

## ORIGINAL ARTICLE

# Reducing bias in bacterial community analysis of lower respiratory infections

Geraint B Rogers<sup>1</sup>, Leah Cuthbertson<sup>2</sup>, Lucas R Hoffman<sup>3,4</sup>, Peter AC Wing<sup>5</sup>, Christopher Pope<sup>3,4</sup>, Danny AP Hooftman<sup>2</sup>, Andrew K Lilley<sup>1</sup>, Anna Oliver<sup>2</sup>, Mary P Carroll<sup>5</sup>, Kenneth D Bruce<sup>1</sup> and Christopher J van der Gast<sup>2</sup>

<sup>1</sup>Molecular Microbiology Research Laboratory, Institute of Pharmaceutical Science, King's College London, London, UK; <sup>2</sup>NERC Centre for Ecology and Hydrology, Wallingford, UK; <sup>3</sup>Department of Pediatrics, University of Washington, Seattle, WA, USA; <sup>4</sup>Department of Microbiology, University of Washington, Seattle, WA, USA and <sup>5</sup>Cystic Fibrosis Unit, Southampton University Hospitals NHS Trust, Southampton, UK

**High-throughput pyrosequencing and quantitative PCR (Q-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, adult cystic fibrosis (CF) sputum samples were used 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.**

*The ISME Journal* (2013) 7, 697–706; doi:10.1038/ismej.2012.145; published online 29 November 2012

**Subject Category:** microbial population and community ecology

**Keywords:** propidium monoazide; pyrosequencing; 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 responses 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 exerted 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.

The 16S rRNA gene pyrosequencing and quantitative PCR (Q-PCR) analysis offer greatly improved accuracy and depth of characterisation of microbial communities. However, although avoiding the selective biases associated with culture-based microbiology, culture-independent approaches will

Correspondence: CJ van der Gast, NERC Centre for Ecology and Hydrology, Benson Lane, Wallingford OX10 8BB, UK.  
E-mail: cjvdg@ceh.ac.uk

Received 1 May 2012; revised 7 October 2012; accepted 19 October 2012; published online 29 November 2012

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, 2008, 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, b). Importantly though, PMA is highly membrane impermeant (Nocker and Camper, 2006). This means that although 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 (Nocker *et al.*, 2007a; Bae and Wuertz, 2009; Kralik *et al.*, 2010; Nam *et al.*, 2011; Taskin *et al.*, 2011), including the assessment of microbiota present in CF airway samples (Rogers *et al.*, 2008, 2010, 2011). Further, this strategy has been used previously in conjunction with bacterial pyrosequencing in waste-water treatment systems (Nocker *et al.*, 2010). Although the benefits of NVCE are leading to its increasing deployment, its impact on microbial community data has not previously been determined.

In this study, the application of NVCE to profiling the bacterial communities present in spontaneously expectorated sputum samples collected from adult CF patients is detailed. 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 airway disease (Ballmann *et al.*, 1998; Kosorok *et al.*, 2001). Comparisons were made first between paired PMA-treated and non-PMA-treated samples (that is, 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 (that is, 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  $-80^{\circ}\text{C}$  before analysis. Each sample was mixed by pipetting and divided into two equal portions, one of which was treated with PMA before DNA extraction.

### PMA cross-linking

PMA was dissolved in 20% dimethyl sulfoxide to create a stock concentration of  $20\text{ mmol l}^{-1}$  with this, added to samples to give a final concentration of  $50\text{ }\mu\text{mol l}^{-1}$ . 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  $5000\text{ g}$  for 5 min before DNA isolation.

