the study.

Resistome

The most frequently detected AMR gene families were *tet* genes (32% of the total detected AMR genes) particularly 25% were *tet* genes which encode ribosomal protection proteins and confer resistance to tetracyclines by target protection such as *tet(O)* and *tet(M)* and 5% were mosaic *tet* genes such as *tet(O/32/O)*, *tet(S/M)*, *tet(W/32/O)*, *tet(O/W/32/O)*, *tet(O/W/32/O/W/O)*. Whereas, the *tet* genes which confer resistance to tetracyclines through efflux pumps such as *tet(A)*, *tet(B)* represented 2% of detected AMR genes. The prevalence of *tet* genes was significantly higher in the antibiotic prophylaxis group (p=0.041).

Macrolide resistance genes were the second most frequent, 14% of the detected AMR genes were *erm*; particularly *erm*(B) and *erm*(F) which encodes 23S ribosomal methyl-transferase and confers Macrolide–Lincosamide–StreptograminB (MLSb) resistant phenotype. Both the major facilitator efflux pump encoding gene *mef(A)* and the MDR ATP binding cassette (ABC-F) encoding gene *msr(D)* which confer resistance to macrolides and multiple antimicrobials that target the protein synthesis process such

as tetracyclines, phenicols, lincosamides, and oxazolidinones, each of these gene families represented 5% of the detected AMR genes. Although there were no statistically significant differences in the prevalence of the above described AMR genes between the two study groups, it is noteworthy to highlight that the prevalence of macrolide resistance genes was also slightly higher in the group of bronchiectasis patients not receiving antibiotic prophylaxis therapy (**Figure 8B**).

Both the MDR ABC-F ribosomal protection protein gene family *lsa(C)* and ciprofloxacin phosphotransferase encoding gene *crpP* were equally detected in both groups. The detected *crpP* gene which confer resistance to fluoroquinolones through antibiotic inactivation was not detected in the patient who was on ciprofloxacin prophylaxis therapy but in a patient on azithromycin prophylaxis having chronic *P. aeruginosa* colonization and was prescribed ciprofloxacin in the rescue pack for the treatment of exacerbations.

The following AMR genes were detected only in the group of bronchiectasis patients not receiving antibiotic prophylaxis. First AMR genes which encode antibiotic inactivating enzymes such as class A ß-lactamase enzymes encoding genes: *blaBRO*, *blaZ* and *cfxA*, macrolide phosphotransferase encoding genes *mphA*, aminoglycosides inactivating enzymes encoding genes: *aac(3)* and *ant(3")*, rifampin ADP-ribosyltransferase encoding gene *arr-3*. Second, the AMR genes which confer resistance through antibiotic target modification: sulfonamide resistant dihydropteroate synthase encoding genes *sul1* and trimethoprim resistant dihydrofolate reductase encoding gene *dfrA17* (**Figure 8B**).

Forty-three percent of the AMR genes detected such as *msr(D*), *mef(A), tet(O)*, *crpP*, *ACC (3)*, and *ANT(3")* have been known to be linked to mobile genetic elements such as plasmids, transposons and integrons. (Alcock et al., 2020)

