Article (refereed) - postprint

Rogers, Geraint B.; **Cuthbertson, Leah**; Hoffman, Lucas R.; Wing, Peter A.C.; Pope, Christopher; **Hooftman, Danny A.P.**; Lilley, Andrew K.; **Oliver, Anna**; Carroll, Mary P.; Bruce , Kenneth D.; **van der Gast, Christopher J.**.
2013 Reducing bias in bacterial community analysis of lower respiratory infections. *ISME Journal*, 7 (4). 10.1038/ismej.2012.145

Towards unbiased bacterial community analysis in lower respiratory infections

Geraint B. Rogers[1], Leah Cuthbertson[2], Lucas R. Hoffman[3], Peter A. C. Wing[4], Christopher Pope[3], Danny A. P. Hooftman[2], Andrew K. Lilley[1], Anna Oliver[2], Mary P. Carroll[4], Kenneth D. Bruce[1], Christopher J. van der Gast[2*]

[1]*Institute of Pharmaceutical Science, Molecular Microbiology Research Laboratory, King's College London, London, SE1 9NH, UK*

[2]*NERC Centre for Ecology and Hydrology, Wallingford, OX10 8BB, UK*

[3]*Departments of Pediatrics and Microbiology, University of Washington, Box 356320, HSB K328A, Seattle, WA 98105, USA*

[4]*Cystic Fibrosis Unit, Southampton University Hospitals NHS Trust, Southampton, SO16 6YD, UK*

* For correspondence: E-mail cjvdg@ceh.ac.uk

Running title: Impact of PMA on CF bacterial community analysis

Subject category: Microbial population and community ecology

The sequence data have been submitted to the Science Research Associates (SRA) database, hosted by the NCBI, under the study accession number SRA051980.

Supplementary information: Figure S1; Tables S1, S2, and S3

Abstract

High-throughput pyrosequencing and quantitative PCR analysis offer greatly improved accuracy and depth of characterisation of lower respiratory infections. However, such approaches suffer from an inability to distinguish between DNA derived from viable and non-viable bacteria. This discrimination represents an important step in characterising microbial communities, particularly in contexts with poor clearance of material or high antimicrobial stress, as non-viable bacteria and extracellular DNA can contribute significantly to analyses. Pre-treatment of samples with propidium monoazide (PMA) is an effective approach to non-viable cell exclusion (NVCE). However, the impact of NVCE on microbial community characteristics (abundance, diversity, composition and structure) is not known. Here, we use adult cystic fibrosis (CF) sputum samples as a paradigm. The effects of PMA treatment on CF sputum bacterial community characteristics, as analysed by pyrosequencing and enumeration by species-specific (*Pseudomonas aeruginosa*) and total bacterial Q-PCR, were assessed. At the local community level, abundances of both total bacteria and of *P. aeruginosa* were significantly lower in PMA treated sample portions. Meta-analysis indicated no overall significant differences in diversity; however, PMA treatment resulted in a significant alteration in local community membership in all cases. In contrast, at the metacommunity level, PMA treatment resulted in an increase in community evenness, driven by an increase in diversity, predominately representing rare community members. Importantly, PMA treatment facilitated the detection of both recognised and emerging CF pathogens, significantly influencing 'core' and 'satellite' taxa group membership. Our findings suggest failure to implement NVCE may result in skewed bacterial community analyses.

Keywords: propidium monoazide, pyrosequencing, 16S ribosomal RNA gene, metacommunity, quantitative PCR, cystic fibrosis, commonness and rarity

Introduction

The first key step of analysing a microbial community is accurately defining the bacterial taxa present. From this, a range of ecological insights can follow in terms of understanding and predicting the impact and response of communities to perturbation. Such approaches have been of importance for some time when assessing the response(s) of communities present in natural environments (Ager et al 2010). More recently, awareness has grown of the benefit of applying these techniques to both normal and pathogenic bacterial communities associated with the human host. In humans, mixed species infections are commonplace (Brogden et al 2005); and, defining the roles played by bacteria in chronic lower respiratory infections is particularly important. Taking cystic fibrosis (CF) as a paradigm, the lung damage that results from such infections drives respiratory failure and death in the overwhelming majority of cases (Davis et al 1996). Understanding the ecological patterns and processes may allow improvement in therapy. However, before being able to do so, it is important to determine accurately the community characteristics of CF airways.

16S rRNA gene pyrosequencing and quantitative PCR analysis offer greatly improved accuracy and depth of characterisation of microbial communities. However, whilst avoiding the selective biases associated with culture-based microbiology, culture-independent approaches will distort to some degree the characteristics of the systems they are being used to describe. For example, standard PCR-based analyses are unable to distinguish between DNA derived from viable and non-viable bacterial cells. In the context of CF lung infections, the combination of poor clearance of airway secretions (Döring and Gulbins 2009) and exposure of bacterial cells to a range of antimicrobial challenges, both in the form of host immune response (Elizur et al 2008) and antibiotic therapy (Ratjen 2001), results in substantial non-viable populations (Rogers et al 2005, Rogers et al 2008, Rogers et al 2010). As such, the potential contribution of DNA, either from the extracellular environment or from non-viable cells, to microbial community characterisation is considerable.

Failing to limit analysis to DNA from viable bacteria could have a number of effects on community characterisation. These include overestimating viable bacterial density; distorting relative viable species abundance, and the masking of less abundant species by dominant species. Further, the impact on community dynamics from antibiotic treatment in this context is less likely to be observed when DNA from bacteria rendered non-viable is not prevented from contributing to molecular analysis. Pre-treatment of samples with propidium monoazide (PMA) is an effective approach to non-viable cell exclusion (NVCE) (Nocker and Camper 2006). In brief, PMA intercalates into double-stranded nucleic acids. On exposure to bright light, cross-linking occurs (Nocker and Camper 2006), preventing the DNA from acting as a PCR

template (Nocker and Camper 2006, Nocker et al 2007a, Nocker et al 2007b). Importantly though, PMA is highly membrane impermeant (Nocker and Camper 2006). This means that whilst PMA is readily able to penetrate dead bacterial cells whose structural integrity has been lost, PMA is excluded from viable cells (Nocker and Camper 2009).

PMA treatment for this purpose has been shown to be effective in a range of contexts (Bae and Wuertz 2009, Kralik et al 2010, Nam et al 2011, Nocker et al 2007a, Taskin et al 2011) including the assessment of microbiota present in CF airways samples (Rogers et al 2008, Rogers et al 2010, Stressmann et al 2011). Further, this strategy has been used previously in conjunction with bacterial pyrosequencing in waste water treatment systems (Nocker et al 2010). Whilst the benefits of NVCE are leading to its increasing deployment, its impact on microbial community data has not previously been determined.

In this study, we detailed the application of NVCE to profiling the bacterial communities present in spontaneously expectorated sputum samples collected from adult CF patients. The aim of this study was to determine the effects of PMA treatment on bacterial community characteristics (abundance, diversity, composition and structure), as analysed by 16S rRNA gene pyrosequencing, total bacterial Q-PCR enumeration, and species-specific Q-PCR enumeration of *Pseudomonas aeruginosa,* a clinically important pathogen in CF airways disease (Ballmann et al 1998, Kosorok et al 2001). Comparisons were made first between paired PMA treated and non-treated samples (i.e., the local community level), before focusing on core and satellite groups of bacterial taxa within the entire collection of PMA and non-PMA treated samples (i.e., the metacommunity).

Materials and methods

Thirty spontaneously expectorated sputum samples were collected from adult CF patients attending the Adult CF Clinic at the Southampton General Hospital, with full ethical approval (Southampton and South West Hampshire Research Ethics Committee (06/Q1704/26)). All patients were judged to be clinically stable by treating physicians at the time of sampling. Sputum samples were collected in sterile containers, placed on ice, and transported to the microbiology laboratory within 60 min. Samples were then frozen and stored at -80°C prior to analysis. Each sample was mixed by pipetting and divided into two equal portions, one of which was treated with PMA prior to DNA extraction.

PMA cross-linking

PMA was dissolved in 20% dimethyl sulfoxide to create a stock concentration of 20 mmol/L with this added to samples to give a final concentration of 50 µmol/L. 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.

DNA extraction

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 was 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).

Quantitative PCR

All quantitative (Q)PCR analyses were performed in triplicate. Total bacterial density was determined using a Taqman assay, in which a 466 bp fragment of the 16S ribosomal RNA gene was amplified, as described previously (Nadkarni et al 2002). *P. aeruginosa* density was determined using a Taqman assay which amplified a 65 bp fragment of the regA gene, as described previously (Rogers et al 2010, Shannon et al 2007). Details of the relevant primers and probes used are shown in Supplementary Table S1.

Pyrosequencing

Bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP) was performed as described previously (using Gray28F 5'-TTTGATCNTGGCTCAG-3' and Gray519r 5'-GTNTTACNGCGGCKGCTG-3'). Initial generation of the sequencing library involved a one-step PCR of 30 cycles, using a mixture of Hot Start and HotStar high fidelity taq polymerase, as described previously (Dowd et al 2008). Tag-encoded FLX amplicon pyrosequencing analyses utilized Roche 454 FLX instrument with Titanium reagents, titanium procedures performed at the Research and Testing Laboratory (Lubbock, TX) using RTL protocols (www.researchandtesting.com).

Sequence processing

Following sequencing, all failed sequence reads, low quality sequence ends and tags and primers were removed. Sequences with ambiguous base calls, sequences with homopolymers > 6bp were removed. Further, any non-bacterial ribosome sequences and chimeras were removed using Black Box Chimera Check software (B2C2) (Gontcharova et al 2010) as has been described previously (Dowd et al 2008). To determine the identity of bacteria in the remaining sequences, sequences were de-noised, assembled into OTU clusters at 97% identity, and queried using a distributed .N*Et al*gorithm that utilizes Blastn+ (KrakenBLAST www.krakenblast.com) against a database of high quality 16S bacterial sequences. Using a .NET and C# analysis pipeline the resulting BLASTn+ outputs were compiled and data reduction analysis performed as described previously (Dowd et al 2008).

Bacterial identification

Based upon the above BLASTn+ derived sequence identity (percent of total length query sequence which aligns with a given database sequence) the bacteria were classified at the appropriate taxonomic levels based upon the following criteria. Sequences with identity scores, to known or well characterized 16S sequences, greater than 97% identity (<3% divergence) were resolved at the species level, between 95% and 97% at the genus level, between 90% and 95% at the family and between 85% and 90% at the order level , 80 and 85% at the class and 77% to 80% at phyla. After resolution based upon these parameters, the percentage of each bacterial identity was individually analyzed for each sample providing relative abundance information within and among the individual samples based upon relative numbers of reads within each (Dowd et al 2008).

Statistical analyses

To determine whether sample sizes were large enough to effectively assess the diversity of bacteria in each of the sputum samples taken, we used finite (without replacement) single sample rarefaction calculated using the Species Diversity and Richness package (version 4.1.2, Pisces Conservation, Lymington, UK).

