Expanded Materials and Methods

Clinical Samples

This observational study of 14 adult CF patients over 12 months was undertaken with ethical approval from Southampton and South West Hampshire Research Ethics Committee (06/Q1704/26). Eligible subjects were aged 18 or over and had experienced a minimum of 3 pulmonary exacerbations (CFPE) requiring antibiotic treatment in the prior 12 months. Samples from these patients were previously used in an analysis of core and satellite bacterial taxa [10].

Patient clinical details are summarised in Supplementary Table 1. The start of a CFPE was defined by the clinician's decision to initiate antibiotic therapy for deteriorating clinical status, broadly based on a range of factors described previously [15]. In turn, the end of CFPE was defined by the decision to cease antibiotic therapy due to stabilisation or improvement in signs and symptoms. A total of 39 episodes of CFPE were experienced by these patients over the year. The number of CFPE ranged from zero (Patients 3 and 8) to 6 (Patient 9) with a mean of 2.8 ± 1.4. In total, 17 different antibiotics combinations were used with five "elective" antibiotic courses given for reasons other than worsening of respiratory symptoms (Supplementary Table 1).

One sputum sample was obtained from each patient at approximately monthly intervals. The majority of samples (65%) were collected at least 21 days prior to or 21 days after stop of antibiotics for CFPE (timings shown in Supplementary Table 2). However, it was not always possible to obtain samples within these limits and 10% of samples were collected that lay 14-20 days prior or post CFPE, 12% of samples were collected 7-13 days prior or post CFPE, 5% of samples were collected 0-6 days prior or post CFPE and 8% of samples were taken during antibiotic courses for CFPE. All sputum samples were stored at 4°C (immediately after expectoration), shipped at 4°C (in accordance with the sputum handling guidelines [16]) and stored at -80°C prior to processing or cultured by the UK Health Protection Agency (HPA) South East, in accordance with HPA standard operating procedures.

Clinical data

Clinical measures (FEV_1 and FEV_6, temperature, and patient reported outcome (PRO) scores ("breathlessness", "cough severity", "sputum production", and "general well-being") were recorded using visual analogue scores (VAS) by means of a symptom sheet. VAS were recorded in millimetres from 0 (no symptoms) to 100 (worst symptoms).

Culture-independent analysis

Exclusion of DNA from non-viable cells in sputa via cross-linking using propidium monoazide [17,18] and subsequent nucleic acid extraction were performed as described previously [19]. Specifically, PMA was dissolved in 20% dimethyl sulfoxide to create a stock concentration of 20 mmol/L with this added to 500 μL of heat-killed cells to give a final concentration of 50 μmol/L. Sputum samples and bacterial suspensions were transferred to 24-well, flat-bottomed, cell culture cluster plates (Corning, Corning, NY) for exposure to light. Following an incubation period of 30 min in the dark with occasional mixing, samples were light exposed using LED Active Blue equipment (IB - Applied Science, Barcelona, Spain). After photo-induced cross-linking, cells were transferred to 1.5-mL microfuge tubes and pelleted at 5000 × g for 5 min prior to DNA isolation.

Nucleic acid extractions were performed on 100 μL portions of sputum. Guanidinium thiocyanate–EDTA–sarkosyl (500 μL) and PBS (500 μL), pH 8.0, were added to samples. Cell disruption was achieved using a Fastprep-24 Instrument (MP Biomedicals Europe, Illkirch, France) 6.5 m/s, 60 s, followed by incubation at 90 °C for 1 min and −20 °C for 5 min. Cell debris was pelleted by centrifugation at 12 000 × g for 2 min at 4 °C. Supernatant was transferred to a fresh microfuge tube. NaCl (to a final concentration of 0.5 mol/L and polyethylene glycol (to a final concentration of 15%)were added and DNA precipitated at 4 °C for 30 min. DNA was pelleted by centrifugation at 12 000 × g for 2min at 4 °C and resuspended in 300 μL of sterile distilled water. Samples were heated at 90 °C for 30 s and vortexed. Phenol/chloroform (1:1) (300 μL) was added, and samples were vortexed for 20 s before centrifugation at 12 000 × g at 4 °C for 3min. The upper