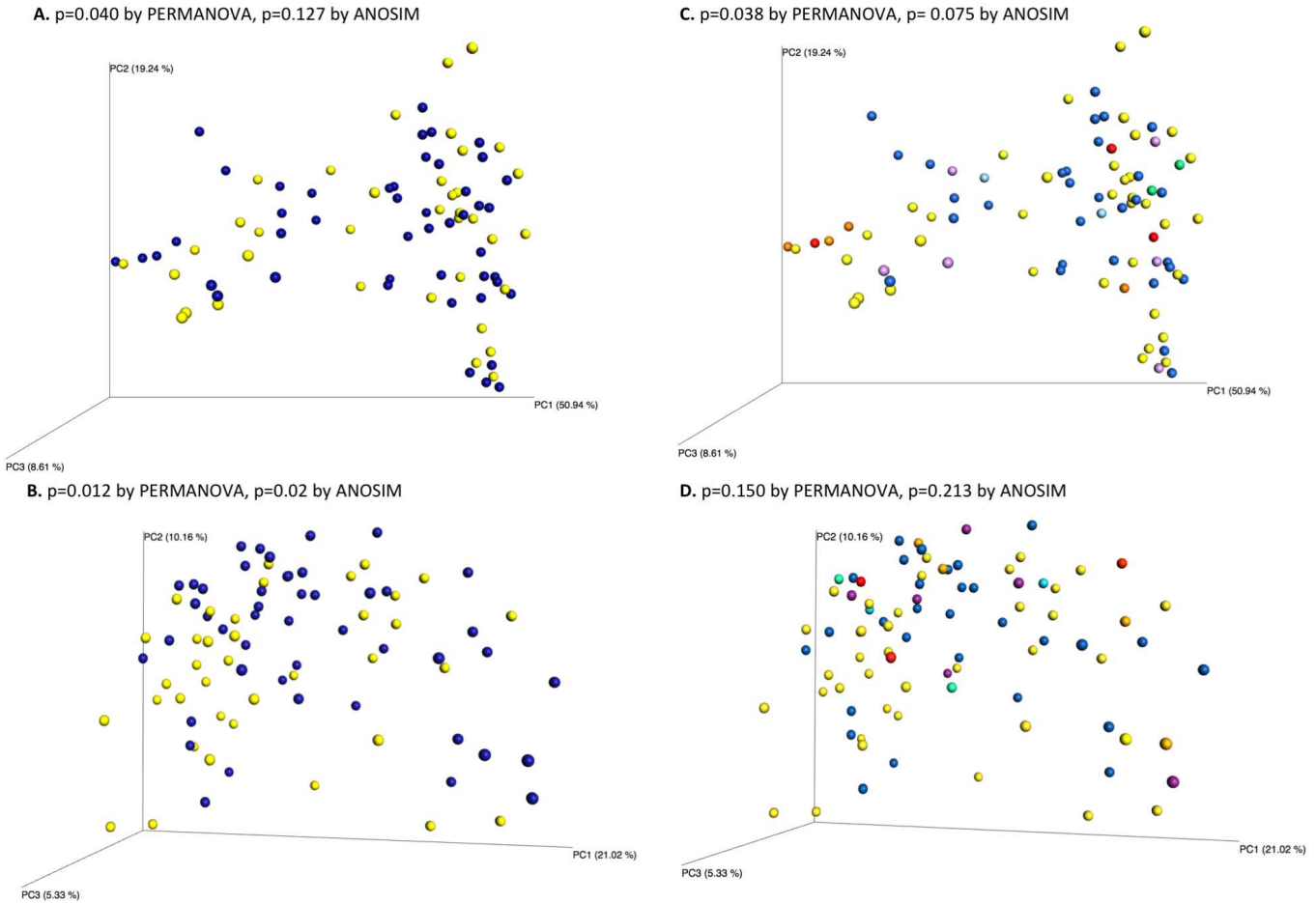


Figure S1: PCoA plot of weighted UniFrac **(A and C)** and unweighted UniFrac **(B and D)**, samples are coloured with respect to**: (A and B)** being on prophylactic antibiotic therapy (navy blue, n=47) or not (comparator group, yellow, n=37) **(C and D**) the prophylactic antibiotic regime: azithromycin (blue, n=29), clarithromycin (light blue, n=2), doxycycline (orange, n=4), ciprofloxacin (green, n=2), ß-lactam (red, n=3) co-trimoxazole (purple, n=7).