We used meta-analyses to summarize the effect sizes of all 30 samples using Hedges' *d* (Borenstein et al 2009), treating the different samples as independent of each other and the two treatments (PMA and non-PMA) as separate groups within samples. Per sample, the effect size (i.e., standardized mean difference (δ)) among treatments was expressed as:

$$\delta = \frac{\overline{X_1} - \overline{X_2}}{\left(\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}\right)}$$

In which X_1 and X_2 are the sample means in the two groups, n_1 and n_2 the sample sizes and S_1 and S_2 the standard deviations within groups. The Standard Error of Mean per sample (SEM_δ) was expressed as:

$$SEM_\delta = \frac{\sqrt{V}}{\sqrt{\frac{(n_1 + n_2)}{2}}}$$

$$\text{in which } V = \left(\frac{n_1+n_2}{n_1 \times n_2} + \frac{\delta^2}{2(n_1+n_2)}\right)$$

Subsequently, we calculated the overall average effect size *M*, weighted by variance of sample *i* (V_i) and the Standard Error of Mean of *M* (SEM_M) as:

$$M = \frac{\sum_{i=1}^{30} \frac{1}{V_i} \times \delta_i}{\sum_{i=1}^{30} \frac{1}{V_i}}$$

$$SEM_M = \frac{\sqrt{\frac{1}{\sum_{i=1}^{30} \frac{1}{Vi}}}}{\sqrt{30}}$$

Three complementary measurements of diversity were used to compare bacterial diversity between paired non-treated and PMA treated samples, as previously described (Edwards et al 2001). The indices used were species richness (*S**), Shannon-Wiener index (*H*'), and Simpson's index (*D*). Differences in *S**, *H*' and *D* (converted to the intuitive reciprocal form (1/*D*)) were computed using the re-sampling randomization method of Solow (1993). It is known that pair wise comparisons will be affected by large differences in sample size (*N*) (Gihring et al 2012). Therefore, an Excel macro-program was written to apply Solow's method to the pair wise comparison of each of the three parameters (*S**, *H*' and *D*) with a uniform sub-sample size (*n* = 1637). The re-sampling was repeated 1000 times and the mean similarity coefficients and standard deviation were taken. The value of *n* = 1637 was chosen to match to the smallest sample size. For metacommunity level analyses, the smallest sample size was chosen for each pair wise comparison.

The community compositions from paired samples were compared pair wise using the Sørensen similarity coefficient calculated from the numbers of shared species. To overcome differences in sample size (Gihring et al 2012), the similarity coefficients were calculated using an Excel macro-program written to resample each sample species list with a uniform sub-sample size (*n* = 1637) and calculate the similarity coefficients (Solow 1993). The re-sampling was repeated 1000 times and the mean similarity coefficients and standard deviation were taken. For metacommunity level analyses, the smallest sample size was chosen for each pair wise comparison.

Poisson distribution tests were carried out according to the method described by Krebs (1999) and applied as previously described (van der Gast et al 2011). Regression analysis, coefficients of determination (r^2), residuals and significance (*P*) were calculated using Minitab software (version 14.2, Minitab, University Park, PA, USA).

Results and discussion

The effect of PMA treatment on abundance

Q-PCR based enumeration using a conserved 16S rRNA gene primer pair and probe was performed on each of the thirty spontaneously expectorated sputum samples to determine the abundance of bacteria in paired PMA treated and non-treated sputum samples (Figure 1A). Mean bacterial abundance across non-PMA samples was $9.05 \times 10^8 \pm 1.06 \times 10^9$ CFU ml^{-1} (standard deviation of the mean, $n = 30$), ranging between 3.27×10^6 and 4.03×10^9 CFU ml^{-1}. In contrast, bacterial abundance across PMA treated samples was $8.91 \times 10^7 \pm 1.79 \times 10^8$ CFU ml^{-1} ($n = 30$), ranging between 1.22×10^5 and 7.50×10^8 CFU ml^{-1}. Further, the abundance of only *P. aeruginosa* in paired samples was determined through Q-PCR using primers and a probe specific for *P. aeruginosa* Toxin A synthesis regulating gene, *regA* (Figure 1C). The mean abundance of *P. aeruginosa* in non-PMA samples was $1.57 \times 10^8 \pm 1.82 \times 10^8$ CFU ml^{-1} compared to $2.50 \times 10^7 \pm 1.82 \times 10^7$ CFU ml^{-1} in the PMA treated samples, ranging from 4.47×10^3 to 6.15×10^8 and $1.64 \times 10^2 \pm 1.49 \times 10^8$ CFU ml^{-1} for non-PMA and PMA treated samples, respectively.

A growing number of studies have observed that bacterial communities are highly variable between CF patients (e.g. Harris et al 2007, Rogers et al 2005, van der Gast et al 2011). To cover the large variation among samples we used meta-analysis (Borenstein et al 2009) to integrate effect sizes among sputum samples, i.e., the difference between with and without prior PMA treatment. Meta-analysis is a statistical synthesis of different individual studies, used in medicine to test among different clinical trials (Borenstein et al 2009) or in ecology among different experimental studies (Marvier et al 2007). Importantly, the number of replicates within individual studies and the variation among these replications are used as weighting of the means. Here, we translated this into using the replications within the same sample as a study and so all paired samples as independent studies, hence, providing us with a powerful tool to study between sample effects rather than simply comparing means of individual samples by, for example, paired t-tests.

Using the meta-analysis approach to assess the effects of PMA treatment on total bacterial abundance and *P. aeruginosa* abundance demonstrated significant and strong negative effects on abundance for all paired samples (Figure 1B&D). Furthermore, the overall effect of PMA treatment demonstrated a significant reduction in abundance (Figure 3). This would indicate that only DNA from extant viable cells was included in the analyses, excluding extracellular DNA and DNA from dead or damaged cells that would otherwise be incorporated into abundance measures using Q-PCR without prior PMA treatment.

The effects of PMA treatment on diversity and composition in local communities

Targeted 454 high-throughput pyrosequencing was used to assess the diversity and composition of bacterial communities within each of the treated and non-treated paired samples. A total of 386213 bacterial sequence reads (mean = 6436.9 ± 634.2 (*n* = 60)) identifying 154 genera and 267 taxa, were generated from all of the sputum samples combined (Supplementary Tables S2 and S3). Rarefaction curves were used to test whether sample sizes per sample were sufficiently large to compare enough of the bacterial diversity (Supplementary Figure S1); this analysis indicated that richness values for the bacterial taxa sampled from all local communities became asymptotic or were approaching asymptotic status, providing stable estimates of diversity. In this way, confirmation that sufficiently large samples had been collected and bacterial diversity was not under-sampled was obtained.

To avoid potential biases in comparisons of diversity between local communities due to varying number of sequences per sample (Gihring et al 2012), a randomised re-sampling method (with a uniform sub-sample size of *n* = 1637) using three indices of diversity (taxa richness (*S**), Shannon-Wiener index (*H*'), and Simpson's reciprocal index (1/*D*)) was employed (Supplementary Table S2). For brevity, the effects of PMA treatment on taxa richness are presented in Figure 2A, and the overall effects of PMA treatment on diversity (*S**, *H*' and 1/*D*) are given in Figure 3. Richness across individual samples regardless of treatment was highly variable, with mean values of 21.0 ± 15.1 and 18.8 ± 15.3, respectively. Meta-analysis of individual samples revealed both significant negative and positive effects on richness by PMA treatment (Figure 2B). Although there was no overall significant effect on richness, and hence diversity (confirmed using *H*' and 1/*D* measures of diversity (Figure 3)), the data here indicate that diversity can be either under- or over-estimated in local communities without prior PMA treatment.

Changes in community composition due to PMA treatment were also investigated. The Sørensen index of similarity with a randomised re-sampling method (uniform sub-sample size of *n* = 1637) revealed a mean change in similarity of 0.46 ± 0.23 in PMA treated samples when compared to paired non-PMA treated samples (Figure 2C). Meta-analysis showed significant negative effects (decreases in similarity) for all individual paired samples (Figure 2D); with an overall significant negative effect on community composition (Figure 3). Therefore, we propose that skewed observations of local community composition are reached without prior PMA treatment.

The effects of PMA treatment at the metacommunity level

At the metacommunity level, it has been established that a dichotomy exists within metacommunity species abundance distributions (SADs), consisting of 'regionally' common and locally abundant core species and infrequent rare satellite species groups (e.g. Dolan et al 2009, Magurran and Henderson 2003, Ulrich and Zalewski 2006, Unterseher et al 2011). Recently, we established that the core-satellite group approach can be applied to CF bacterial communities and that partitioning the two groups from a metacommunity revealed important aspects of SADs, which would otherwise be neglected without such a distinction (van der Gast et al 2011).

In the current study, we employed the same approach to focus on PMA treatment effects at the metacommunity level. To establish that the defined metacommunity, comprised of either the 30 PMA- or non-treated local communities, was coherent we first plotted the relationship between mean abundance of taxa and local community occupancy / persistence (Figure 4A&B). The expectation for a coherent metacommunity would be a 'regional' community comprised of trophically similar individuals and species each of which is perceived to exist in a series of local communities from similar habitats linked by rationale dispersal of potentially interacting species between or into those local communities (i.e. by dispersal from outside of the metacommunity) (Prosser et al 2007), resulting in a positive relationship between distribution among sites and abundance. Here we observed significant distribution-abundance relationships for both the PMA and non-PMA treated metacommunity (Figure 4A&B). Therefore, as observed previously, the commonness and rarity of bacterial taxa in the treated or non-treated metacommunity was found to be related to their permanence (local communities occupied) (Hanski 1982, Magurran and Henderson 2003, van der Gast et al 2011).

Next we objectively divided the treated and non-treated metacommunity SADs into core and satellite groups by decomposing the overall distribution using the ratio of variance to the mean abundance for each bacterial taxon (van der Gast et al 2011). The variance to mean ratio, or index of dispersion, is an index used to model whether species follow a Poisson distribution, falling between 2.5 and 97.5 % confidence limits of the χ^2 distribution (Krebs 1999). Plotting the indices of dispersion against persistence throughout the 30 samples without prior PMA treatment, 32 bacterial taxa were randomly distributed through space (i.e., those taxa that fall below the 2.5 % confidence limit line in Figure 4C). Bacterial taxa that occurred only in a single sputum sample were excluded from this analysis, as their dispersion in space would have no variance. However, those 49 taxa were classified into the satellite group (81 taxa in total) and the remaining 68 non-randomly distributed taxa classified as core group taxa, in the non-PMA treated metacommunity (see

309 Supplementary Table S3 for core and satellite identification). For the PMA treated metacommunity, the
310 satellite group was comprised of 163 taxa (with 112 taxa each occurring in only one sample) and 70 core
311 group taxa (Figure 4D).
312 Following the categorisation of the CF bacterial taxa into the two groups, the metacommunity SADs
313 with and without prior treatment were visualised as rank-abundance plots to examine differences in diversity
314 and metacommunity structure (Figure 5). The slope values for each plot were used as a descriptive statistic
315 of evenness, where a value of zero represents perfect evenness (Ager et al 2010). A change in community
316 structure was observed in the PMA treated metacommunity (b = -0.0137) when compared to the non-treated
317 metacommunity (b = -0.0244). Using the t-distribution test (Fowler et al 1998) allowed us to determine that
318 the slopes were significantly different (t = 12.86; $d.f.$ = 1,378; P < 0.0001); indicating an increase in
319 community structure evenness following PMA treatment. Furthermore, the changes in evenness were driven
320 by changes in richness, where 149 taxa were observed in the non-PMA treated metacommunity compared to
321 233 taxa in the PMA treated metacommunity (Figure 5). We also determined whether previously undetected
322 extant taxa, without prior PMA treatment, were either core (regionally common and locally abundant) or
323 satellite (infrequent and rare) members of the metacommunity. A total of 118 taxa were previously
324 undetected, of which 106 were categorised as satellite group members and 12 taxa as core group members.
325 Furthermore, a total of 34 taxa comprised of 31 satellite group and only 3 core group members, were not
326 detected after prior PMA treatment. Despite the inclusion of a small number of core taxa group members,
327 we could conclude that the majority of taxa, both newly detected and undetected after PMA treatment, were
328 predominately infrequent rare members of the metacommunity.
329 Subsequently, we examined to what extent this drove overall patterns of diversity and composition of
330 the metacommunity, by comparing mean richness and changes in composition for the whole
331 metacommunity, and the core and satellite groups between treatments (Figure 6A). Again to avoid potential
332 biases in comparisons of diversity and composition between samples due to varying mean numbers of
333 sequences, we used a randomised re-sampling method (with a uniform sub-sample size, using the lowest
334 number of sequences in each pair-wise comparison for the metacommunity (n = 8172), core (n = 6370), and
335 satellite (n = 214) taxa groups). Richness was significantly higher in the PMA treated metacommunity when
336 compared to the non-treated metacommunity, confirmed using meta-analysis (Figure 6C). When partitioned
337 into groups, PMA treatment did not lead to a significant change in core group richness, but a significant
338 change in richness was observed in the satellite group, again demonstrating that the majority of previously
339 undetected taxa belonged to the satellite group, which drove the changes in diversity of the overall

metacommunity (Figure 6A&C). Likewise, patterns similar to the observed changes in taxa richness were found for taxa composition, with overall significant changes in similarity in the PMA treated metacommunity (change of 0.23 ± 0.01) and the satellite group (0.63 ± 0.03). Conversely, the composition of the core group was not significantly affected by PMA treatment (0.03 ± 0.04) (Figure 6B&D).