phase was then transferred to a fresh microfuge tube. Total DNA was then precipitated by the addition of an equal volume of isopropanol, a 0.1-volume 10 mol/L ammonium acetate, and 1 μL of GenElute linear polyacrylamide (Sigma-Aldrich, Gillingham, UK) and incubated at −20 °C for 25 min. DNA was pelleted by centrifugation at 12 000 × g at 4 °C for 5 min. Pelleted DNA was then washed 3 times in 70% ethanol, dried, and resuspended in 50 μL of sterile distilled water. DNA extracts were quantified using the Picodrop Microlitre Spectrophotometer (GRI, Braintree, UK).

PCR amplification and T-RFLP profiling were carried out as previously described [1]. Specifically, the oligonucleotide primers used to amplify a region of the 16S rRNA gene for members of the domain Bacteria, 8f700 and 926r (5'-CCG TCA ATT CAT TTG AGT TT-3'), were described previously (24). Primer 8f700 was labelled at the 5' end with IRD700; primer 926r was unlabeled. PCR mixtures comprised 1x PCR buffer, 1.5 mM MgCl2, each deoxynucleoside triphosphate at a concentration of 0.2 mM, each primer at a concentration of 0.2 mM, and 1 U of RED*Taq* DNA polymerase (Sigma-Aldrich) in a final volume of 50 μl. An initial denaturation step of 94°C for 2 min was followed by 32 cycles of denaturation at 94°C for 1 min, annealing at 56°C for 1 min, and extension at 72°C for 2 min, with a final extension step at 72°C for 10 min. Amplification was carried out by using a GeneAmp PCR System 2400 (Perkin-Elmer), with PCR products for T-RFLP analysis stored at -20°C after verification on Tris-acetate-EDTA–agarose gels as described above. PCR products (ca. 20 ng) were digested by using the restriction endonuclease CfoI (Roche, Lewes, United Kingdom) for 3 h at 37°C with the reaction buffer supplied by the manufacturer. All restriction endonuclease digestions were carried out to complete digestion as shown by comparing PCR products after various digestion incubation times (data not shown). The restriction endonuclease was inactivated by heating at 90°C for 20 min. An approximately 0.7-μg portion of T-RFLP PCR products was separated by length by using a 25-cm SequagelXR denaturing acrylamide gel (National Diagnostics) prepared in accordance with the manufacturer's instructions with the addition of 8.3 M urea and 10% (final concentration) formamide and by using an IR2 automated DNA sequencer (LI-COR Biosciences) at 55°C and 1,200 V.

T-RFLP gels were analyzed by using Phoretix one-dimensional advanced software, version 5.10 (Nonlinear Dynamics, Newcastle upon Tyne, United Kingdom). The sizes of the bands resolved by T-RFLP were determined by comparing their relative positions with two sets of size markers, one set that formed bands at 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, and 1,000 bases (microSTEP 15a [700 nm]) and one set that formed bands at 155, 209, 214, 238, and 364 bases (microSTEP custom GR [700 nm]). Both sets of size markers were obtained from Microzone (Lewes, United Kingdom). In addition, this software was also used to determine the volume of each band (with band volume being the product of the area over which a band was detected and the intensity of signal recorded over that area). Band volume was expressed as a percentage of the total volume of bands detected in a given electrophoretic profile. T-RFLP bands were resolved over the region between 50 and 958 bases. No bands shorter than 50 bases in length were recorded, as they were in the region susceptible to high levels of signal stemming from the IR tag on either unattached or nonused primer 8f700IR. In this study, the threshold used to detect bands was 0.01% of the total signal between the 50- and 958-base region.

MapSort (Wisconsin Package, version 10.3; Accelrys), was used to identify *Hha*I restriction sites *in silico* in all of the 2,137 16S rRNA gene sequences from clone library construction, and to predict T-RFLP band lengths for each clone. Species or genus level identities were assigned to T-RF bands from sputum sample profiles on the basis of these predicted band lengths. 16S rRNA gene clone data generated previously from a subset of the samples analysed here was published previously [6].

