Airway Microbiome in Adult Survivors of Extremely Preterm Birth (The EPICure Study)

Sylvia A D Rofael, Timothy D McHugh, Rachael Troughton, Joanne Beckmann, David Spratt, Neil Marlow, John R Hurst

Online Data Supplement

Sample Collection

Samples were collected in the period between January 2014 and March 2015 at University College London Hospital (UCLH).

Sputum induction was carried out by inhalation of nebulised hypertonic saline: 4% hypertonic saline over 5 or 10 min increasing to 7% over a further 10 min as tolerated in non-asthmatics; or 0.9% hypertonic saline over 5 min increasing to 3% over another 5 min and then 4% for a further 10 min as tolerated in asthmatics. FEV1 and oxygen saturation (SPO2) was monitored every 5 min after baseline and the procedure was stopped if FEV1 dropped to <80% baseline level or if SPO2 dropped to <92% or if the participant requested to stop the procedure, and if required in the event of bronchospasm, 400 µg salbutamol was administered.

Participants were asked to take a sip of water and blow their nose before expectorating to minimise contamination from saliva and post-nasal drip.

As soon as a satisfactory sample was collected, the sample was placed immediately on ice and transported on dry ice to the UCL-RFH biobank lab for processing and freezing within an hour. The samples were split into aliquots and stored at -80°C in the UCL-RFH biobank. One aliquot of each sample was unprocessed and these were selected for the microbiome study. Samples were sorted into three groups by the clinical team in the EPICure study based on medical history of participants; those who were born prematurely with no history of bronchopulmonary dysplasia; those who were born premature with history of bronchopulmonary dysplasia; and full term born controls.

The samples were removed from -80°C freezers and allowed to thaw at room temperature before 500µL of each sample was aliquoted for testing. The samples were treated with an equal volume of freshly diluted Sputasol® (as per the manufacturer's instructions; Oxoid, UK). The samples were thoroughly mixed using a vortex (Clifton™ Cyclone vortex mixer, Nickel-Electro Ltd, UK) for 10 seconds and incubated at room temperature (24-25°C) for 15 min with vortexing 2 or 3 times before they were heated at 95°C for 30 min.

Multiplex qPCR for respiratory pathogens

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, was used at final concentration of 0.04 pM to test for PCR inhibition [1].

16S rRNA gene sequencing

A sequence library was created by amplification of V5-V7 regions of the bacterial 16S *rrna* gene through conventional PCR on the extracted metagenomic DNA using 785 forward primer (785F: GGATTAGATACCCBRGTAGTC) and 1175 reverse primer (1175R: ACGTCRTCCCCDCCTTCCTC). 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; 0.4 µM for each of the forward and reverse primers, 0.625 units Mol

Taq 16S/18S basic Master-mix (Molzym, VH Bio Limited, UK) with additional 0.5 mM magnesium chloride and 800 µM deoxynucleosides triphosphate (dNTPs) mixture. The amount of DNA template added was adjusted such that the final DNA input per reaction was around 300ng. 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. Seventy-four samples produced an amplicon at the expected size of 504 bp. The PCR products 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 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 10 nM into one library. The library was checked on bioanalyzer. Sequencing was performed using Illumina MiSeq Platform using costume sequencing primers for read 1: ACGTACGTACGTGGATTAGATACCCBRGTAGTC, read 2: AGTCAGTCAGCCACGTCRTCCCCDCCTTCCTC and index i7: GAGGAAGGDGGGGARGACGTGGCTGACTGACT, 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.

