Supplementary material

Committee names and approval numbers

The research was approved by the following ethics committees (approval number is given in brackets);

Sir Charles Gairdner Group Human Research ethics committee (2008-147)

Hunter New England Research Ethics Committee (08/11/19/3.03)

The Prince Charles Hospital Metro North Hospital and Health Services (HREC/08/QPCH/4)

Exclusion criteria

Participants with an FEV_1 <40% predicted, current smokers, ex-smokers who had ceased smoking in the previous year and those with a recent (past four weeks) exacerbation or respiratory infection were excluded. Those with significant smoking related air-space disease (ex-smokers >10 pack year history and DLCO/VA <70% predicted OR smoking history >10 pack years and exhaled carbon monoxide >10ppm) were also excluded.

Smoking assessment

A smoking history was taken and smoking pack-years determined. Participants underwent exhaled carbon monoxide (eCO) measurements, determined by electrochemical detection with a Smokerlyzer® (Bedfont Scientific Ltd, Kent, UK; detection limit of 1ppb). All included participants had an eCO of less than 10ppm confirming their non-smoking status [33].

DNA preparation

Total DNA was extracted from selected sputum (QIAamp DNA Mini Kit Qiagen, Hilden, Germany) and quantitated (Quant-iT™ PicoGreen dsDNA Assay Kit Invitrogen, Carlsbad, CA, USA).

PCR of 16S rRNA and amplicon pyrosequencing

The V8 and V9 regions of the 16S rRNA gene were amplified using a two-step PCR assay and sequenced as described. In the first step, template DNA was amplified using the primers 803Fmix and 1392wR. Each 50µL PCR reaction contained 2µL of sputum template DNA at 10ng/µL, 5µL of 10X buffer (Invitrogen, Carlsbad, CA, USA), 1µL of 10mM dNTP mix (Invitrogen), 1.5µL BSA (New England BioLabs, Ipswich, MA, USA), 2µL 50mM MgCl (Invitrogen), 1µL of each 10uM primer (803Fmix forward, 1392wR reverse), 1 unit (0.2µL) of Taq Polymerase (Invitrogen) and 36.3µL of nuclease free water. Cycling conditions were one cycle of 95°C for 3 min, followed by 35 cycles of 95°C for 30 s, 55°C for 45 s and 72°C for 90 s followed by a final extension of 72°C for 10 min. Single PCR products were confirmed for amplification using a standard SYBR green real-time PCR with a dissociation curve run on the ABI 7500 Real Time PCR System (Applied Biosystems, Foster City, USA). To prepare amplicons for sequencing, a second round of 10 cycles of PCR under the same conditions was conducted using fusion primers containing barcoded 454 adaptor sequences ligated to the 803Fmix and 1392wR primers [34]. Amplicons were sequenced from the reverse primer using the GS-FLX platform at the Australian Centre for Ecogenomics.

Bioinformatics and statistics for 16S amplicon pyrosequencing

Amplicon sequences underwent read quality filtering and error correction using Acacia (ver. 1.50) [35]. Libraries were normalized to 1050 sequences and sequences clustered into operational taxonomic units (OTUs) at 97% sequence similarity threshold [36] and classified against Greengenes (ver. Aug 2012) [37] using BLAST to produce community profiles for each library (PMID: 2231712).

Pyrosequencing bioinformatics and statistics

Amplicon sequences were processed as described previously[12]. Heatmaps from the normalized OTU table were produced using the R package g-plots [13]. Alpha-diversity was

calculated using rarefaction and the Shannon diversity index normalized at a depth of 1,000 sequences in QIIME [14]. Beta-diversity analysis was performed using FastUniFrac [15]. Principal component analysis (PCA) plots were produced in R and the relationships between the microbial community data and metadata were analyzed by permutational multivariate analysis of variance with 10,000 permutations using the R package vegan [16].