Impact of PMA treatment on the detection of recognised and emergent CF pathogens

Analysis of the detection of recognised and emergent CF pathogens in the metacommunity indicated that *P. aeruginosa*, *Stenotrophomonas maltophilia*, and *Streptococcus pneumoniae* remained core group members regardless of PMA treatment. The most abundant taxon in the metacommunity, *P. aeruginosa*, was detected across the same 29 individual local communities regardless of prior treatment; accounting for 76.1% (178952 of 235119) total sequence reads without PMA, and 77.1% (116444 of 151094) reads with PMA. However, the number of local communities in which *S. maltophilia* was detected increased from 10 (35 reads) to 16 (112) with PMA. In contrast, the abundance and distribution of *S. pneumoniae* decreased from 14 (408 reads) to 7 (36). *Staphylococcus aureus* (2 samples) and the candidate pathogen *Streptococcus intermedius* (part of the *Streptococcus milleri* group [3 samples]), both initially classified as satellite group members, were not detected with prior PMA treatment. Interestingly, four previously undetected taxa were detected in the metacommunity following PMA treatment. The recognised CF pathogens, *Achromobacter xylosoxidans* and *Haemophilus influenzae* were found to belong to the satellite group, being observed in 3 and 1 local communities, respectively. *Mycobacterium* sp., some of which are regarded as pathogens in CF, was also found to belong to the satellite group, being detected in one local community. Conversely, members of the *Burkholderia* genus were found to belong to the core taxa group (7 samples and 83 reads) yet were undetected without prior PMA treatment (Supplementary Table S3).

The distribution of taxa in the metacommunity that are associated with the oral cavity according to the Human Oral Microbiome Database (Dewhirst et al 2010) was also investigated. These taxa were found to be widely distributed across both the non-PMA and PMA treated metacommunity SADs (and therefore throughout both the core and satellite groups) (see Supplementary Table S3).

Conclusions

The lower respiratory system is a challenging environment for bacterial colonisation and persistence. Bacterial populations are subject to pressures from the host immune system and a wide variety of antimicrobial drugs. Antibiotic treatment is the mainstay of CF therapy, particularly in response to acute

worsening of respiratory symptoms. In order to assess the effectiveness of such interventions, it is essential that numbers of viable bacteria, and any reduction achieved as a result of therapy, can be accurately quantified. Our data indicate that DNA from non-viable bacteria and accumulated extracellular DNA make a significant contribution to estimations of bacterial abundance unless PMA treatment is employed. As such, this suggests that the inclusion of a PMA treatment step is essential for accurate Q-PCR based enumeration.

As CF bacterial communities are known to contain a wide phylogenetic range of species, all of which can be expected to be affected to different degrees by antimicrobial therapy and immune response, the populations of those species are likely to be present in differing relative viable and non-viable proportions. Determining the impact of non-viable cell exclusion (NVCE) treatment on community characteristics was therefore vital in the present study. Our data clearly demonstrate that contribution of DNA from non-viable sources prevents the accurate characterisation of CF bacterial communities by pyrosequencing, and that NVCE approaches to limit this contribution is critical for community analysis.

We demonstrate that suppressing the contribution of DNA from non-viable sources is essential for accurate characterisation of lower airway bacterial communities, in part because this method further identifies bacterial taxa that would otherwise go undetected. Important from a clinical perspective is the detection of known CF pathogens (both core and satellite group members) only with NVCE, increasing the sensitivity of molecular analyses for these pathogens, and providing vital information for prognosis and treatment. Here we have used CF bacterial communities as a paradigm for lower respiratory infections in general. Our findings suggest that a failure to implement a NVCE approach to analysing microbial communities may result in skewed results. As this effect is due to the contribution of DNA from non-viable sources, such approaches are appropriate where communities are under antimicrobial stress, and where a characterisation of the viable bacterial community is sought.

Acknowledgments

This study was funded by the Anna Trust. CJvdG, LC, DAPH, and AO are funded by the UK Natural Environment Research Council. LRH is funded by the Cystic Fibrosis Foundation.

References

Ager D, Evans S, Li H, van der Gast CJ (2010). Anthropogenic disturbance affects the structure of bacterial communities. *Environ Microbiol* **12:** 670-678.

Bae SW, Wuertz S (2009). Discrimination of viable and dead fecal Bacteroidales bacteria by quantitative PCR with propidium monoazide. *Appl Environ Microbiol* **75:** 2940-2944.

Ballmann M, Rabsch P, von der Hardt H (1998). Long term follow up of changes in FEV1 and treatment intensity during *Pseudomonas aeruginosa* colonisation in patients with cystic fibrosis. *Thorax* **53:** 732-737.

Borenstein M, Hedges LV, Higgins JP, Rothstein HR (2009). *Introduction to Meta-Analysis*. Wiley & Sons: Chichester, UK

Brogden KA, Guthmiller JM, Taylor CE (2005). Human polymicrobial infections. *Lancet* **365:** 253-255.

Davis PB, Drumm M, Konstan MW (1996). Cystic fibrosis. *Am J Resp Crit Care Med* **154:** 1229-1256.

Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner ACR, Wen-Han Y *et al* (2010). The human oral microbiome. *J Bacteriol* **192:** 5002-5017.

Dolan JR, Ritchie ME, Tunin-Ley A, Pizay MD (2009). Dynamics of core and occasional species in the marine plankton: tintinnid ciliates in the north-west Mediterranean Sea. *J Biogeog* **36:** 887-895.

Döring G, Gulbins E (2009). Cystic fibrosis and innate immunity: how chloride channel mutations provoke lung disease. *Cell Microbiol* **11:** 208-216.

Dowd SE, Wolcott RD, Sun Y, McKeehan T, Smith E, Rhoads D (2008). Polymicrobial nature of chronic diabetic foot ulcer biofilm infections determined using bacterial tag encoded FLX amplicon pyrosequencing (bTEFAP). *PLOS One* **3:** e3326.

Edwards ML, Lilley AK, Timms-Wilson TH, Thompson IP, Cooper I (2001). Characterisation of the culturable heterotrophic bacterial community in a small eutrophic lake (Priest Pot). *FEMS Microbiol Ecol* **35:** 295-304.

Elizur A, Cannon CL, Ferkol TW (2008). Airway inflammation in cystic fibrosis. *Chest* **133:** 489-495.

Fowler J, Cohen L, Jarvis P (1998). *Pratical Statistics For Field Biologists*. John Wiley and Sons: Chichester, UK.

Gihring TM, Green SJ, Schadt CW (2012). Massively parallel rRNA gene sequencing exacerbates the potential for biased community diversity comparisons due to variable library sizes. *Environ Microbiol* **14:** 285-290.

Gontcharova V, Youn E, Sun Y, Wolcott RD, Dowd SE (2010). A comparison of bacterial composition in diabetic ulcers and contralateral intact skin. *Open Microbiol J* **4:** 8-19.

Hanski I (1982). Dynamics of regional distribution: the core and sateillte species hypothesis. *Oikos* **38:** 210-221.

Harris JK, De Groote MA, Sagel SD, Zemanick ET, Kapsner R, Penvari C *et al* (2007). Molecular identification of bacteria in bronchoalveolar lavage fluid from children with cystic fibrosis. *Proc Natl Acad Sci USA* **104:** 20529-20533.

Kosorok MR, Zeng L, West SEH, Rock MJ, Splaingard ML, Laxova A *et al* (2001). Acceleration of lung disease in children with cystic fibrosis after *Pseudomonas aeruginosa* acquisition. *Pediatr Pulmonol* **32:** 277-287.

Kralik P, Nocker A, Pavlik I (2010). *Mycobacterium avium* subsp *paratuberculosis* viability determination using F57 quantitative PCR in combination with propidium monoazide treatment. *Int J Food Microbiol* **141:** S80-S86.

Krebs CJ (1999). *Ecological Methodology*, 2nd edn. Harper and Row: New York, USA.

Magurran AE, Henderson PA (2003). Explaining the excess of rare species in natural species abundance distributions. *Nature* **422:** 714-716.

Marvier M, McCreedy C, Regetz J, Kareiva P (2007). A meta-analysis of effects of Bt cotton and maize on non-target invertebrates. *Science* **316:** 1475-1477.

Nadkarni MA, Martin FE, Jacques NA, Hunter N (2002). Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology* **148:** 257-266.

Nam S, Kwon S, Kim MJ, Chae JC, Maeng PJ, Park JG *et al* (2011). Selective detection of viable *Helicobacter pylori* using ethidium monoazide or propidium monoazide in combination with real-time polymerase chain reaction. *Microbiol Immunol* **55:** 841-846.

Nocker A, Camper AK (2006). Selective removal of DNA from dead cells of mixed bacterial communities by use of ethidium monoazide. *Appl Environ Microbiol* **72:** 1997-2004.

Nocker A, Camper AK (2009). Novel approaches toward preferential detection of viable cells using nucleic acid amplification techniqiues. *FEMS Microbiol Lett* **291:** 137-142.

Nocker A, Sossa-Fernandez P, Burr MD, Camper AK (2007a). Use of propidium monoazide for live/dead distinction in microbial ecology. *Appl Environ Microbiol* **73:** 5111-5117.

Nocker A, Sossa KE, Camper AK (2007b). Molecular monitoring of disinfection efficacy using propidium monoazide in combination with quantitative PCR. *J Microbiol Methods* **70:** 252-260.

Nocker A, Richter-Heitmann T, Montijn R, Schuren F, Kort R (2010). Discrimination between live and dead cells in bacterial communities from environmental water samples analyzed by 454 pyrosequencing. *Int Microbiol* **13:** 59-65.

Prosser JI, Bohannan BJM, Curtis TP, Ellis RJ, Firestone MK, Freckleton RP *et al* (2007). The role of ecological theory in microbial ecology. *Nature Rev Microbiol* **5:** 384-392.

Ratjen F (2001). Changes in strategies for optimal antibacterial therapy in cystic fibrosis. *Int J Antimicrob Agents* **17:** 93-96.

Rogers GB, Carroll MP, Serisier DJ, Hockey PM, Kehagia V, Jones GR *et al* (2005). Bacterial activity in cystic fibrosis lung infections. *Respir Res* **6:** 49-60.