Bioinformatics and Statistical Analyses

In bioinformatic analysis we adopted the workflow established by Microbiome helper [2]. Briefly, the sequencing reads were primary analysed and demultiplexed and exported to Illumina cloud-based BaseSpace. The paired end reads were stitched together using PEAR.[3]. The low-quality reads with quality score <30 over 10% of its bases and length less than 350 bp were filtered out using FASTX-toolkit (v.0.0.14) [4]. The reads were then screened for possible chimeras that may have resulted from PCR using VSEARCH (v1.11.1) [5]. The subsequent steps were through QIIME pipeline v1.9.1 [6] 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 4000 reads removing all samples having number of reads less than 1000 reads (4 samples). Alpha and beta diversity indices were calculated on the rarefied OTU table using QIIME. The appropriate statistical significance tests were calculated using SPSS v. 23 or QIIME wrapper scripts after checking the normal distribution assumption of the continuous variables through Shapiro-Wilk test, Skewness and Kurtosis z-scores, Normal Q-Q Plot and Levene test for homogeneity of variance. Both PERMANOVA and ANOSIM tests were performed on weighted Unifrac distance matrix through QIIME. STAMP (v2.1.3) [7] was used to visualize the results and explore the OTUs showing significant differences across the groups. Whenever applicable the *P*-values were corrected using Benjamini-Hochberd False Discovery Rate (FDR) method for multiple comparisons on filtered OTU tables to compare highly abundant taxa (RA>5%).

Results

Sensitivity and Specificity of methods

Conventional bacteriology has a diagnostic cut-off of 10^6 CFU/mL [8], the qPCR had a sensitivity of 3700 CFU/mL for *S. pneumoniae,* 1000 CFU/mL for *H. influenzae* and 500 CFU/mL for *M. catarrhalis* [9]. The specificity of the multiplex qPCR was previously determined by in-silico analysis of the primers specificity and validated by screening against a range of airway bacteria and viruses (data not showed). 16S rRNA sequencing is a semi-quantitative method in which the results are expressed as the relative abundance. Currently, the resolution of taxonomic classification cannot go beyond the genus level for most OTUs. Comparing the sequencing results with the qPCR results the sensitivity and specificity of 16S rRNA v5-v7 sequencing were 82% and 35% respectively for *H. influenzae*, 100% and 74% respectively for *M. catarrhalis* and 100% and 0% respectively for *S. pneumoniae*. Sensitivity in this context is defined as true positive rate and specificity as the true negative rate [10].