Real-time PCR bacterial confirmation

The presence of *Tropheryma whipplei* and *H. influenzae* in sputum samples was determined using real-time PCR. Detection of *T. whipplei* was similar to that previously described[38,39]. PCR was performed with 1µL of extracted DNA, RealMasterMix Probe ROX (5 Prime, Deutschland, Germany), 250nM forward and reverse primers (GeneWorks, Thebarton, SA, Australia) and 500nM probe (GeneWorks) labelled with the 5' reporter dye carboxyfluorescein (FAM) and the 3' quencher dye carboxytetramethylrhodamine (TAMRA) (total PCR volume 12.5µL). Amplification and detection were performed with an ABI 7500 Real Time PCR System (Applied Biosystems) under the following conditions: 2 min at 95°C and 40 cycles of 15 s at 95°C and 1 min at 60°C.

Detection of *H. influenzae* was carried out using Qiagen's Microbial DNA qPCR Assay for *H. influenzae* according to the manufacturer's instructions. Amplification and detection were performed with an ABI 7500 Real Time PCR System (Applied Biosystems). The presence of bacteria in sputum samples were reported as positive if the average cycle threshold (Ct) values were <38 whilst samples with average Ct values ≥ 38 were considered negative.

Clinical data analysis

Results are reported as mean (SD) or median (quartile 1, quartile 3) unless indicated. Analysis was performed using the two-sample Wilcoxon rank-sum test and the Kruskal-Wallis test was used for more than two groups. Fisher's exact test was used for categorical data.

Correlations between clinical data, inflammatory cell data and bacteria OTUs were examined using Spearman's rank correlation coefficient.

Results

Supplementary Table 1: Shannon and Simpson indices by study centre location (city/state). Data are median (q1, q3).

City/State	Newcastle, New South Wales	Brisbane 1, Queensland	Brisbane 2, Queensland	Perth, Western Australia	p
Number	11	5	10	4	
Shannon index	3.8 (3.4, 4.2)	4.1 (3.5, 4.2)	2.9 (1.1, 3.8)	4.0 (2.4, 4.3)	0.317
Simpson index	0.8 (0.8, 0.9)	0.9 (0.9, 0.9)	0.7 (0.3, 0.9)	0.9 (0.6, 0.9)	0.329

Supplementary Table 2: Clinical and inflammatory cell parameters for the additional 16 participants with severe neutrophilic asthma used for real-time PCR confirmation of *H. influenzae*.

N	16
Age, mean (SD)	61 (15)
Male, n (%)	7 (78%)
Atopic, n (%)	13 (81%)
Previous smoker, n (%)	7 (44%)
Smoking pack years, median (q1,q3)	4.5 (3.7,21.0)
FEV_1 % predicted, mean (SD)	72.0 (21.0)
FEV_1/FVC, mean (SD)	71.0 (12.0)
ACQ6, mean (SD)	2.2 (1.1)
Taking ICS, n (%)	15 (100%)

ICS dose, μg, median (q1,q3)	1800 (800,2000)
Total cell count, x 10^6/mL, median (q1,q3)	11.2 (4.5,11.8)
Viability, %, median (q1,q3)	85.0 (75.0,96.0)
Neutrophils, %, median (q1,q3)	75.9 (69.1,88.1)
Eosinophils, %, median (q1,q3)	0.7 (0.4,1.3)
Macrophages, %, median (q1,q3)	19.9 (9.1,30.2)
Lymphocytes, %, median (q1,q3)	0.3 (0.0,0.9)
Columnar epithelial cells, %, median (q1,q3)	0.0 (0.0,1.1)
Squamous cells, %, median (q1,q3)	3.5 (0.9,7.1)

ACQ6: asthma control questionnaire 6 question format, FEV_1: forced expiratory volume in 1 second, FVC: forced vital capacity, ICS: inhaled corticosteroids, SD: standard deviation, q1,q3: quartile 1, quartile 3.