Rogers GB, Stressmann FA, Koller G, Daniels T, Carroll MP, Bruce KD (2008). Assessing the diagnostic importance of nonviable bacterial cells in respiratory infections. *Diagn Microbiol Infect Dis* **62:** 133-141.

Rogers GB, Marsh P, Stressmann FA, Allen CE, Daniels TW, Carroll MP *et al* (2010). The exclusion of dead bacterial cells is essential for accurate molecular analysis of clinical samples. *J Clin Microbiol Infect* **16:** 1656-1658.

Shannon KE, Lee DY, Trevors JT, Beaudette LA (2007). Application of real-time quantitative PCR for the detection of selected bacterial pathogens during municipal wastewater treatment. *Sci Total Environ* **382:** 121-129.

Solow AR (1993). A simple test for change in community structure. *J Anim Ecol* **62:** 191-193.

Stressmann FA, Rogers GB, Marsh P, Lilley AK, Daniels TWV, Carroll MP *et al* (2011). Does bacterial density in cystic fibrosis sputum increase prior to pulmonary exacerbation? *J Cyst Fibros* **10:** 357-365.

Taskin B, Gozen AG, Duran M (2011). Selective quantification of viable *Escherichia coli* bacteria in biosolids by quantitative PCR with propidium monoazide modification. *Appl Environ Microbiol* **77:** 4329-4335.

Ulrich W, Zalewski M (2006). Abundance and co-occurrence patterns of core and satellite species of ground beetles on small lake islands. *Oikos* **114:** 338-348.

Unterseher M, Jumponen A, Öpik M, Tedersoo L, Moora M, Dormann CF *et al* (2011). Species abundance distributions and richness estimations in fungal metagenomics - lessons learned from community ecology. *Mol Ecol* **20:** 275-285.

van der Gast CJ, Walker AW, Stressmann FA, Rogers GB, Scott P, Daniels TW *et al* (2011). Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. *ISME J* **5:** 780-791.

Figure legends

Fig. 1. The effect of PMA treatment on (A & B) total bacteria and (C & D) *Pseudomonas aeruginosa* abundance. (A & C) Abundance (log_{10} scale CFU ml^{-1}) measured by Q-PCR for paired samples that have been PMA treated (closed circles) and not treated (open circles). Error bars represent the standard deviation of the mean (n = 3). (B & D) Meta-analysis of abundance using Hedges' d effect size measure between PMA treated and non-treated paired samples. Columns represent effect size and error bars represent the standard error (v_i) of the effect size (n = 6). Negative values indicate lower abundance in PMA treated samples. Standard error bars that cross zero indicate no significant effect by PMA treatment.

Fig. 2. The effect of PMA treatment on taxa richness and community similarity. (A) Taxa richness for paired samples that have been PMA treated (closed circles) and not treated (open circles). (C) Similarity (Sørensen index of similarity) of PMA treated (closed circles) samples compared to paired non-treated samples (assuming a similarity of 1 for comparison of a non-treated sample with itself). Richness and similarity were calculated with a uniform re-sample size of 1637 sequences following 1000 iterations in each instance. Error bars represent the standard deviation of the mean (n = 1000). Meta-analysis of (B) richness and (D) similarity using Hedges' d effect size measure between PMA treated and non-treated paired samples. Columns represent effect size and error bars represent the standard error (v_i) of the effect size (n = 2000). Standard error bars that cross zero indicate no significant effect by PMA treatment.

Fig. 3. Overall meta-analysis for measures of abundance, diversity, and community similarity using Hedges' d effect size measure between all PMA treated versus non-treated samples. Columns represent effect size and error bars represent the standard error (v_i) of the effect size (n = 30). Standard error bars that cross zero indicate no significant effect by PMA treatment.

Fig. 4. Distribution and dispersal of bacterial taxa across the (A & C) Non-PMA and (B & D) PMA treated metacommunity. (A & B) The number of samples for which each bacterial taxon was observed, plotted against the mean abundance (log_{10} scale) across all 30 samples ([A] r^2 = 0.86, $F_{1,147}$ = 679.4, P < 0.0001; [B] r^2 = 0.65, $F_{1,231}$ = 425.7, P < 0.0001). (C & D) Random and non-random dispersal through space visualised by decomposing the overall distribution using an index of dispersion based on the ratio of variance to the mean abundance for each bacterial taxon from the 30 samples. The line depicts the 2.5 % confidence limit for the χ^2 distribution. The 97.5% confidence limit was not plotted, as no taxon fell below that line.

Fig. 5. Rank-abundance plots to determine patterns of taxa frequency for the (A) non-PMA and (B) PMA treated metacommunity. In each instance, the relative rank positions of core group (closed grey circles) and satellite group (open diamonds) taxa are given. Each plot has been fitted with a slope, and the slope values (*b*) are given, to determine changes in evenness: (A) $r^2 = 0.88$, $F_{1,147} = 1030.5$, $P < 0.0001$; and (B) $r^2 = 0.88$, $F_{1,231} = 1669.1$, $P < 0.0001$.

Fig. 6. The effect of PMA treatment on metacommunity (A) richness and (B) community similarity (Sørensen). In each instance, results are given for all taxa, the core and satellite groups. Values and error bars ($n = 1000$) for richness and similarity were derived as described in the legends for Figs 1 & 2. Also given are the meta-analysis of (C) richness and (D) similarity using Hedges' *d* effect size measure. Columns represent effect size and error bars represent the standard error (v_i) of the effect size ($n = 2000$). Standard error bars that cross zero indicate no significant effect by PMA treatment.

Supplementary information

Fig. S1. Rarefaction curves for paired non-PMA and PMA treated sputum samples. Number denotes sample number and P denotes PMA treatment. Standard error of mean represented by shaded area.

Table S1 Quantitative PCR primer and probe sequences.

Table S2 Bacterial diversity measure (*S**, *D*, and *H*') values estimated for paired non–PMA and PMA (P) sputum samples, and adjusted for uniform sample sizes (*n*) of 1637 sequence reads. Each value was estimated with a re-sample size of 1637 sequence reads and taking the mean from 1000 re-samplings. Given for each sample is: *N*, full sample size; *S*, observed species richness at full sample size; and estimates of species richness (*S**), Simpson's index of diversity (*D*), and Shannon-Wiener index of diversity (*H*'). Also given are confidence intervals for each of the three measures.

Table S3 Bacterial taxa sampled across the 30 cystic fibrosis lung sputum samples. Ae, denotes aerobe; An, Anerobe[1]; O, taxa associated with the oral microbiome[2]; P, recognised CF pathogen. Also given for the non-PMA and PMA treated metacommunity are indications of core (C) and satellite (S) taxon group membership; lack of C or S designation indicates absence of detection of that taxon in the metacommunity.

Figure 1

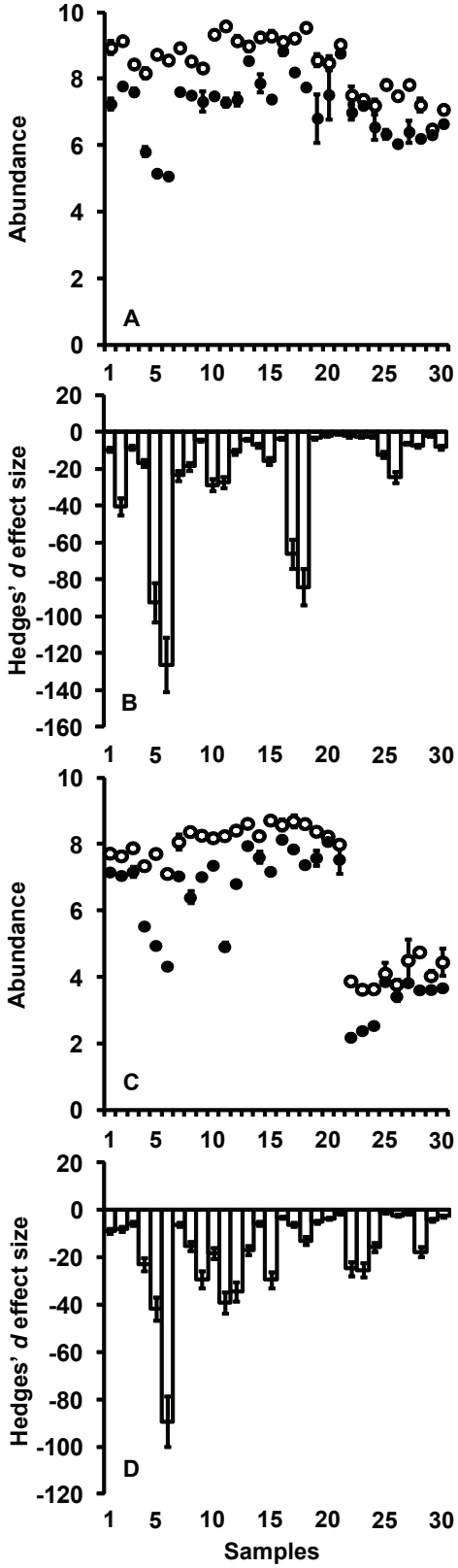
Abundance
A
Hedges' d effect size
B
Abundance
C
Hedges' d effect size
D
Samples

Figure 2

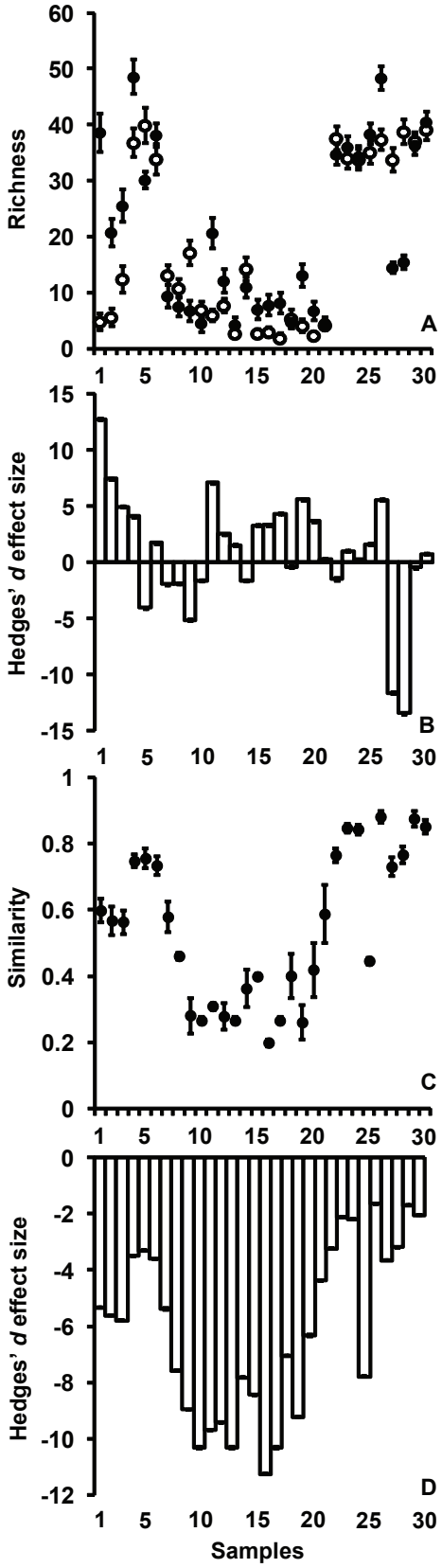
Richness
A
Hedges' d effect size
B
Similarity
C
Hedges' d effect size
D
Samples

