Supplementary data file

Rapid detection of emerging pathogens and the loss of microbial diversity associated with severe lung disease in cystic fibrosis

William G. Flight,[1,2,3] Ann Smith,[4] Christopher Paisey,[4] Julian R. Marchesi,[4,5] Matthew J. Bull, Phillip J. Norville, [4] Ken J. Mutton,[2,6] A. Kevin Webb,[1,2] Rowland J. Bright-Thomas,[1,2] Andrew M. Jones[1,2] and Eshwar Mahenthiralingam[4]*

Affiliations:
[1] University Hospital of South Manchester NHS Foundation Trust
[2] Institute of Inflammation & Repair, University of Manchester
[3] Oxford University Hospitals NHS Trust
[4] Organisms and Environment Division, School of Biosciences, Cardiff University
[5] Department of Hepatology & Gastroenterology, St. Mary's Hospital, Imperial College London
[6] Central Manchester University Hospitals NHS Foundation Trust

***Corresponding author:** Prof. Eshwar Mahenthiralingam, Organisms and Environment Division, School of Biosciences, Cardiff University, Main Building, Room 0.23, Museum Avenue, Cardiff CF10 3AT, United Kingdom
Tel. +44 (0)29 20875875; Fax. +44 (0)29 20874305;
Email: MahenthiralingamE@cardiff.ac.uk

Supplementary Figures

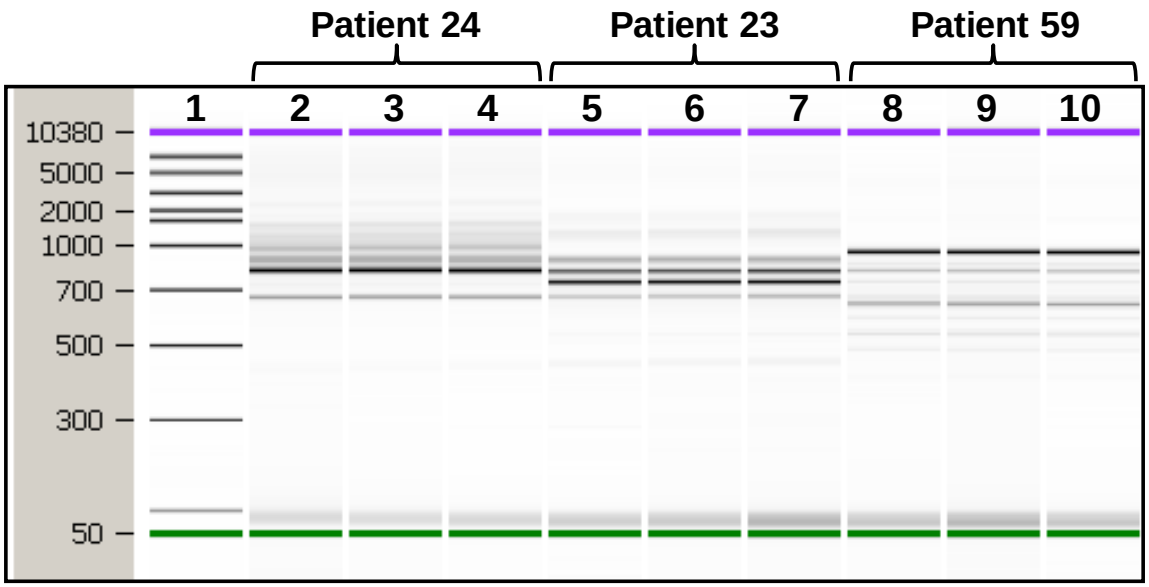

Figure S1. Reproducibility of RISA profiling from total sputum DNA. RISA PCR was repeated on the same sample and the ITS amplicons examined on a single microfluidics chip. Diversity profiles amplified on three separate occasions for patient 24 (lanes 2 to 4), patient 23 (lanes 5 to 7) and patient 59 (lanes 8 to 9) are shown, with molecular size markers run in lane 1 (size indicated in bp).