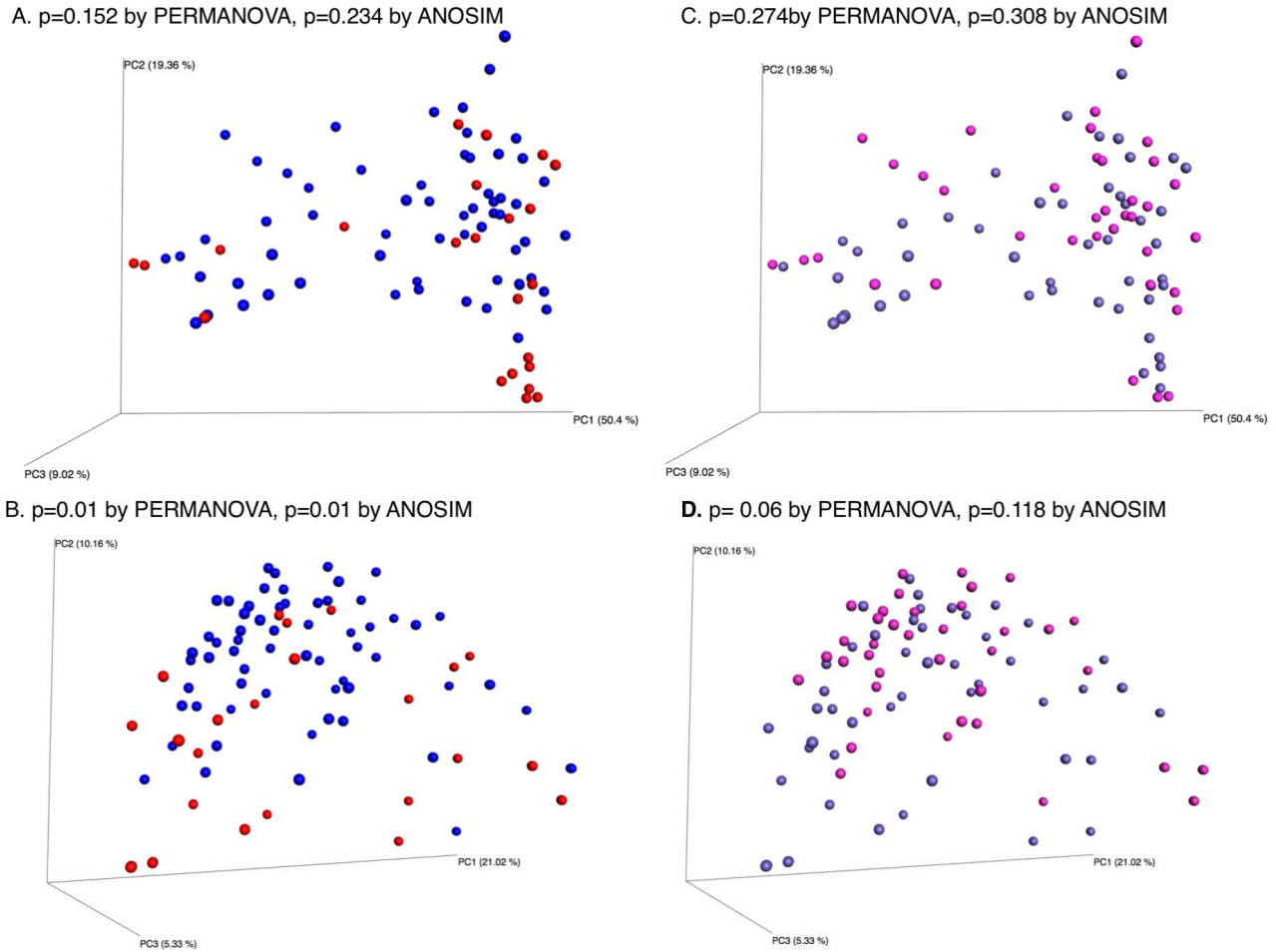
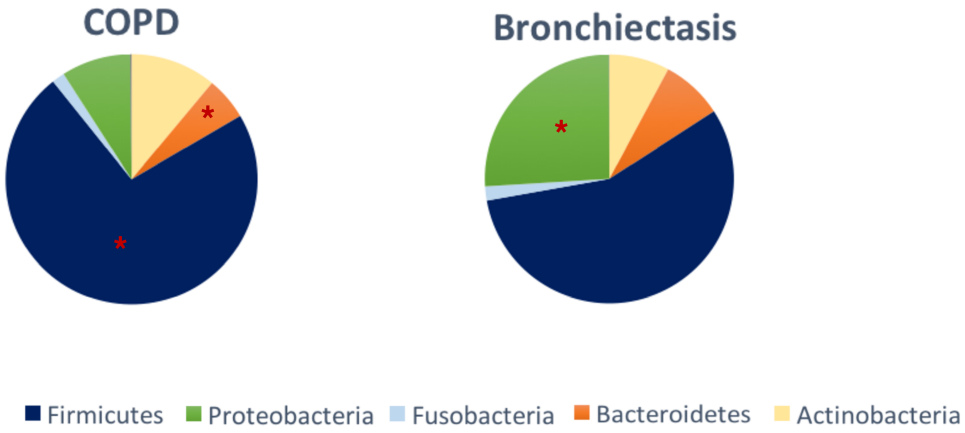


Figure S2: PCoA plot of weighted UniFrac (A and C) and unweighted UniFrac (B and D), samples are coloured with respect to:
(A and B) **chronic respiratory condition (blue: bronchiectasis, n=61), (red: COPD, n=23)**
(C and D) **CVID status pink: CVID, n=38) (purple: immunocompetent, n= 4**6).



B.