Figure 3

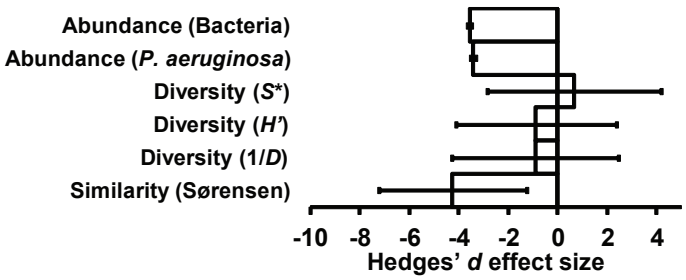
Abundance (Bacteria)
Abundance (P. aeruginosa)
Diversity (S*)
Diversity (H')
Diversity (1/D)
Similarity (Sørensen)
-10
-8
-6
-4
-2
0
2
4
Hedges' d effect size

Figure 4

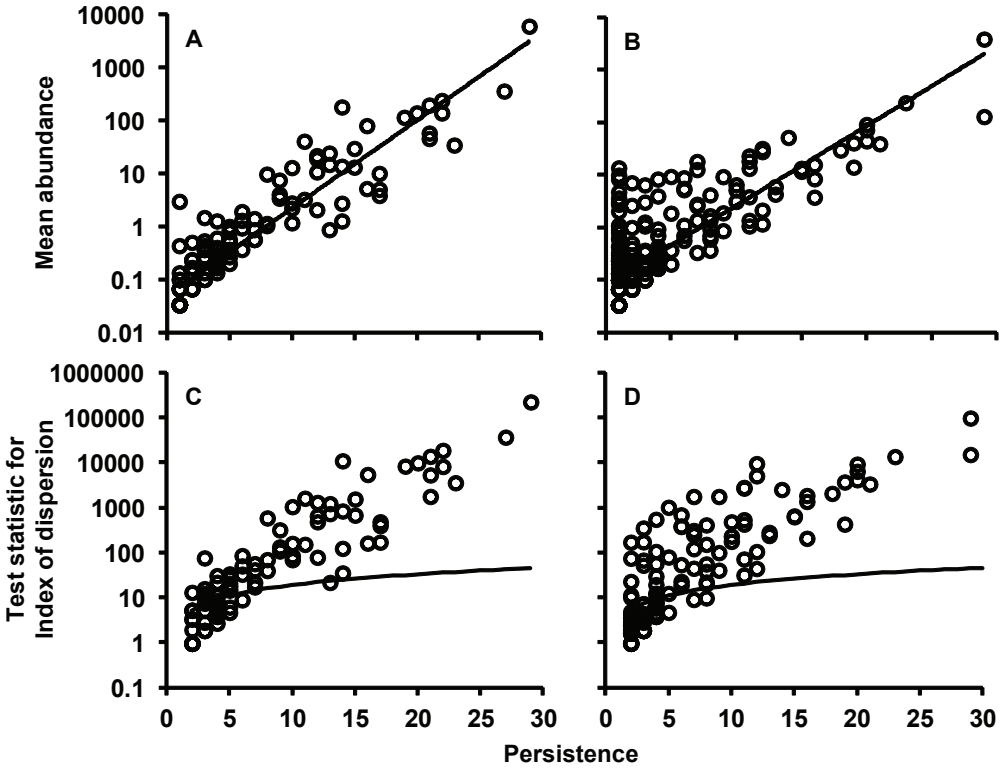
A
B
C
D
Mean abundance
Test statistic for Index of dispersion
Persistence
10000
1000
100
10
1
0.1
0.01
1000000
100000
10000
1000
100
10
1
0.1
0 5 10 15 20 25 30

Figure 5

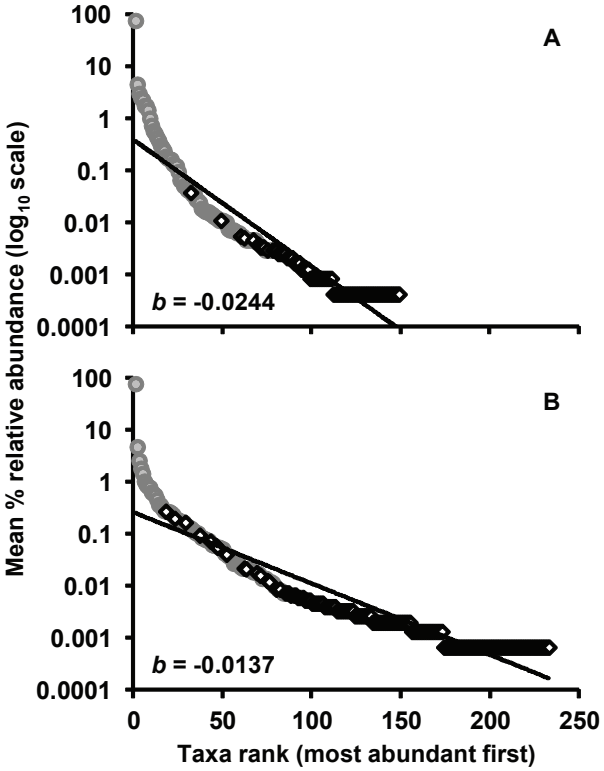
A
B
100
10
1
0.1
0.01
0.001
0.0001
b = -0.0244
b = -0.0137
Mean % relative abundance (log10 scale)
0
50
100
150
200
250
Taxa rank (most abundant first)

Figure 6

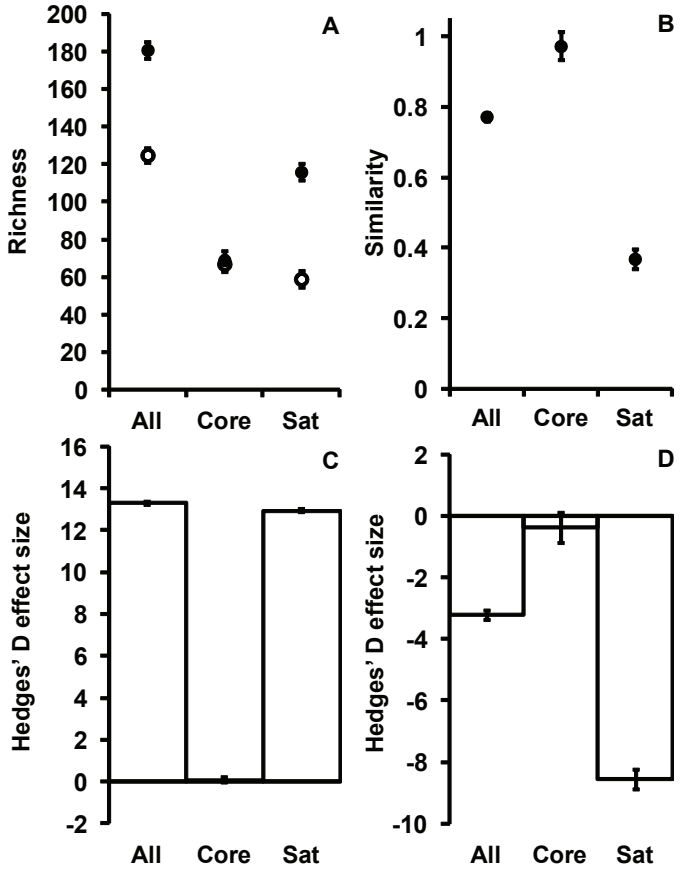
A
Richness
200
180
160
140
120
100
80
60
40
20
0
All
Core
Sat
B
Similarity
1
0.8
0.6
0.4
0.2
0
All
Core
Sat
C
Hedges' D effect size
16
14
12
10
8
6
4
2
0
-2
All
Core
Sat
D
Hedges' D effect size
2
0
-2
-4
-6
-8
-10
All
Core
Sat

Fig. S1. Rarefaction curves for paired non-PMA and PMA treated sputum samples. Number denotes sample number and P denotes PMA treatment. Standard error of mean represented by shaded area.