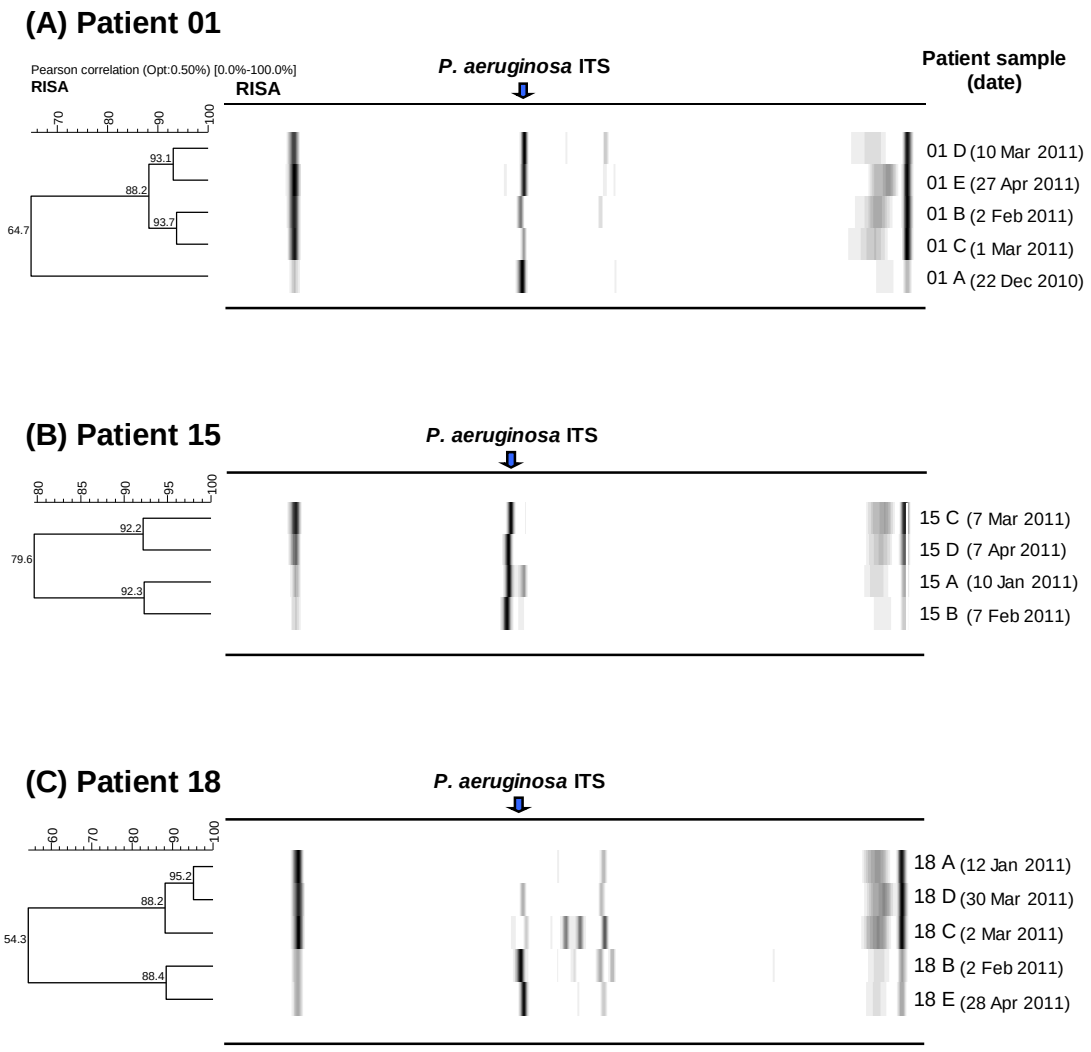

Figure S2: Cluster analysis of RISA bacterial diversity profiles amplified from sequential sputum samples. The clustered RISA diversity profiles for sequential samples for 3 *P.* aeruginosa culture positive patients are shown in each respective panel. The sample number and date are provided on the right and the 753 bp ITS amplicon correlating to that predicted for *P. aeruginosa* is shown. The percentage similarity of the diversity profiles is shown on the scale bar and for each node of the similarity dendogram.