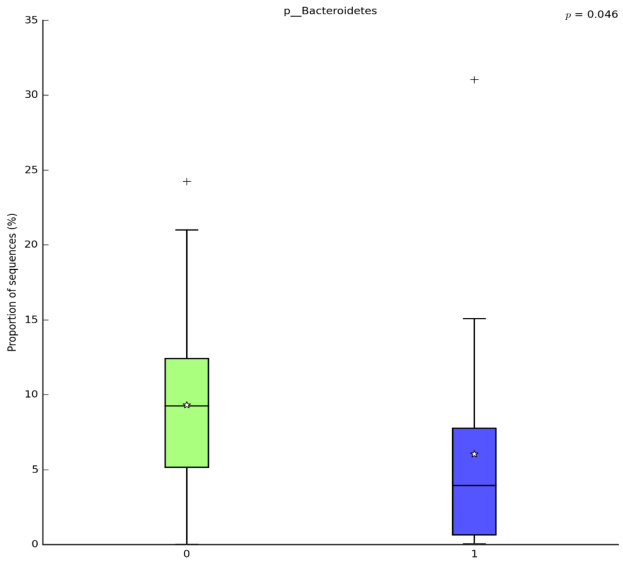


Figure S3: Comparison between the airway microbiome profiles of COPD and bronchiectasis patients at phylum level (A) In COPD, *Firmicutes* (p=0.048) was significantly higher while *Bacteroidetes* was lower (p=0.049). On the other, hand *Proteobacteria* (p=0.041) was significantly higher in bronchiectasis **(B)** *Bacteroidetes* was significantly lower in patients with airflow obstruction (p=0.046) (purple (group 1): airflow obstruction, n= 22), (green (group 0): no airflow obstruction, n= 23).

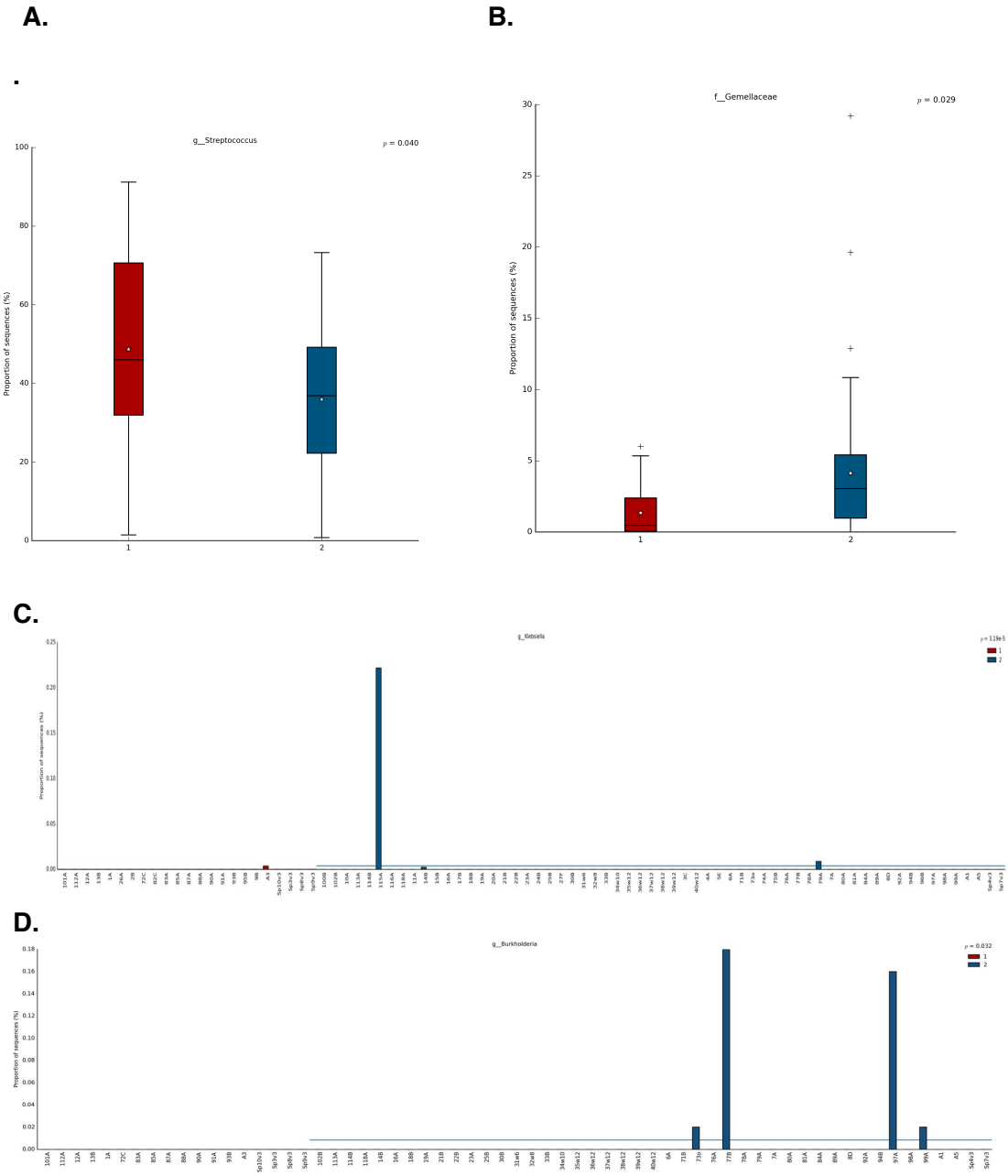


Figure S4: Differential abundance of the following taxa in the airways of COPD (group 1: red, n=23) and bronchiectasis patients (group 2: blue, n=61) **(A)** *Streptococcus* (p=0.04) more abundant in COPD; whereas, in bronchiectasis, **(B)** Family *Gemellaceae* (p=0.029), **(C)** *Klebsiella* (p=0.003) (D) *Burkholderia* (p=0.032) were more abundant and/or frequent.