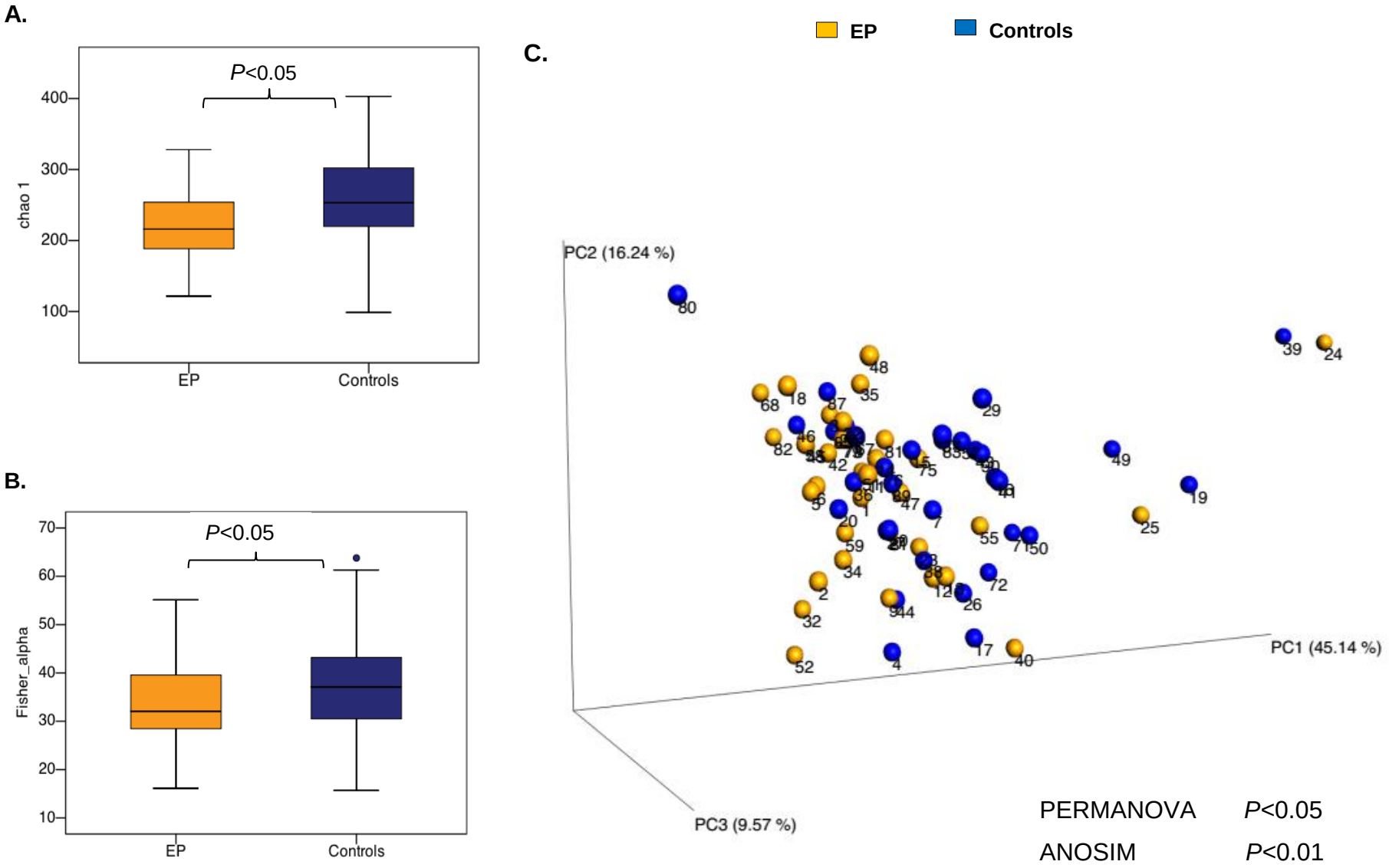


Figure E1: **Comparison of the richness and diversity of microbial communities in sputum between the pre-term birth survivors (EP) (orange) and controls (blue)** Richness and α diversity measured by **(A)** Chao 1 and **(B)** Fisher-alpha diversity index, **(C)** Principal Coordinate Analysis (PCoA) of weighted UniFrac β-diversity index. Sample size: 37 EP and 33 Controls