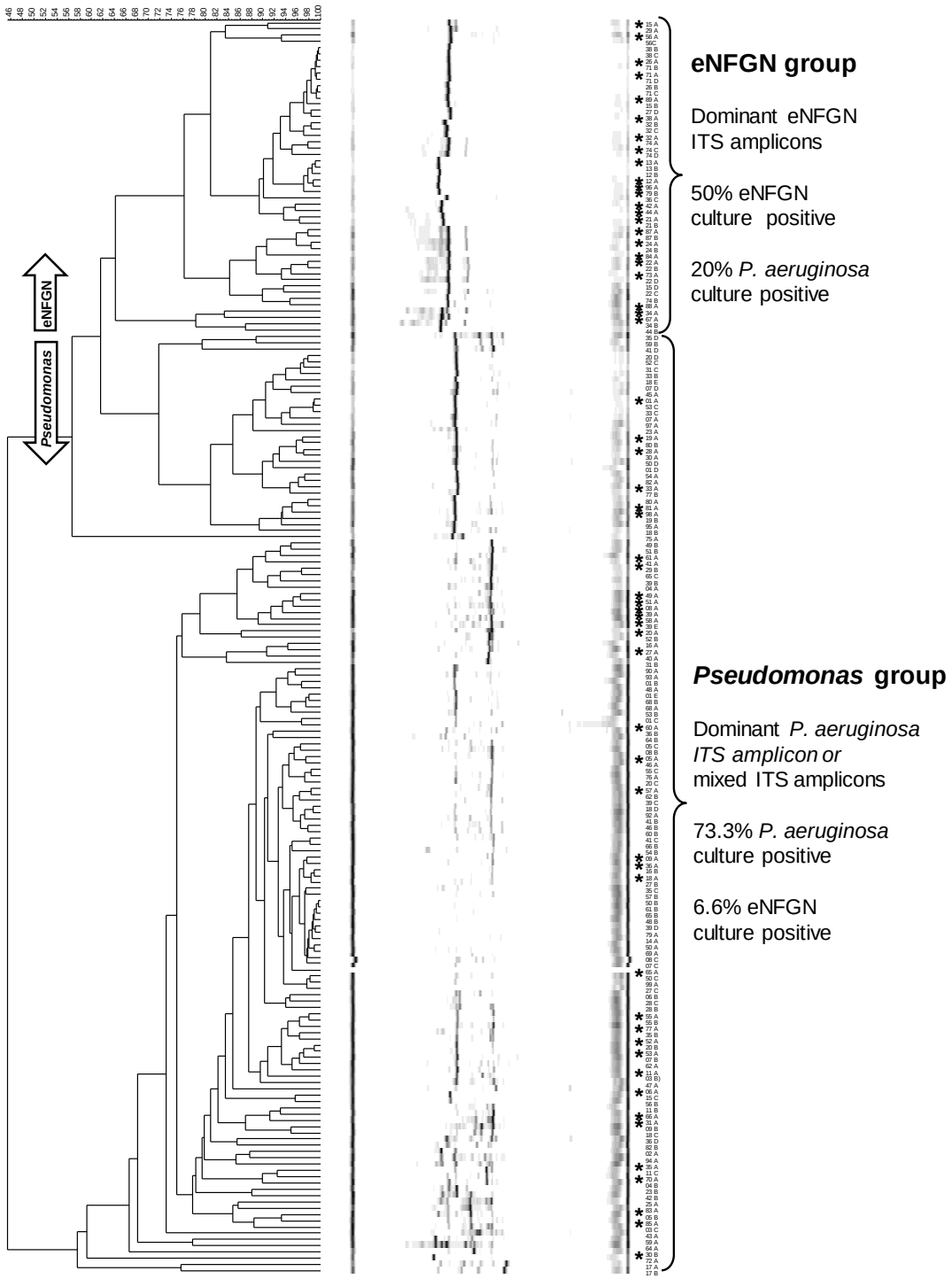

Figure S3: Cluster analysis of RISA bacterial diversity profiles for the study cohort. The RISA profiles amplified from all 200 sputum samples for the 93 patient study cohort were subjected to cluster analysis as described. The two profile groups most correlated to the presence of specific cultured pathogens (eNFGNs and *Pseudomonas*) are shown by the brackets. The percentage similarity of the diversity profiles is shown on the scale bar and the 60 samples selected for pyrosequencing analysis are indicated by the asterisks adjacent to the sample numbers on the right.

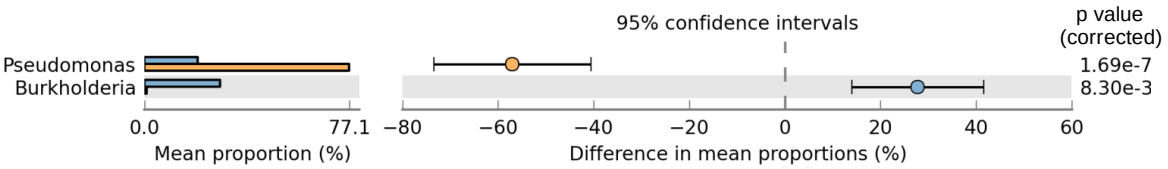

Figure S4. Significant correlation of *Pseudomonas* and *Burkholderia* 16S rRNA reads with RISA profile clusters. Microbial diversity within the *Pseudomonas* and eNFGN RISA profile groups (see Figure 2) was examined for statistically significant bacterial abundance correlations (t-test with Bonferroni correction). 