A.

AB	*S. pneumoniae*								*S. aureus*		*E. faecalis*
AZM	R	R	R	R	R	R	R	S	S	S	S
E									S	S	I
CIP	S								S	S	S
TE	R	S	S	R	S	R	R	S	S	S	R
AMP	S	S	S	S	S	S	S	S	R		S
CTX	S	S	S	S	S	S	S	S			
C	S	S	S	S	S	S	S	S		S	I
RIF	S	S	S	S	S	S	S	S	S		I
VA	S	S	S	S	S	S	S	S			S
OX	R	S	R	S	S	S	R	R		R	
DA	R									S	

B.

AB	*P. aeruginosa*							*P. aeruginosa*						
CIP	S	R	R	R	S	S	S	S	S	S	S	S	S	S
LEV	S	R	S	R	S	S	S	S	S	S	S	S	S	S
TZP	S	S	S	S	S	S	S	S	S	S	S	S	S	S
AK	S	S	S	S	S	S	S	S	S	S	S	S	S	S
CAZ	S	S	S	S	S	S	S	S	S	S	S	S	S	S
CTX	R	R	S	S	R	S	S	S	S	S	S	S	S	S
MEM	S	S	S	S	S	S	S	I	S	S	S	S	S	S

Figure S5: Comparison between the AMR prevalence in the Gram positive and *P. aeruginosa* sputum isolates from patients receiving antibiotic prophylaxis therapy and those not (**A)** Antibiogram of Gram positive bacteria: *S. pneumoniae* (n=8), *S. aureus* (n=2) and *E. faecalis* (n=1) (p=0.014 by Fisher exact test) and **(B)** Antibiogram of *P. aeruginosa* isolates (n=14) (p=0.03 by Fisher exact test); antibiotic prophylaxis group (Blue) and the comparator group (orange). R: resistant (red), I: intermediate sensitivity (peach), S: sensitive (green), AB: antibiotic, AZM: azithromycin, E: erythromycin, CIP: ciprofloxacin, LEV: levofloxacin, TE: tetracycline, AMP: ampicillin, TZP: piperacillin/tazobactam, CTX: cefotaxime, CAZ: ceftazidime, MEM: meropenem, C: chloramphenicol, RIF: rifampicin, VA: vancomycin, OX: oxacillin, DA: clindamycin, AK: amikacin, grey shade: data not available.

AB	H. influenzae								H. influenzae				H. para		H. para	Pc	Mc	*M. catarrhalis*				*E. coli*		E. coli			Pr.	Pr	Ck	K. pn		Kv	Ko	K. aer	Mm
CIP	R	S	S	S	S	S	S	S	S	S	S	S	S	R	S	S	S	S	S	S	S	S	R	R	R	I	S	S	S	S	R	S	S	S	S
TE	S	R	S	S	S	S	S	S	S	I	S	S	S	R	R	S	S	S	S	S	R	S	R	R	R	R	R	R	S	S	S	S	S	S	R
AZM	R	S	S	S	S	S	S	S	R	S	S	S	S	R	S	S	S	S	S	S	S	R	R	R	S	R	R	R	S	S	R	S	S	S	S
AMP	S	R	R	R	R	R	S	S	S	S	R	R	S	R	R	R	S	S	S	S	R			R	R	R			R	R	R	R	R	R	
AMC	S	S	S	S	S	S	S	S	S	S	S	S	S	S	R	S	S	S	S	S	S	S	R	S	R	R	S	S	S	S	R	S	S	R	
CAZ																						S	R				S	S							S
CTX	S	S	S	R	S	S	S	S	S	S	S	S	S	S	S	R	S	R	S	S	S	I	R	S	S	S	S	S	S		S	S	S	S	
FEB									S								S			S	S	I					S	S					S	S	
MEM		S	S	S	S	S	S	S	S	S	S	S					S			S	S	S	S	S			S	S					S	S	
C	S	S	S	S	S	S	S	S	S	S	S	S	S	S	I				S	S	R	S	S	S	S		S			S	R	S	S	S	
SXT	S								S	S			S	S	S	S	S	S	S	S	S		R	R	S	R			S	S	R	S	S		
CN																						S		R			S	S						S	S