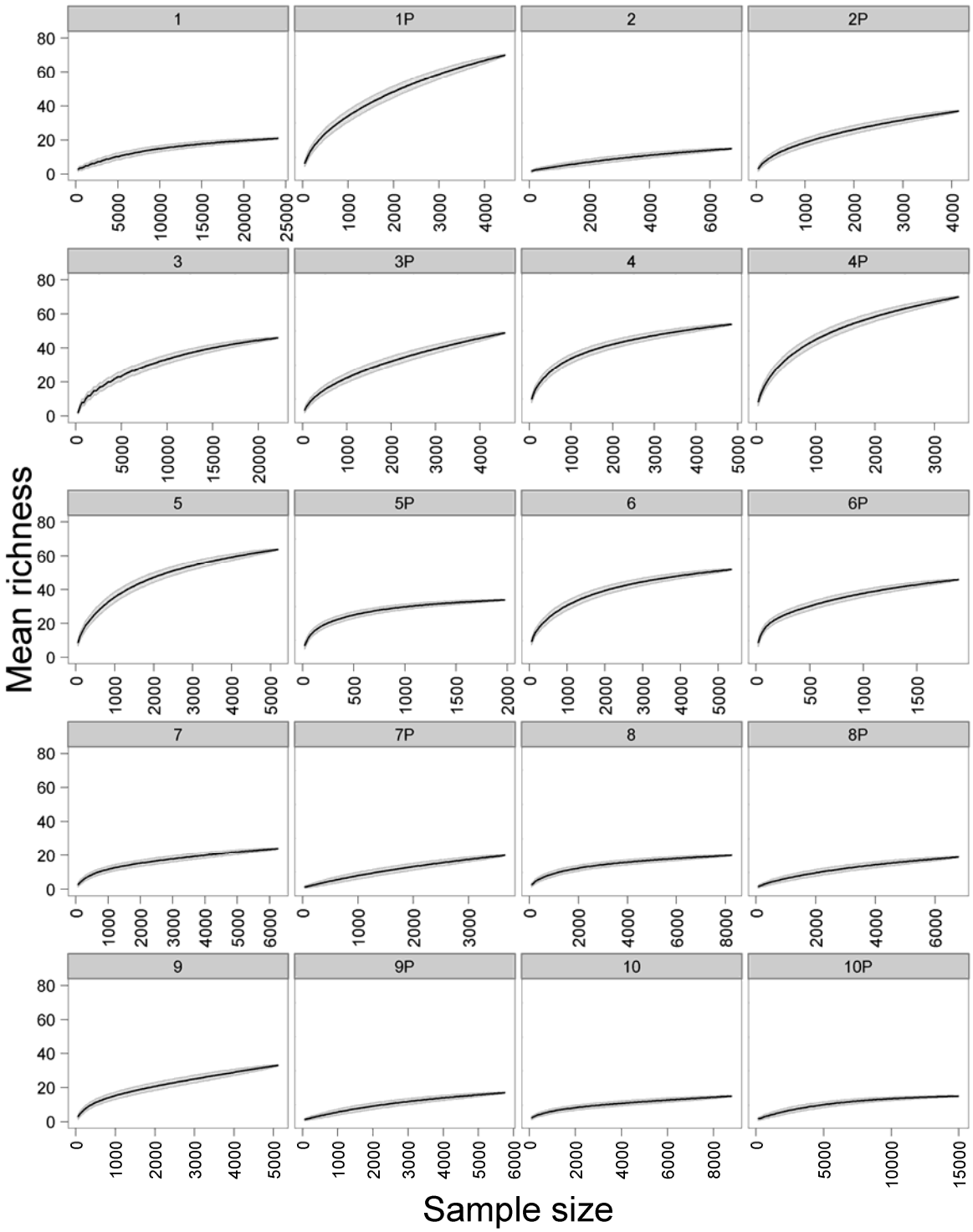

Fig. S1. Continued

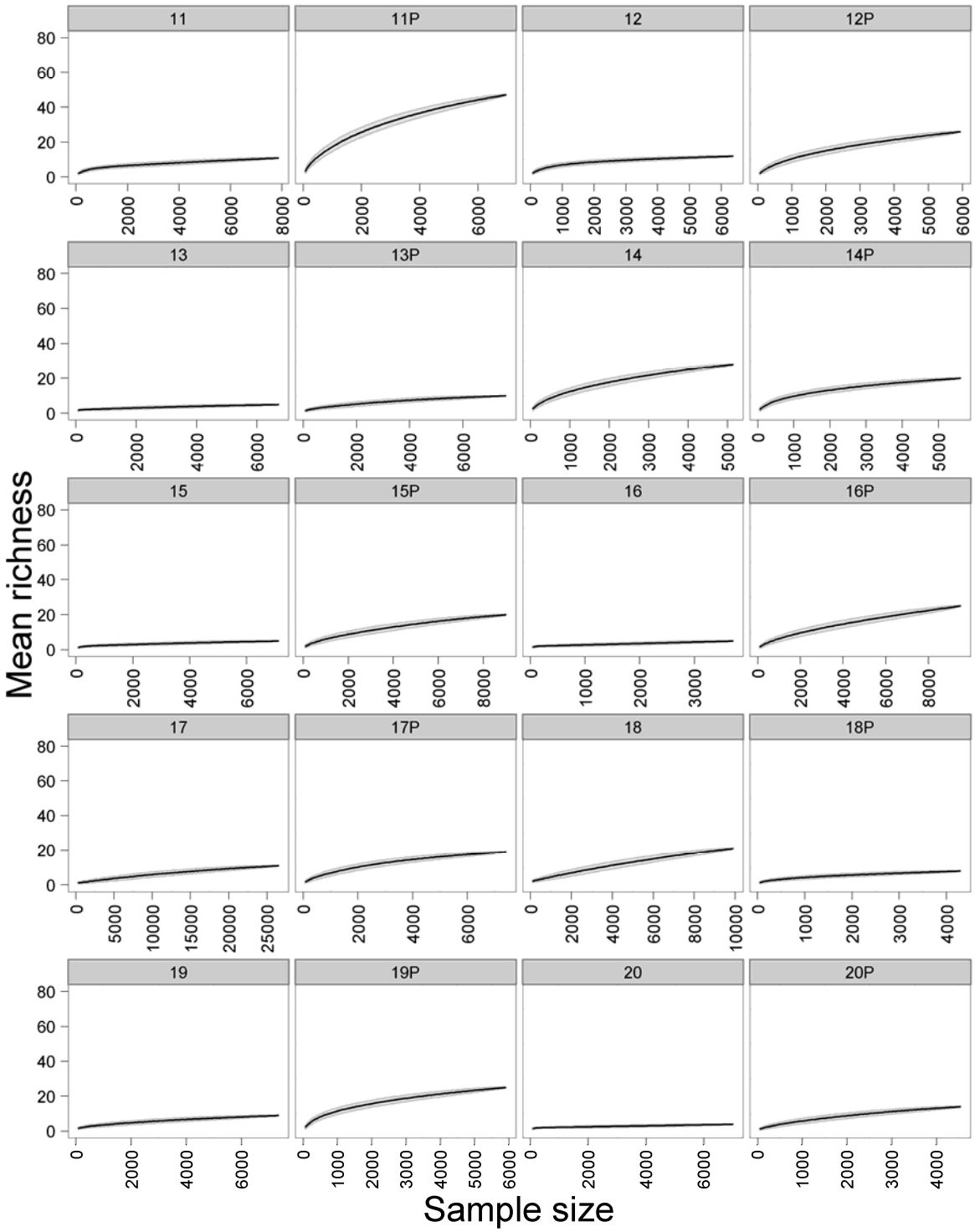

Fig. S1. Continued

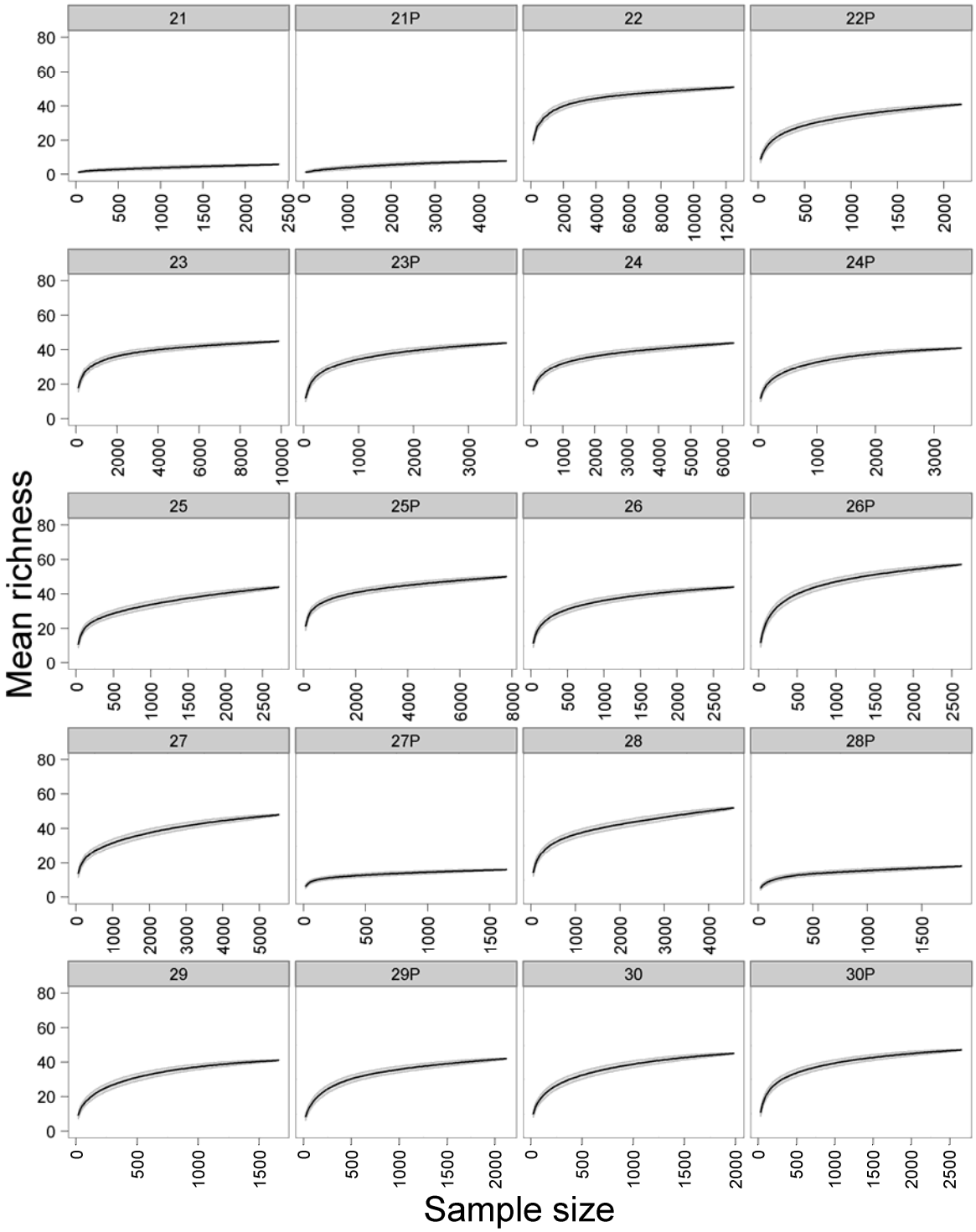

Table S1 Quantitative PCR primer and probe sequences.

Primer	Primer type	Sequence (5' to 3')	Reference
EubF	Total bacterial forward	TCCTACGGGAGGCAGCAGT	Nadkarni, 2002
EubR	Total bacterial reverse	GGACTACCAGGGTATCTAATCCTGTT	Nadkarni, 2002
EubPR	Total bacterial Taqman probe	FAM-CGTATTACCGCGGCTGCTGGCAC-TAMRA	Nadkarni, 2002
PaerF	*P. aeruginosa* forward	TGCTGGTGGCACAGGACAT	Shannon, 2007
PaerR	*P. aeruginosa* reverse	TTGTTGGTGCAGTTCCTCATTG	Shannon, 2007
PaerPR	*P. aeruginosa* Taqman probe	FAM-CAGATGCTTTGCCTCAA-TAMRA	Shannon, 2007

All quantitative (Q)PCR analyses were performed in triplicate. Total bacterial density was determined using a Taqman assay, in which a 466 bp fragment of the 16S ribosomal RNA gene was amplified, as described previously (Nadkarni et al 2002). *P. aeruginosa* density was determined using a Taqman assay which amplified a 65 bp fragment of the regA gene, as described previously (Rogers et al 2010, Shannon et al 2007). Details of the relevant primers and probes used are shown in the Table above. Bacterial primers and probe were used at a concentration of 100 nM each, whereas *P. aeruginosa*-specific primers were used at a concentration of 1000 nM each, and the probe at a concentration of 250 nM. 1 µl (~800 ng) of mixed template DNA (human and microbial) was used in the *P. aeruginosa* assay. 1 µl of a 100 fold dilution (~8 ng) was used in the total bacterial assay. All PCR reactions were carried out in a total volume of 25 µl in Taqman® Universal PCR Mastermix (Applied Biosystems, Warrington, UK). Quantitative PCR assays were carried out using the Rotorgene 6000 (Qiagen,Crawley,UK) with a temperature profile of 50 °C for 2 min, 95 °C for 10 min, followed by 45 cycles at 95 °C for 15 s and 60 °C for 60 s.

For both total bacterial and *P. aeruginosa*-specific quantitative PCR assays, densities (cfu/ml) were determined by comparison with standard curves generated from bacterial isolates. Nutrient broth cultures of *P. aeruginosa* (NCTC 12934/ ATCC 27853) were incubated at 37 °C for 16 h, with cfu ml^{-1} estimated by incubation of dilutions (n = 4) on Columbia Blood Agar at 37° for 24 h, followed by colony counts. DNA was extracted from tenfold dilutions of these broth cultures in the same way as for the sputum samples, and RT-PCR was carried out as above on the DNA extracts. The standard curve generated using *P. aeruginosa* was used for both *P. aeruginosa* and total bacterial enumeration to allow direct comparisons to be made.

Nadkarni MA, Martin FE, Jacques NA, Hunter N (2002). Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology* **148:** 257-266.

Shannon KE, Lee DY, Trevors JT, Beaudette LA (2007). Application of real-time quantitative PCR for the detection of selected bacterial pathogens during municipal wastewater treatment. *Sci Total Environ* **382:** 121-129.

Table S2 Bacterial diversity measure (*S**, *D*, and *H*') values estimated for paired non–PMA and PMA (P) sputum samples, and adjusted for uniform sample sizes (*n*) of 1637 sequence reads. Each value was estimated with a re-sample size of *n* = 1637 sequence reads and taking the mean from 1000 re-samplings. Given for each sample is: *N*, full sample size; *S*, observed species richness at full sample size; and estimates of species richness (*S**), Simpson's index of diversity (*D*), and Shannon-Wiener index of diversity (*H*'). Also given are confidence intervals for each of the three measures.