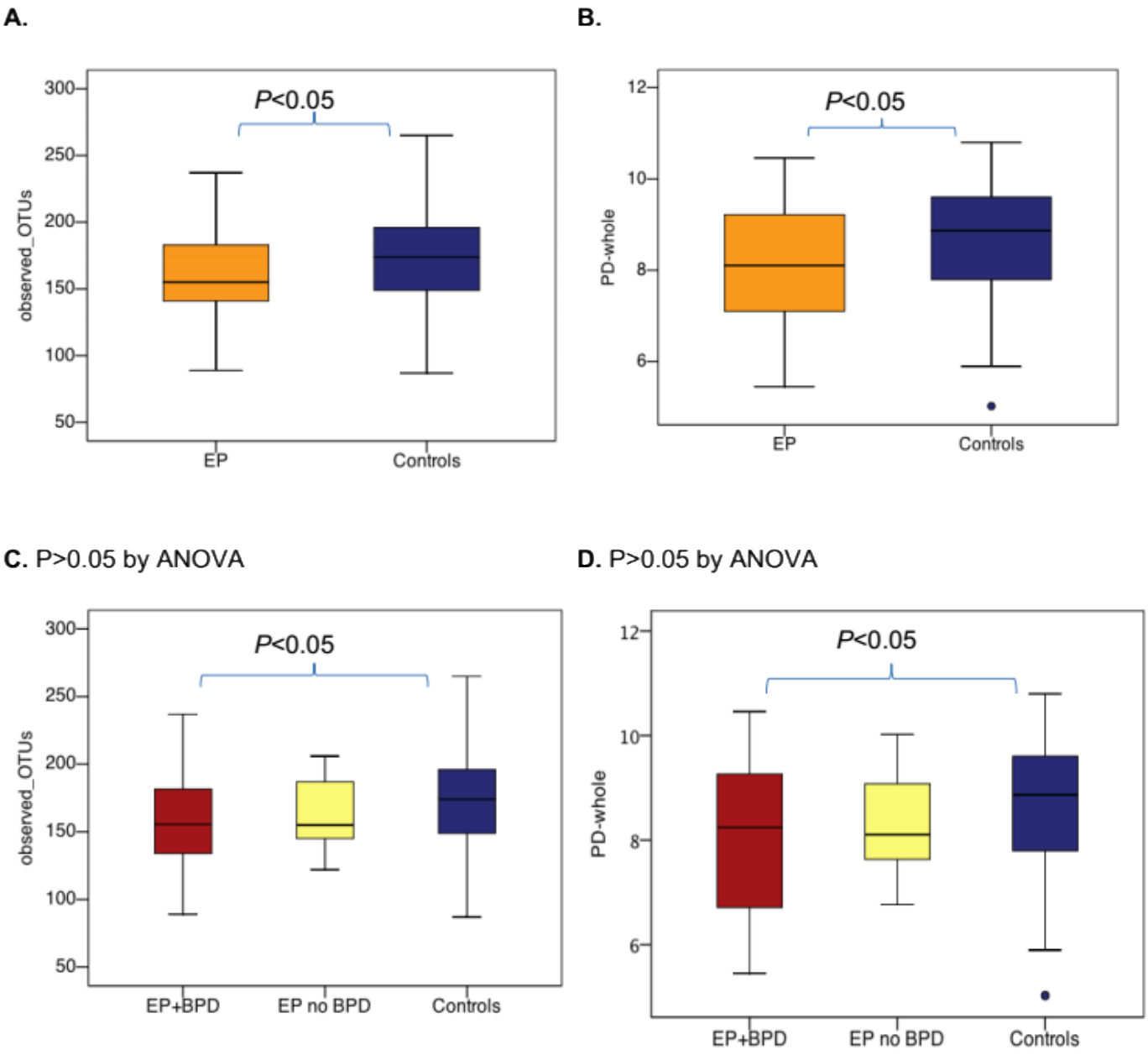
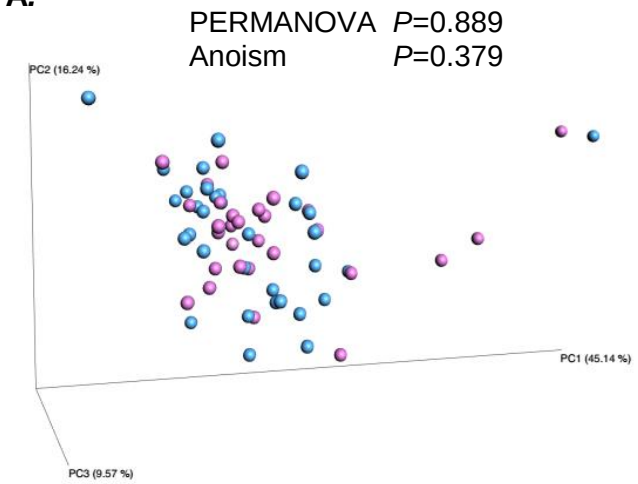


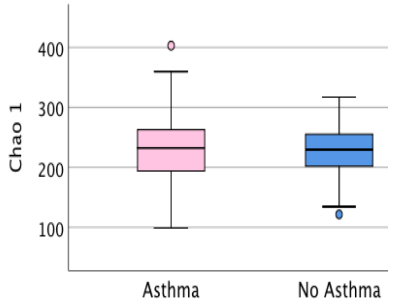
Figure E2: Richness and alpha diversity of the airway microbial communities measured by total number of observed OTU **(A &C)** and PD whole tree **(B&D)** respectively, both were significantly lower in the whole extremely pre-term born (EP)(orange) group (*P*<0.05, T-test) **(A&B)**, the BPD group had significantly less diverse microbial community compared to controls (*P*<0.05, T-test) **(C&D)**, EP+BPD extremely preterm born group with Bronchopulmonary Dysplasia history (n=28), EP no BPD: extremely preterm born group without BPD history (n=9), Controls (blue) (n=33)

Whole Cohort

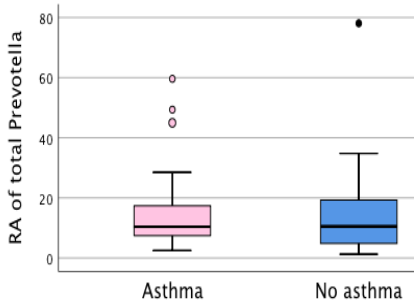
A.