16S rRNA gene reads for *Pseudomonas* were proportionally more dominant in the *Pseudomonas* group (mean = 77%) compared to the eNFGN group (mean = 20%). In contrast, the proportion of reads for *Burkholderia* was significantly greater in the eNFGN group (mean = 28.4%) than the *Pseudomonas* group (mean = 0.73%).

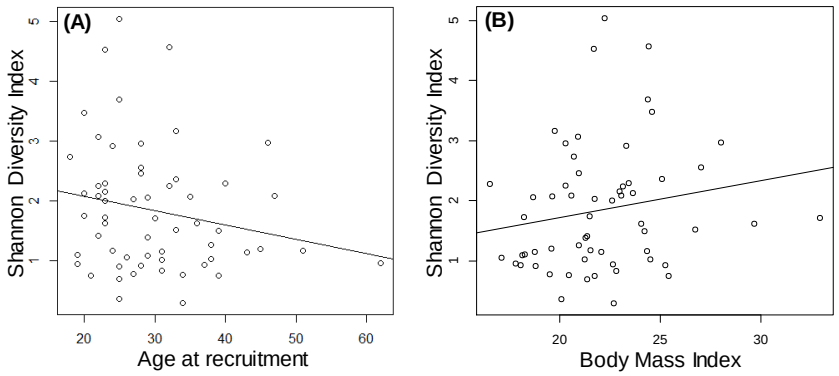

Figure S5. Correlation of patient age at recruitment and BMI with sputum microbial diversity. The nonparametric Shannon Diversity index for 59 patients was plotted against the age at recruitment (panel A) and BMI (panel B). Trends towards lower microbial diversity with increasing age and reduced BMI, respectively, were observed, however in each case the correlation was not significant.