Name	*N*	*S*	*S**	$S^*_{(2.5\%)}$	$S^*_{(97.5\%)}$	*D*	$D_{(2.5\%)}$	$D_{(97.5\%)}$	*H'*	$H'_{(2.5\%)}$	$H'_{(97.5\%)}$
1	24108	21	5	2.0	8.0	0.968	0.956	0.980	0.09	0.06	0.12
1P	4448	70	39	32.0	45.0	0.601	0.568	0.631	1.06	0.97	1.15
2	6728	15	6	3.0	9.0	0.967	0.953	0.979	0.10	0.06	0.13
2P	4138	37	21	16.0	26.0	0.883	0.862	0.904	0.37	0.31	0.43
3	22181	46	12	8.0	18.0	0.960	0.945	0.974	0.14	0.09	0.18
3P	4570	49	26	20.0	31.0	0.878	0.855	0.900	0.40	0.34	0.48
4	4860	54	37	32.0	42.0	0.389	0.362	0.416	1.61	1.54	1.70
4P	3408	70	49	43.0	55.0	0.443	0.415	0.469	1.56	1.48	1.66
5	5200	64	40	34.0	46.0	0.497	0.466	0.528	1.36	1.28	1.45
5P	1979	34	30	27.0	33.0	0.341	0.315	0.367	1.78	1.70	1.86
6	5359	52	34	29.0	39.0	0.414	0.388	0.442	1.51	1.43	1.59
6P	1888	46	38	34.0	42.0	0.207	0.189	0.227	2.28	2.20	2.35
7	6269	24	13	10.0	17.0	0.938	0.918	0.954	0.20	0.16	0.26
7P	3656	20	9	6.0	14.0	0.984	0.974	0.992	0.06	0.03	0.10
8	8260	20	11	7.0	14.0	0.945	0.926	0.960	0.18	0.13	0.23
8P	6789	19	8	4.0	11.0	0.983	0.972	0.991	0.06	0.03	0.10
9	5119	33	17	13.0	22.0	0.915	0.895	0.935	0.28	0.22	0.34
9P	5755	17	7	3.0	11.0	0.989	0.981	0.996	0.04	0.02	0.07
10	8797	15	7	5.0	10.0	0.964	0.951	0.976	0.11	0.08	0.15
10P	14981	15	5	2.0	8.0	0.995	0.988	0.999	0.02	0.01	0.05
11	7878	11	6	4.0	8.0	0.974	0.962	0.984	0.09	0.05	0.12
11P	6967	47	21	16.0	26.0	0.918	0.898	0.936	0.28	0.22	0.34
12	6379	12	8	5.0	10.0	0.954	0.939	0.968	0.14	0.10	0.18
12P	5942	26	12	8.0	16.0	0.960	0.946	0.974	0.14	0.10	0.19
13	6728	5	3	2.0	4.0	0.968	0.954	0.979	0.08	0.06	0.11
13P	7616	10	4	2.0	7.0	0.985	0.976	0.993	0.05	0.02	0.08
14	5147	28	14	10.0	19.0	0.937	0.919	0.953	0.21	0.15	0.26
14P	5612	20	11	8.0	15.0	0.958	0.941	0.972	0.15	0.10	0.20
15	7135	5	3	2.0	4.0	0.989	0.980	0.996	0.04	0.01	0.06
15P	9007	20	7	4.0	10.0	0.979	0.968	0.988	0.07	0.04	0.11
16	3714	5	3	2.0	5.0	0.965	0.952	0.978	0.09	0.06	0.12
16P	9504	25	8	5.0	12.0	0.981	0.971	0.991	0.07	0.04	0.10
17	26563	11	2	1.0	4.0	0.999	0.996	1.000	0.01	0.00	0.02
17P	7416	19	8	5.0	12.0	0.974	0.962	0.984	0.09	0.06	0.13
18	9909	21	5	3.0	9.0	0.970	0.958	0.981	0.09	0.06	0.12
18P	4298	8	5	3.0	7.0	0.980	0.968	0.989	0.07	0.04	0.10
19	7360	9	4	2.0	6.0	0.981	0.970	0.989	0.06	0.03	0.09
19P	5915	25	13	9.0	17.0	0.949	0.932	0.963	0.18	0.13	0.23
20	7020	4	2	2.0	4.0	0.981	0.971	0.989	0.05	0.03	0.08
20P	4531	14	7	4.0	10.0	0.984	0.975	0.992	0.06	0.03	0.09
21	2399	6	4	3.0	6.0	0.971	0.959	0.981	0.08	0.06	0.11
21P	4635	8	5	3.0	7.0	0.985	0.976	0.993	0.05	0.03	0.08
22	12488	51	38	34.0	42.0	0.119	0.111	0.128	2.57	2.52	2.63
22P	2191	41	35	31.0	38.0	0.145	0.137	0.153	2.34	2.28	2.40
23	9897	45	34	30.0	38.0	0.111	0.105	0.118	2.52	2.46	2.58
23P	3692	44	36	32.0	40.0	0.151	0.140	0.163	2.39	2.33	2.46

Table S2 (continued)

Name	***N***	***S***	***S****	***S**** $_{(2.5\%)}$	***S**** $_{(97.5\%)}$	***D***	***D*** $_{(2.5\%)}$	***D*** $_{(97.5\%)}$	***H***	***H*** $_{(2.5\%)}$	***H*** $_{(97.5\%)}$
24	6362	44	34	30.0	37.0	0.108	0.101	0.116	2.59	2.53	2.64
24P	3468	41	34	31.0	38.0	0.125	0.117	0.132	2.46	2.40	2.52
25	2721	44	35	31.0	39.0	0.144	0.135	0.156	2.46	2.39	2.51
25P	7787	50	38	35.0	42.0	0.083	0.077	0.089	2.87	2.81	2.91
26	2783	44	37	34.0	41.0	0.140	0.129	0.151	2.50	2.44	2.57
26P	2625	57	49	44.0	52.0	0.128	0.117	0.139	2.68	2.61	2.75
27	5539	48	34	30.0	38.0	0.176	0.163	0.190	2.27	2.20	2.33
27P	1637	16	15	13.0	16.0	0.178	0.171	0.187	1.93	1.88	1.97
28	4569	52	39	35.0	43.0	0.107	0.101	0.114	2.61	2.56	2.67
28P	1865	18	16	13.0	18.0	0.304	0.286	0.323	1.56	1.51	1.61
29	1663	41	37	34.0	40.0	0.099	0.093	0.105	2.63	2.57	2.68
29P	2121	42	37	33.0	40.0	0.194	0.181	0.208	2.21	2.14	2.27
30	1984	45	39	36.0	43.0	0.100	0.095	0.107	2.63	2.57	2.69
30P	2653	47	41	36.0	44.0	0.166	0.154	0.179	2.43	2.36	2.50

Table S3 Bacterial taxa sampled across the 30 cystic fibrosis lung sputum samples. Ae, denotes aerobe; An, Anerobe[1]; O, taxa associated with the oral microbiome[2]; P, recognised CF pathogen. Also given for the non-PMA and PMA treated metacommunity are indications of core (C) and satellite (S) taxon group membership; lack of C or S designation indicates absence of detection of that taxon in the metacommunity.

Class	Family	Taxon name	Code	Non-PMA	PMA
Acidobacteria	Acidobacteriaceae	*Acidobacterium sp*	Ae	S	S
Actinobacteria	Actinomycetaceae	*Actinomyces meyeri*	Ae, O	S	S
		Actinomyces odontolyticus	Ae, O	C	C
		Actinomyces sp	Ae, O	C	C
	Corynebacteriaceae	*Corynebacterium durum*	Ae, O		S
		Corynebacterium jeikeium	Ae	S	
		C. pseudodiphtheriticum	Ae, O	S	
		Corynebacterium segmentosum	Ae	S	S
		Corynebacterium sp	Ae, O	S	C
	Geodermatophilaceae	*Geodermatophilus sp*	Ae		S
	Gordoniaceae	*Gordonia rubripertincta*	Ae	S	S
	Intrasporangiaceae	*Janibacter sp*	Ae		S
	Microbacteriaceae	*Microbacterium lacticum*	Ae	S	
	Micrococcaceae	*Arthrobacter psychrolactophilus*	Ae		S
		Arthrobacter sp	Ae		C
		Citricoccus sp	Ae		S
		Kocuria sp	Ae, O	S	S
		Micrococcus sp	Ae	C	S
		Rothia dentocariosa	Ae, O	C	C
		Rothia mucilaginosa	Ae, O	C	C
		Rothia sp	Ae, O	C	C
	Mycobacteriaceae	*Mycobacterium sp*	Ae		S
	Nocardioidaceae	*Nocardioides alkalitolerans*	Ae		S
	Propionibacteriaceae	*Propionibacterium acnes*	Ae	S	C
		Propionibacterium avidum	Ae, O		S
		Propionibacterium propionicum	Ae, O	C	S
		Propionibacterium sp	Ae, O	S	C
	Sporichthyaceae	*Sporichthya polymorpha*	Ae		S
	Bifidobacteriaceae	*Scardovia sp*	An, O	S	
	Coriobacteriaceae	*Atopobium parvulum*	An, O	S	S
		Eggerthella sp	An, O		S
Bacteroidia	Bacteroidaceae	*Bacteroides acidifaciens*	An	S	
		Bacteroides ovatus	An		S
		Bacteroides salyersiae	An		S
		Bacteroides sp	An		C
		Bacteroides thetaiotaomicron	An		S
	Porphyromonadaceae	*Parabacteroides distasonis*	An		S
		Porphyromonas bennonis	An	C	
		Porphyromonas catoniae	An, O	S	S
		Porphyromonas sp	An	C	C
	Prevotellaceae	*Prevotella copri*	An		S
		Prevotella denticola	An, O	C	C
		Prevotella histicola	An, O	C	S
		Prevotella melaninogenica	An, O	C	C
		Prevotella nanceiensis	An	C	C
		Prevotella oris	An, O	S	S
		Prevotella oulorum	An, O	C	S

Table S3 (Continued)

Class	Family	Taxon name	Code	Non-PMA	PMA
		Prevotella pallens	An, O	C	C
		Prevotella ruminicola	An	S	
		Prevotella salivae	An, O	C	C
		Prevotella sp	An, O	C	C
		Prevotella veroralis	An, O	C	S
	Bacteroidia	*Bacteroidales bacterium*	An		S
Cytophagia	Cytophagaceae	*Flexibacter sp*	Ae		S
Flavobacteria	Flavobacteriaceae	*Bergeyella zoohelcum*	Ae, O	C	S
		Capnocytophaga gingivalis	Ae, O	C	C
		Capnocytophaga granulosa	Ae, O	S	
		Capnocytophaga sp	Ae, O	C	C
		Chryseobacterium sp	Ae		S
		Maribacter sp	Ae		S
Sphingobacteria	Chitinophagaceae	*Flavisolibacter sp*	Ae		S
	Sphingobacteriaceae	*Mucilaginibacter sp*	Ae		S
		Pedobacter duraquae	Ae	S	
Caldilineae	Caldilineaceae	*Caldilinea sp*	An		S
Cyanobacteria	Chroococcaceae	*Gloeocapsa sp*	Ae		S
	Nostocaceae	*Anabaena augstumalis*	Ae		S
		Anabaena cylindrica	Ae		S
		Nostoc commune	Ae	S	
	Scytonemataceae	*Brasilonema terrestre*	Ae		S
		Scytonema sp	Ae	S	
	Stigonemataceae	*Symphyonemopsis sp*	Ae		S
Elusimicrobia	Elusimicrobiaceae	*Elusimicrobium sp*	An		S
Bacilli	Bacillaceae	*Anoxybacillus flavithermus*	Ae		S
		Bacillus cereus	Ae		C
		Bacillus licheniformis	Ae	S	
		Bacillus sp	Ae	S	C
		Gemella haemolysans	Ae, O	C	C
		Gemella sanguinis	Ae, O	C	C
		Gemella sp	Ae, O	C	S
	Paenibacillaceae	*Paenibacillus sp*	Ae, O	S	C
	Planococcaceae	*Planococcus sp*	Ae		S
	Staphylococcaceae	*Staphylococcus aureus*	Ae, P	S	
		Staphylococcus auricularis	Ae		S
		Staphylococcus epidermidis	Ae, O	S	C
		Staphylococcus haemolyticus	Ae	S	S
		Staphylococcus sp	Ae		C
		Staphylococcus warneri	Ae, O	S	
	Aerococcaceae	*Abiotrophia defectiva*	Ae, O		S
		Abiotrophia sp	Ae, O	C	C
		Facklamia sp	Ae		S
	Carnobacteriaceae	*Dolosigranulum sp*	Ae, O	S	
		Granulicatella adiacens	Ae, O	C	C
		Granulicatella elegans	Ae, O	S	S
		Granulicatella sp	Ae, O	C	C
	Enterococcaceae	*Enterococcus sp*	Ae	S	
	Lactobacillaceae	*Lactobacillus delbrueckii*	Ae		S
		Lactobacillus mucosae	Ae		S
		Lactobacillus rhamnosus	Ae, O	S	S
		Lactobacillus sp	Ae, O	S	C