Figure S6: Comparison between the AMR prevalence in the Gram negative sputum isolates from patients receiving antimicrobial prophylaxis and those not**.** Each column represents antibiogram of a Gram negative isolate: *H. influenzae* (Hi, n=12), *H. parainfluenzae* (H. para, n=3), *Pasteurella canis* (Pc:, n=1), *M. catarrhalis* (Mc, n=5) *Proteus mirabilis* (Pr, n=2), *Citrobacter koseri* (Ck, n=1,) *K. pneumoniae* (K.pn, n=2), *K. variicola* (Kv, n=1), *K. oxytoca* (Ko, n=1), *K. aerogenes* (K.aer, n=1), *M. morganii* (M. mor, n=1):isolated from the sputum of chronic lung disease patients who were receiving antimicrobial prophylaxis (Blue) and those not (orange) (p=0.390 by Chi square test)

R: resistant (red), I: intermediate sensitivity (peach), S: sensitive (green), grey shade: data not available.
AB: Antibiotics, CIP: ciprofloxacin, TE: tetracycline, AZM: azithromycin, AMP: ampicillin, AMC: amoxicillin/clavulanic acid, CAZ: ceftazidime, CTX: cefotaxime, FEB: cefepime, MEM: Meropenem, C: chloramphenicol, SXT: Sulfamethoxazole/ Trimethoprim, CN: gentamycin, Hi: *H. influenzae*, H. para: *H. parainfluenzae,* Pc: *Pasteurella canis*, Mc: *M. catarrhalis*, Pr: *Proteus mirabilis*, Ck: *Citrobacter koseri*, K.pn: *K. pneumoniae*, Kv: *K. variicola*, Ko: *K. oxytoca*, K.aer: *K. aerogenes*, m: *M. morganii.*

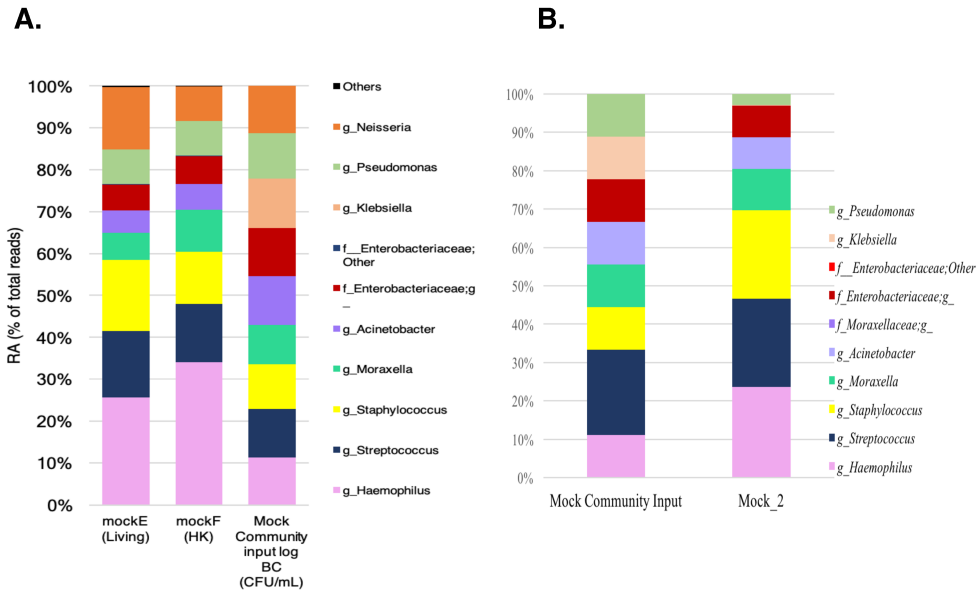


Figure S7: A. Mock community E and F used to assess the efficiency of the DNA extraction procedure and the effect of the heat killing step on the bacterial community composition representation **B.** The microbiome profile of laboratory prepared mock community; Mock_2 is composed of equal proportions of DNA concentrations from each of the constituting bacteria measured by Qubit™ HS Kit.

Table S2:The limit of quantification (LOQ) and limit of detection (LOD) of the qPCR for respiratory pathogens

	P. aeruginosa	***S. pneumoniae***	***H. influenzae***	***M. Catarrhalis***
LOQ	700 CFU/mL 2.9 log_{10}CFU/mL	25,000 CFU/mL 4.4 log_{10} CFU/mL	20,000 CFU/mL 4.3 log_{10}CFU/mL	2500 CFU/mL 3.4 log_{10}CFU/mL
LOD	350 CFU/mL 2.5 log_{10}CFU/mL	3700 CFU/mL 3.7 log_{10}CFU/mL	1000 CFU/mL 3 log_{10}CFU/mL	500 CFU/mL 2.7 log_{10}CFU/mL

Table S3: Sensitivity, Specificity and Accuracy of the 16S rRNA sequencing relativeto the qPCR

taxa	**Sensitivity**	**Specificity**	**Accuracy**
Pseudomonas	90.5%	77.2%	80.7
Moraxella	82.5%	84.1%	83.9
Haemophilus	100%	0%	41.2%
H. influenzae	92.7%	81.1%	85.8%