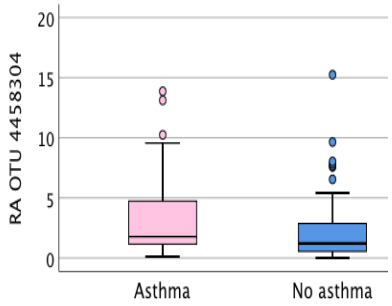
B. *P*=0.463[1]



C. *P*=0.496[2]

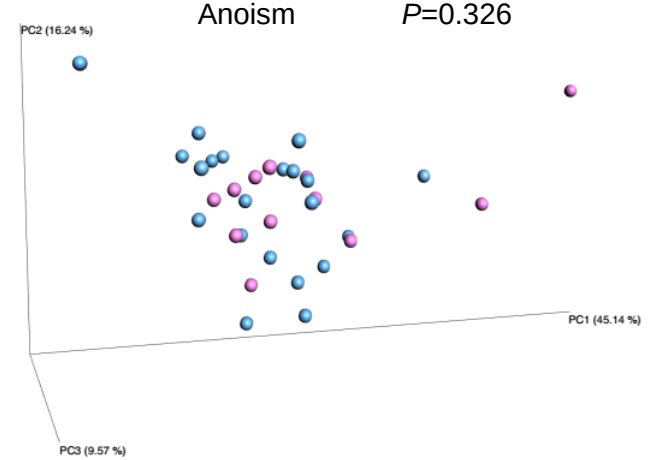


D. *P*=0.217[2]

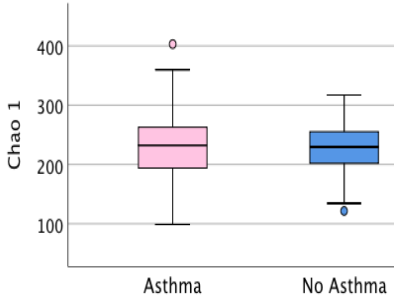


Control Group

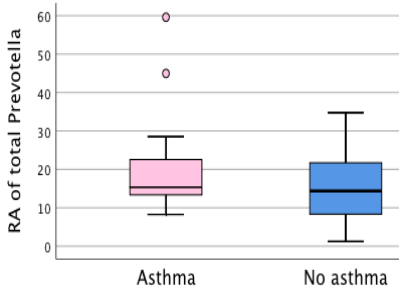
E.



F. *P*=0.859[1]



G. *P*=0.388[2]



H. *P*=0.367[2]

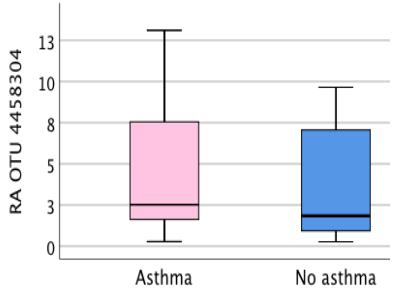


Figure E3: Comparison of weighted Unifrac β diversity index, Chao 1 α diversity index, genus *Prevotella* relative abundance (RA) and RA of OTU 4458304 identified as *Prevotella melaninogenica* in sputum samples from participants who were labelled with asthma and those who were not within the whole cohort (A, B, C and D respectively) and within our control group (E, F, G and H respectively), No significant differences were detected. 1: *P*-values by T test, 2: *P*-values by Mann-Whitney Test