Table S3 (Continued)

Class	Family	Taxon name	Code	Non-PMA	PMA
	Streptococcaceae	*Streptococcus australis*	Ae	C	C
		Streptococcus cristatus	Ae	C	C
		Streptococcus infantis	Ae, O	C	C
		Streptococcus intermedius	Ae, O	S	
		Streptococcus mitis	Ae, O	C	C
		Streptococcus mutans	Ae, O		S
		Streptococcus oralis	Ae, O	S	S
		Streptococcus parasanguinis	Ae, O	C	C
		Streptococcus pneumoniae	Ae, O,P	C	C
		Streptococcus salivarius	Ae, O	C	C
		Streptococcus sanguinis	Ae, O	C	C
		Streptococcus sp	Ae, O	C	C
		Streptococcus thermophilus	Ae, O	S	S
		Streptococcus vestibularis	Ae, O	C	C
Clostridia	Clostridiaceae	*Anaerococcus sp*	An, O	S	S
		Blautia hansenii	An		S
		Blautia sp	An		S
		Clostridium neopropionicum	An		S
		Clostridium orbiscindens	An		S
		Clostridium sp	An		C
		Finegoldia sp	An, O	C	S
		Mogibacterium vescum	An, O		S
		Peptoniphilus sp	An, O	S	C
		Tissierella sp	An		S
	Eubacteriaceae	*Eubacterium eligens*	An	S	
		Eubacterium saburreum	An, O	S	C
		Eubacterium sp	An		S
	Lachnospiraceae	*Anaerostipes sp*	An	C	S
		Catonella sp	An, O		S
		Dorea sp	An, O		S
		Lachnospira pectinoschiza	An	S	S
		Oribacterium sinus	An, O	S	S
		Oribacterium sp	An, O		S
		Roseburia intestinalis	An		S
		Roseburia inulinivorans	An		S
	Oscillospiraceae	*Oscillibacter valericigenes*	An	S	S
	Peptostreptococcaceae	*Peptostreptococcus sp*	An		S
	Ruminococcaceae	*Faecalibacterium prausnitzii*	Ae	S	S
		Faecalibacterium sp	Ae		S
		Oscillospira guilliermondii	An		S
		Ruminococcus obeum	An		C
	Veillonellaceae	*Dialister sp*	An, O		S
		Megamonas sp	An		S
		Megasphaera micronuciformis	An, O	S	S
		Selenomonas dianae	An, O	C	S
		Selenomonas sp	An, O	S	S
		Selenomonas sputigena	An, O	C	C
		Veillonella atypica	An, O	C	C
		Veillonella dispar	An, O	C	C
		Veillonella parvula	An, O	C	C
		Veillonella sp	An, O		S
	Thermoanaerobacteraceae	*Gelria sp*	An	S	

Table S3 (Continued)

Class	Family	Taxon name	Code	Non-PMA	PMA
		Thermoanaerobacter sp	An	S	
Erysipelotrichi	Erysipelotrichaceae	*Bulleidia sp*	An, O		S
		Catenibacterium sp	An	C	S
		Solobacterium moorei	An, O	S	
		Solobacterium sp	An, O		S
		Turicibacter sanguinis	An		S
Fusobacteria	Fusobacteriaceae	*Fusobacterium nucleatum*	An, O		S
		Fusobacterium periodonticum	An, O		S
		Fusobacterium sp	An, O	C	S
		Leptotrichia sp	An, O	S	S
		Leptotrichia wadei	An, O	S	
Gemmatimonadetes	Gemmatimonadaceae	*Gemmatimonas sp*	Ae		S
Lentisphaerae	Victivallaceae	*Victivallis sp*	An		S
Nitrospira	Nitrospiraceae	*Nitrospira sp*	Ae		S
Alphaproteobacteria	Caulobacteraceae	*Brevundimonas diminuta*	Ae, O	S	
		Brevundimonas sp	Ae	C	S
		Caulobacter sp	Ae, O	S	
	Bradyrhizobiaceae	*Bradyrhizobium sp*	Ae, O		C
	Hyphomicrobiaceae	*Devosia sp*	Ae		S
		Hyphomicrobium sp	Ae	S	
	Methylobacteriaceae	*Methylobacterium sp*	Ae	C	
		Methylobacterium zatmanii	Ae		S
	Rhizobiaceae	*Rhizobium sp*	Ae		S
	Rhodobacteraceae	*Pannonibacter phragmitetus*	Ae		S
		Paracoccus sp	Ae	S	S
		Rhodobacter sp	Ae	S	S
		Roseobacter sp	Ae		C
	Acetobacteraceae	*Roseomonas gilardii*	Ae		S
	Rhodospirillaceae	*Azospirillum sp*	Ae		S
		Defluviicoccus sp	Ae		S
		Thalassospira sp	Ae		S
	Anaplasmataceae	*Wolbachia sp*	Ae		S
	Sphingomonadaceae	*Novosphingobium sp*	Ae	S	
		Sphingomonas sp	Ae	S	S
Betaproteobacteria	Alcaligenaceae	*Achromobacter sp*	Ae		S
		Achromobacter xylosoxidans	Ae, P		S
		Alcaligenes faecalis	Ae		S
		Alcaligenes sp	Ae		C
		Sutterella sp	Ae		S
	Burkholderiaceae	*Burkholderia sp*	Ae, P		C
		Lautropia mirabilis	Ae, O	C	C
		Lautropia sp	Ae, O	C	S
		Ralstonia sp	Ae, O	C	C
		Mitsuaria sp	Ae		S
	Comamonadaceae	*Acidovorax defluvii*	Ae		S
		Acidovorax sp	Ae		S
		Comamonas sp	Ae		S
		Diaphorobacter sp	Ae		S
		Pelomonas sp	Ae	S	
		Ramlibacter sp	Ae		S
	Oxalobacteraceae	*Duganella sp*	Ae		S

Table S3 (Continued)

Class	Family	Taxon name	Code	Non-PMA	PMA
		Herbaspirillum sp	Ae	S	S
		Janthinobacterium sp	Ae		S
		Massilia sp	Ae	C	C
		Massilia timonae	Ae	S	S
	Neisseriaceae	*Eikenella corrodens*	Ae, O		S
		Kingella denitrificans	Ae, O	S	S
		Kingella oralis	Ae, O	S	S
		Kingella sp	Ae, O	C	S
		Neisseria bacilliformis	Ae, O	C	S
		Neisseria mucosa	Ae, O	S	S
		Neisseria sp	Ae, O	S	S
	Rhodocyclaceae	*Sterolibacterium sp*	Ae		S
		Zoogloea oryzae	Ae		S
Deltaproteobacteria	Desulfobulbaceae	*Desulfofustis sp*	Ae		S
Epsilonproteobacteria	Campylobacteraceae	*Campylobacter concisus*	Ae, O	S	S
		Campylobacter curvus	Ae, O	S	S
		Campylobacter gracilis	Ae, O	C	S
		Campylobacter sp	Ae, O	C	
Gammaproteobacteria	Aeromonadaceae	*Aeromonas sp*	Ae	S	S
	Enterobacteriaceae	*Citrobacter sp*	Ae		S
		Enterobacter hormaechei	Ae, O		S
		Enterobacter sp	Ae, O	S	C
		Escherichia coli	Ae, O		C
		Escherichia sp	Ae, O	S	C
		Klebsiella sp	Ae	S	C
		Pantoea sp	Ae		S
		Pectobacterium carotovorum	Ae	S	
		Proteus mirabilis	Ae, O		S
		Proteus sp	Ae, O		S
		Salmonella enterica	Ae		S
		Serratia sp	Ae	S	
		Shigella flexneri	Ae		S
		Shigella sp	Ae		C
	Alcanivoracaceae	*Alcanivorax dieselolei*	Ae	S	
	Pasteurellaceae	*Actinobacillus pleuropneumoniae*	Ae	C	S
		Actinobacillus sp	Ae	S	
		Aggregatibacter aphrophilus	Ae, O	S	
		Aggregatibacter sp	Ae, O		S
		Haemophilus influenzae	Ae, P		S
		Haemophilus parahaemolyticus	Ae, O	C	S
		Haemophilus sp	Ae, O	C	C
		Pasteurella skyensis	Ae		S
		Terrahaemophilus aromaticivorans	An, O	C	S
	Moraxellaceae	*Acinetobacter radioresistens*	Ae	S	C
		Acinetobacter sp	Ae	S	C
		Psychrobacter fulvigenes	Ae		S
		Psychrobacter sp	Ae		S
	Pseudomonadaceae	*Pseudomonas aeruginosa*	Ae, P	C	C
		Pseudomonas delhiensis	Ae	S	S
		Pseudomonas fluorescens	Ae, O	C	C

Table S3 (Continued)

Class	Family	Taxon name	Code	Non-PMA	PMA
		Pseudomonas pertucinogena	Ae		S
		Pseudomonas pseudoalcaligenes	Ae, O	C	C
		Pseudomonas putida	Ae	C	S
		Pseudomonas tropicalis	Ae		S
	Vibrionaceae	*Vibrio harveyi*	Ae		S
		Vibrio sp	Ae	S	S
	Xanthomonadaceae	*Lysobacter panaciterrae*	Ae		S
		Pseudoxanthomonas suwonensis	Ae	S	S
		Stenotrophomonas maltophilia	Ae, P	C	C
		Stenotrophomonas sp	Ae	C	C
		Xanthomonas sp	Ae, O	C	C
Mollicutes	Mollicutes(family)	*Mollicutes sp*	An		S
	Mycoplasmataceae	*Ureaplasma urealyticum*	Ae		S
TM7(class)	TM7(family)	*TM7 uncultured*	Ae, O	C	C
Verrucomicrobiae	Verrucomicrobiaceae	*Akkermansia muciniphila*	An		S

[1] Only strict anaerobes were classified as anaerobes, whereas aerobes, facultative anaerobes, and microaerophiles were classified as aerobes, as described previously (van der Gast *et al.*, 2011).

[2] Classification of oral taxa was according to the Human Oral Microbiome Database (Dewhirst *et al.*, 2010).

References

Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner ACR, Wen-Han Y *et al* (2010). The human oral microbiome. *J Bacteriol* **192:** 5002-5017.

van der Gast CJ, Walker AW, Stressmann FA, Rogers GB, Scott P, Daniels TW *et al* (2011). Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. *ISME J* **5:** 780-791.