Sensitivity is the true positive rate
Specificity is the true negative rate [19]
Accuracy is the true positives true negatives rate [20]

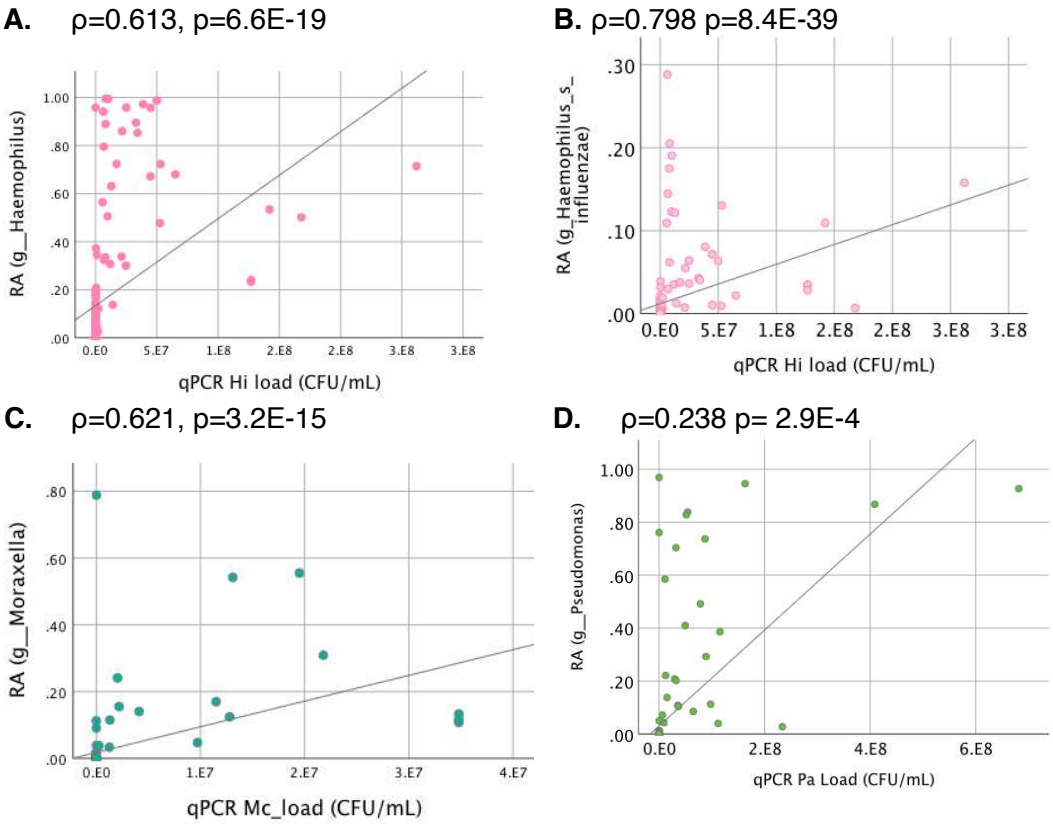


Figure S8. Correlation between the Relative Abundance (RA) results of V3-V4 16S rRNA sequencing and the quantitative loads of qPCR results (CFU/mL) in 182 tested sputum samples (A) Correlation between relative abundance of genus *Haemophilus* and load of *H. influenzae* (Hi) (B) Correlation between relative abundance of *H. influenzae* OTUs and load of *H. influenzae* (C) Correlation between relative abundance of genus *Moraxella* and load of *M. catarrhalis* (Mc) (D) Correlation between relative abundance of genus *Pseudomonas* and load of *P. aeruginosa* (Pa). ρ: Spearman's rho Correlation.