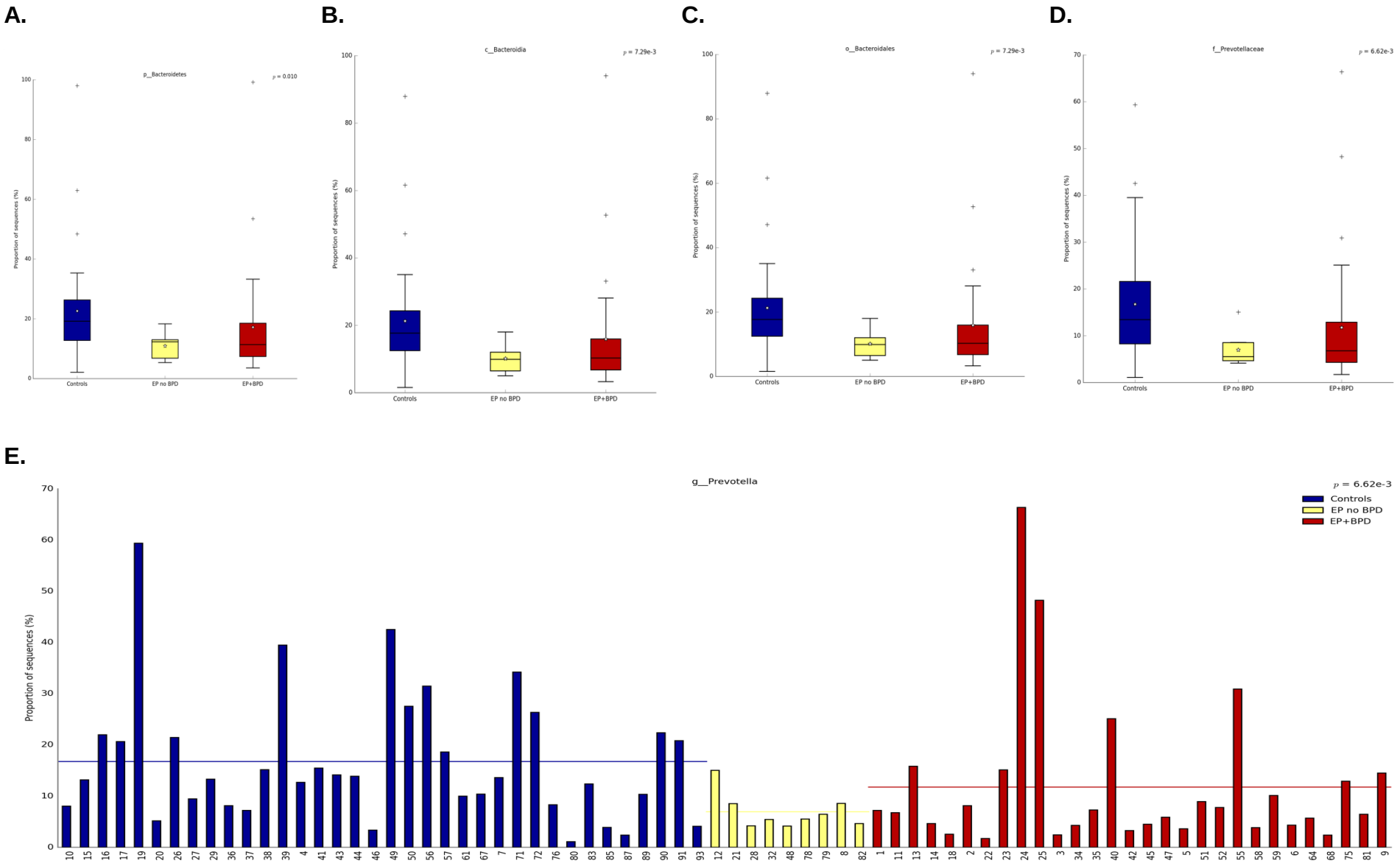


Figure E4: Comparison of relative abundancies of **A**. Phylum *Bacteroidetes*, **B.** Class: *Bacteroidia*, **C.** Order *Bacteroidales*, **D.** Family: *Prevotellaceae*, **E.** Genus*: Prevotella* across the three study groups the Extremely Preterm (EP) group with neonatal Bronchopulmonary Dyspalsia (BPD) (red), EP group without BPD (yellow) and the control group (blue); sample size:29, 9 and 36 respectively.