Statistical analysis

Bacterial species richness (the number of bacterial species in a sample) was inferred from the number of T-RF bands present in a sample as previously described [20]. The cumulative bacterial taxon richness (the total number of different bacterial taxa identified after each successive time point) was assessed using taxa-time relationships (TTR), which describe how the taxon

richness of a community increases with the length of time over which the community is monitored [20,21]. TTR are modelled with the power law equation, $S = cT^w$, where S is the cumulative number of observed taxa over time T, c is an empirically derived taxon- and time-specific constant, and w is the slope of the fitted curve or temporal scaling exponent (the rate of observation of new bacterial taxa over the course of sampling); therefore, increasing values of w reflect greater turnover rates. Power law regression coefficients of determination (r^2) and significance (P) values were calculated using Minitab software (version 14.20; University Park, PA).

To examine similarities and differences in bacterial community composition over time, we employed distance-decay relationships (DDR). DDR describes how similarity in taxa composition between two communities varies with a measure of distance (here, the time interval between pairs of sample collections) [22]. Similarities and differences in community composition were determined using the Bray-Curtis quantitative index of similarity using the PAST program available from the University of Oslo website link (http://folk.uio.no/ohammer/past) run by Øyvind Hammer. Power law coefficients of determination (r^2) and significance (P) were calculated using Minitab software (version 14.20, Minitab, University Park, PA, USA).

Pairwise Pearson's correlation analyses were performed to assess whether clinical or microbiological factors showed significant relationships i) across patients, and ii) across samples independent of the patient from which they came. All Pearson's correlation analysis and Bonferroni corrections where applied were carried out in SPSS version 15.0.1 (SPSS Inc., Chicago, IL). Correlations across the patients (n = 14) assessed the following parameters – clinical: age, genotype, gender, BMI, number of CFPE, mean FEV_1, mean FEV_6 and mean temperature over the 12 month study period, mean individual and summed PRO scores over the year, sample status, antibiotics and long-term medications administered over the year and time, - microbiological: averaged diagnostic microbiology results, mean species number, mean Bray-Curtis similarity index, TTR and SDR slope, mean rank abundance slope and mean relative abundance of *P. aeruginosa* in the T-RFLP profile. Correlations across all samples (n = 168) assessed relationships between data

corresponding to each of the 12 sampling points for clinical parameters - FEV_1, FEV_6, temperature, individual and sum of the PRO scores, sample status, and antibiotics and long-term medications administered; and microbiological variables - diagnostic microbiology results, species number, Bray-Curtis similarity index, rank abundance slope of the community and relative abundance of *P. aeruginosa* in the T-RFLP profile.

To test the robustness of Pearson's correlations, Mantel tests were performed. Specifically, Mantel proximity matrices were constructed for community similarity (Bray-Curtis index of similarity) and then FEV_1, FEV_6, temperature, summed PRO scores, sample status (whether samples were taken during clinical stability or how long (days) before/after CFPE), antibiotics (or combinations thereof) administered, temporal distance between samples, species number, rank abundance slope (see below), and relative abundance of *P. aeruginosa* within in the T-RFLP profile, performed as previously described [10] using XLSTAT 2010 (Addinsoft, France). The results of these Mantel tests corroborated the findings from the Pearson's correlations, with no evidence of auto-correlation was observed ($P > 0.05$).

Principal component analysis (PCA), used to identify and quantify differences in microbial community structures, was carried out with software developed in-house, using the Python programming language (version 2.7.2) with the Numpy package (version 1.5.1) as described previously (Louic S Vermeer, Gilbert O. Fruhwirth, Pahini Pandya, Tony Ng, and Andrew James Mason. NMR metabolomics of MTLn3E breast cancer cells identifies a role for CXCR4 in lipid and choline regulation *J. Proteome Res* In press). Information of the development of these packages is available at www.python.org and www.numpy.scipy.org, respectively. The principal components were calculated using eigenvalue decomposition of mean-centred and auto-scaled T-RFLP profile band percentage values. Source code for in-house software can be provided on request.