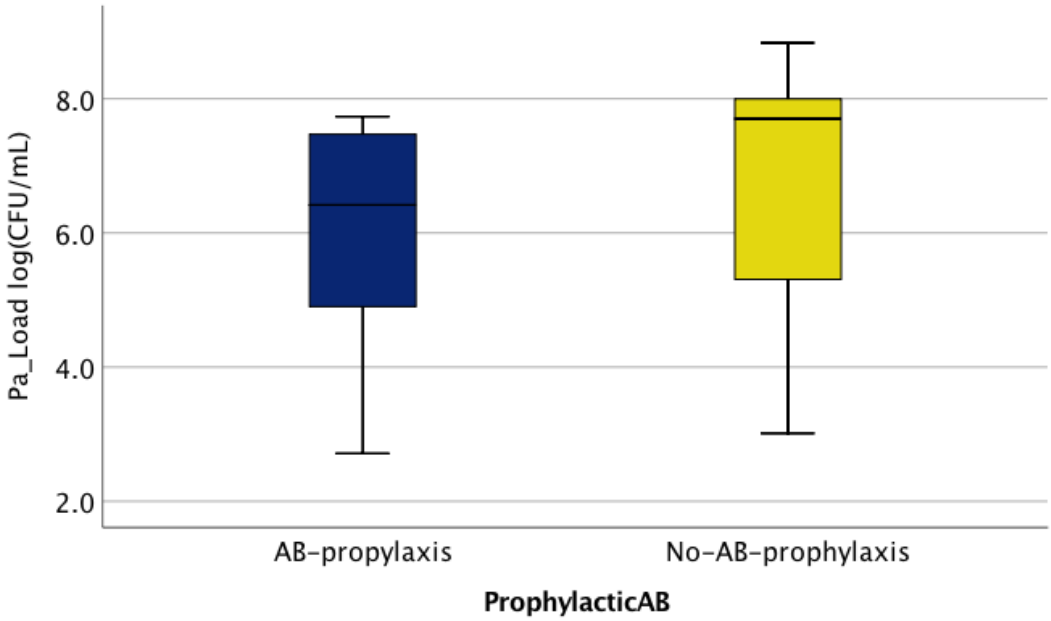


Figure S9: Comparison between the *P. aeruginosa* load in the sputum of patients in the antibiotic (AB) prophylaxis group (navy blue, n= 6) and the comparator group not receiving antibiotic prophylaxis (yellow, n=12) after excluding the CVID patients in both study groups (p=0.223 by MW)

References:

1. Garcha DS, Thurston SJ, Patel AR, et al. Changes in prevalence and load of airway bacteria using quantitative PCR in stable and exacerbated COPD. *Thorax* 2012: 67(12): 1075-1080.

2. Nolan T, Hands RE, Ogunkolade W, et al. SPUD: a quantitative PCR assay for the detection of inhibitors in nucleic acid preparations. *Analytical biochemistry* 2006: 351(2): 308-310.

3. Stressmann FA, Rogers GB, Marsh P, et al. Does bacterial density in cystic fibrosis sputum increase prior to pulmonary exacerbation? *Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society* 2011: 10(5): 357-365.

4. Klindworth A, Pruesse E, Schweer T, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic acids research* 2013: 41(1): e1.

5. Hugerth LW, Wefer HA, Lundin S, et al. DegePrime, a program for degenerate primer design for broad-taxonomic-range PCR in microbial ecology studies. *Applied and environmental microbiology* 2014: 80(16): 5116-5123.

6. Comeau AM, Douglas GM, Langille MG. Microbiome Helper: a Custom and Streamlined Workflow for Microbiome Research. *mSystems* 2017: 2(1).

7. Zhang J, Kobert K, Flouri T, et al. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics (Oxford, England)* 2014: 30(5): 614-620.

8. A. G. FASTX-Toolkit: FASTQ/A short-reads pre-processing tools. Cold Spring Harbor Laboratory Cold Spring Harbor, NY, 2009.

9. Rognes T, Flouri T, Nichols B, et al. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 2016: 18(4).

10. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. *Nature methods* 2010: 7(5): 335-336.

11. Jari Oksanen, F. Guillaume Blanchet, Michael Friendly, et al. Community Ecology Package: Ordination, Diversity and Dissimilarities. Package 'vegan. Package 'vegan', 2018.

12. Parks DH, Tyson GW, Hugenholtz P, et al. STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics (Oxford, England)* 2014: 30(21): 3123-3124.

13. White JR, Nagarajan N, Pop M. Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput Biol* 2009: 5(4): 10.

14. IBM. IBM SPSS Statistics for Windows. Version 25.0 ed. IBM Corp, Armonk, NY, 2017.

15. Petrie A. Medical Statistics at a Glance. 4th ed.. ed. Newark: John Wiley & Sons, Incorporated, 2019.

16. Charalampous T, Kay GL, Richardson H, et al. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. *Nature biotechnology* 2019: 37(7): 783.

17. Li H. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics (Oxford, England)* 2016: 32(14): 2103-2110.

18. Zankari E, Hasman H, Cosentino S, et al. Identification of acquired antimicrobial resistance genes. *Journal of Antimicrobial Chemotherapy* 2012: 67(11): 2640-2644.

19. Lalkhen AG, McCluskey A. Clinical tests: sensitivity and specificity. *Continuing Education in Anaesthesia Critical Care & Pain* 2008: 8(6): 221-223.

20. Baratloo A, Hosseini M, Negida A, et al. Part 1: Simple Definition and Calculation of Accuracy, Sensitivity and Specificity. *Emerg (Tehran)* 2015: 3(2): 48-49.