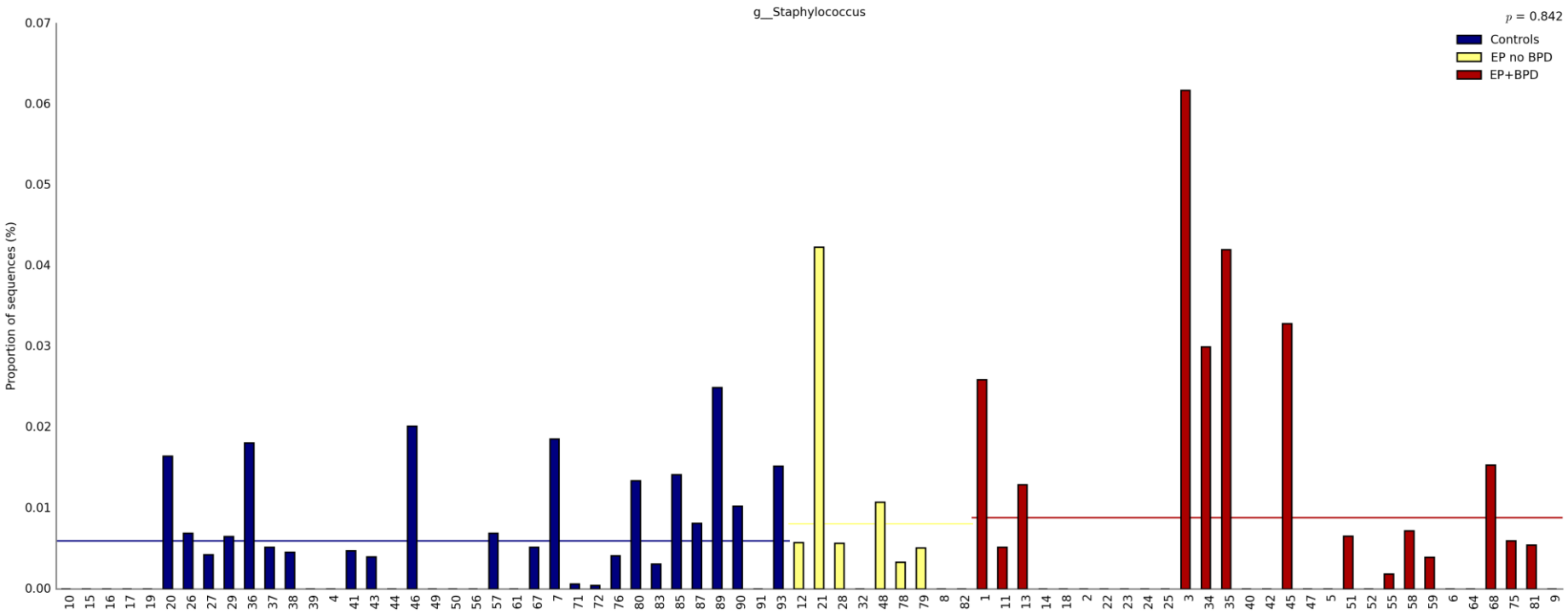


Figure E5: Comparison of the relative abundance of *Staphylococcus* species between study groups EP+BPD: extremely preterm born group with Bronchopulmonary Dysplasia (BPD) history (n=29), EP no BPD: extremely preterm born group without BPD history (n=9) (yellow), Controls (n=36) (blue).