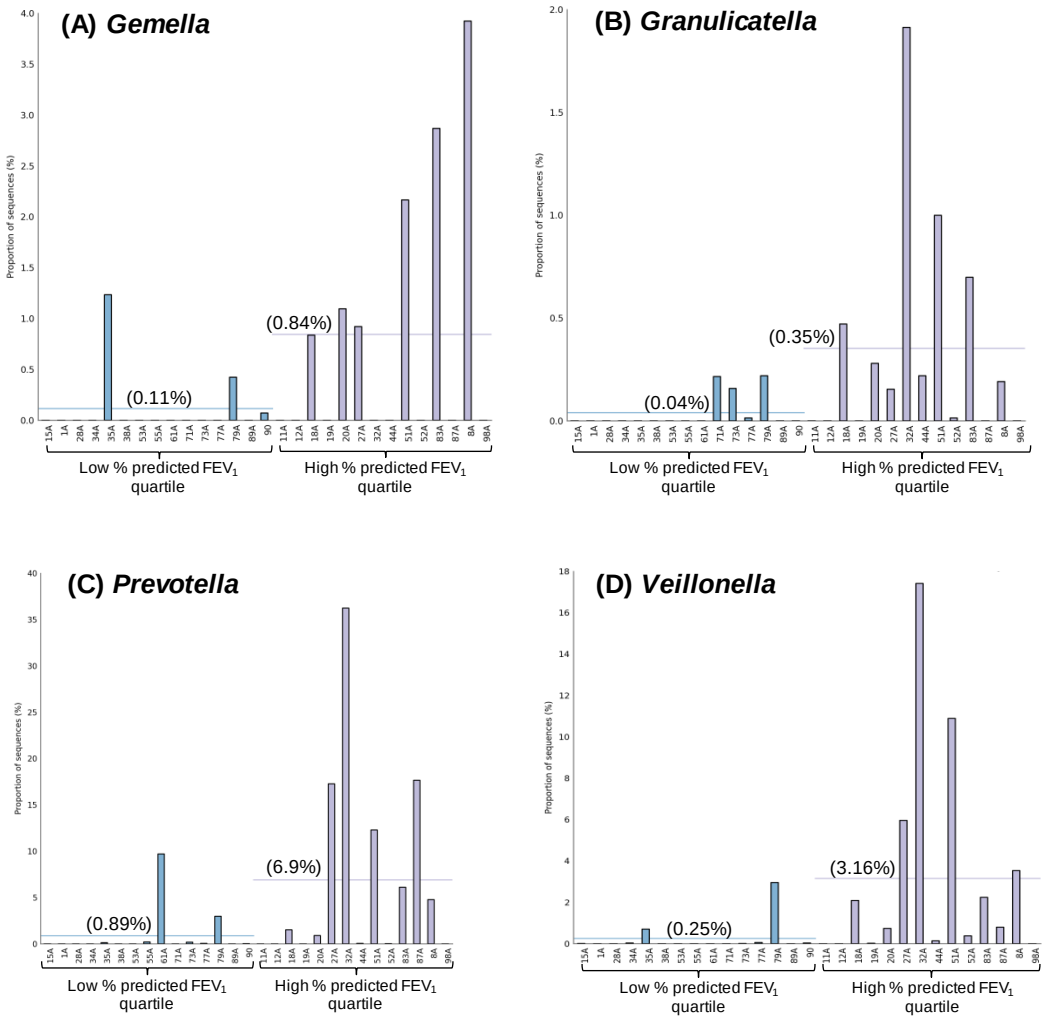

Figure S6. Bacterial genera with reduced relative abundance in patients with low lung function. Bacterial diversity for patients with lung function in the lowest quartile (n = 15) was compared by ANOVA to that for individuals with top quartile % predicted FEV_1 (n = 14). Prior to correction, four genera were observed to be less abundant ($p < 0.05$) in the samples from patients with low lung function: (A) *Gemella*, (B) *Granulicatella*, (C) *Prevotella* and (D) *Veillonella*. The relative abundance for each patient and the low and high quartile lung function groups are plotted for each genera in the panels above.