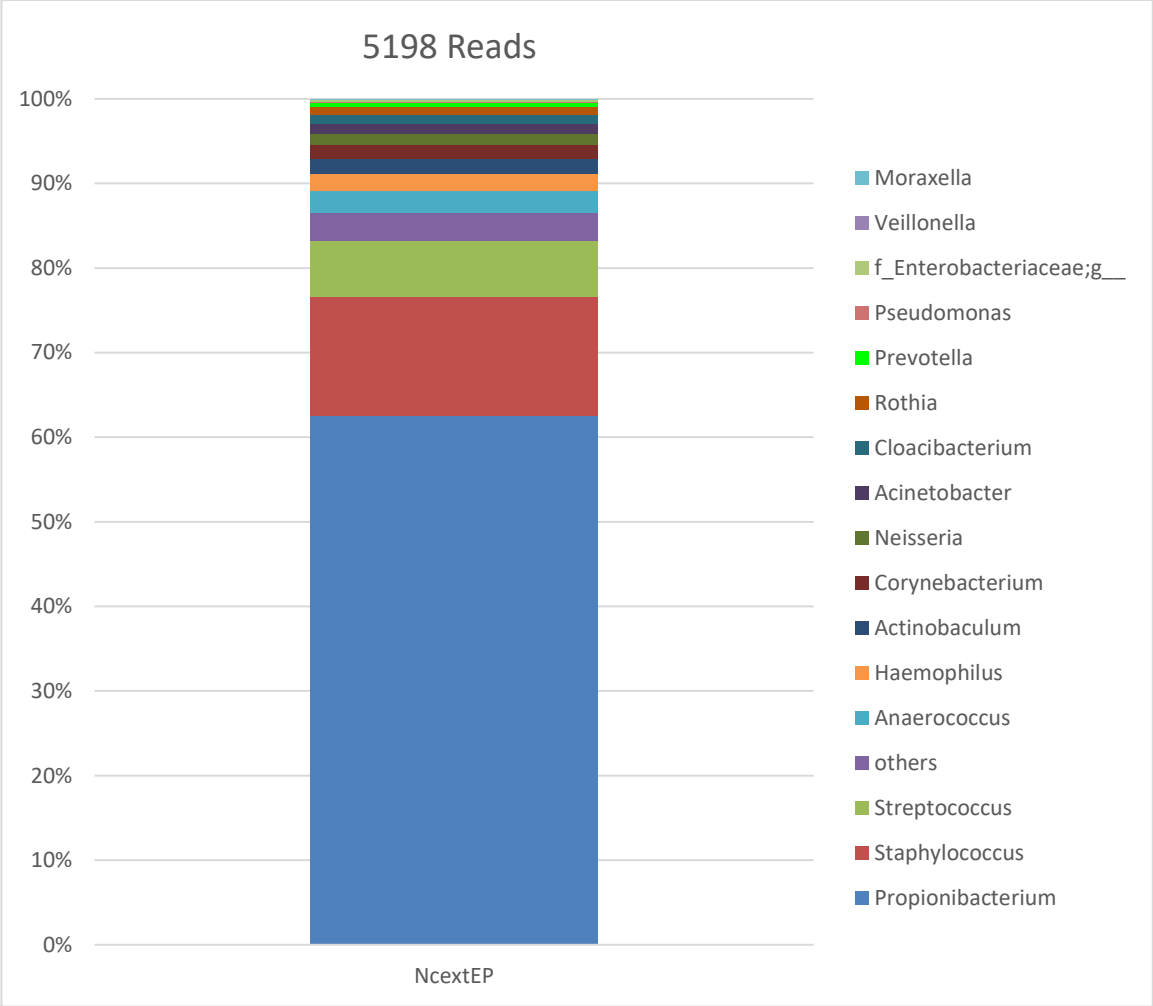


Figure E6: Microbiome profile of the extraction negative control of the saline used for sputum induction (composed of Nebusal 7% by Forest©, and Sodium chloride 0.9% w/v BP by B. Braun© and water of injection BP by B. Braun©)

References

1. Nolan T, Hands RE, Ogunkolade W, Bustin SA. SPUD: a quantitative PCR assay for the detection of inhibitors in nucleic acid preparations. Anal Biochem 2006; 351: 308-310.
2. Comeau AM, Douglas GM, Langille MG. Microbiome Helper: a Custom and Streamlined Workflow for Microbiome Research. mSystems 2017; 2.
3. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 2014; 30: 614-620.
4. A. G. FASTX-Toolkit: FASTQ/A short-reads pre-processing tools. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory 2009.
5. Rognes T, Flouri T, Nichols B, Quince C, Mahe F. VSEARCH: a versatile open source tool for metagenomics. PeerJ 2016; 18.
6. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 2010; 7: 335-336.
7. Parks DH, Tyson GW, Hugenholtz P, Beiko RG. STAMP: statistical analysis of taxonomic and functional profiles. Bioinformatics 2014; 30: 3123-3124.
8. da Silva RM, Teixeira PJ, Moreira Jda S. The clinical utility of induced sputum for the diagnosis of bacterial community-acquired pneumonia in HIV-infected patients: a prospective cross-sectional study. Braz J Infect Dis 2006; 10: 89-93.
9. Kralik P, Ricchi M. A Basic Guide to Real Time PCR in Microbial Diagnostics: Definitions, Parameters, and Everything. Front Microbiol 2017; 8: 108.
10. Lalkhen AG, McCluskey A. Clinical tests: sensitivity and specificity. Continuing Education in Anaesthesia Critical Care & Pain 2008; 8: 221-223.