### DNA extraction

Nucleic acid extractions were performed on 100  $\mu\text{l}$  portions of sputum. Guanidinium thiocyanate–EDTA–sarkosyl (500  $\mu\text{l}$ ) and phosphate-buffered saline (500  $\mu\text{l}$ ), pH 8.0, were added to samples. Cell disruption was achieved using a Fastprep-24 Instrument (MP Biomedicals Europe, Illkirch, France)  $6.5\text{ m s}^{-1}$ , 60 s, followed by incubation at  $90^{\circ}\text{C}$  for 1 min and  $-20^{\circ}\text{C}$  for 5 min. Cell debris was pelleted by centrifugation at  $12\text{ }000\text{ g}$  for 2 min at  $4^{\circ}\text{C}$ . Supernatant was transferred to a fresh microfuge tube. NaCl (to a final concentration of  $0.5\text{ mol l}^{-1}$ ) and polyethylene glycol (to a final concentration of 15%) were added, and DNA was precipitated at  $4^{\circ}\text{C}$  for 30 min. DNA was pelleted by centrifugation at  $12\text{ }000\text{ g}$  for 2 min at  $4^{\circ}\text{C}$  and resuspended in 300  $\mu\text{l}$  of sterile distilled water. Samples were heated at  $90^{\circ}\text{C}$  for 30 s and vortexed. Phenol/chloroform (1:1) (300  $\mu\text{l}$ ) was added, and samples were vortexed for 20 s before centrifugation at  $12\text{ }000\text{ g}$  at  $4^{\circ}\text{C}$  for 3 min. 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 of  $10 \text{ mol l}^{-1}$  ammonium acetate and  $1 \mu\text{l}$  of GenElute linear polyacrylamide (Sigma-Aldrich, Gillingham, UK) and incubated at  $-20^\circ\text{C}$  for 25 min. DNA was pelleted by centrifugation at  $12\,000 \text{ g}$  at  $4^\circ\text{C}$  for 5 min. Pelleted DNA was then washed three times in 70% ethanol, dried and resuspended in  $50 \mu\text{l}$  of sterile distilled water. DNA extracts were quantified using the Picodrop Microtitre Spectrophotometer (GRI, Braintree, UK).

#### Quantitative PCR

All 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 (Shannon *et al.*, 2007; Rogers *et al.*, 2010). Details of the relevant primers and probes used are shown in Supplementary Table S1.

#### Pyrosequencing

Bacterial tag-encoded FLX amplicon pyrosequencing 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 utilised Roche 454 FLX instrument (Roche, Branford, CT, USA) with Titanium reagents, titanium procedures performed at the Research and Testing Laboratory (Lubbock, TX, USA) using RTL protocols ([www.researchandtesting.com](http://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 and sequences with homopolymers  $>6$  bp 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 NET algorithm that utilises Blastn+ (Kraken BLAST; [www.krakenblast.com](http://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 was performed as described previously (Dowd *et al.*, 2008).

#### Bacterial identification

Based upon the above BLASTn+ -derived sequence identity (percent of total length query sequence that 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-characterised 16S sequences,  $>97\%$  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 77–80% at phyla. After resolution based upon these parameters, the percentage of each bacterial identity was individually analysed 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, a finite (without replacement) single-sample rarefaction was employed and calculated using the Species Diversity and Richness package (version 4.1.2; Pisces Conservation, Lymington, UK).

Meta-analyses were used to summarise the effect sizes of all 30 samples using Hedges' *d* effect size measure (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 (that is, standardized mean difference ( $\delta$ )) among treatments was expressed as:

$$\delta = \frac{\bar{X}_1 - \bar{X}_2}{\left( \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \right)}$$

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 s.d. within groups. The s.e. of mean per sample (s.e.m. <sub>$\delta$</sub> ) was expressed as:

$$\text{s.e.m.}_{\delta} = \frac{\sqrt{V}}{\sqrt{\frac{(n_1 + n_2)}{2}}}$$

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

Subsequently, the overall average effect size *M* was calculated, weighted by variance of sample *i* ( $V_i$ ) and

the s.e. of mean of  $M$  (s.e.m.<sub>M</sub>) as:

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

$$\text{s.e.m.}_M = \frac{\sqrt{\frac{1}{\sum_{i=1}^{30} \frac{1}{V_i^2}}}}{\sqrt{30}}$$

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 randomisation method of Solow (1993). It is known that pairwise 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 pairwise 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 s.d. 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 pairwise comparison.

The community compositions from paired samples were compared pairwise 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 re-sample 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 s.d. were taken. For metacommunity level analyses, the smallest sample size was chosen for each pairwise 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 30 spontaneously expectorated sputum samples to determine the abundance of bacteria in paired PMA-treated and non-treated sputum samples (Figure 1a). Mean bacterial abundance

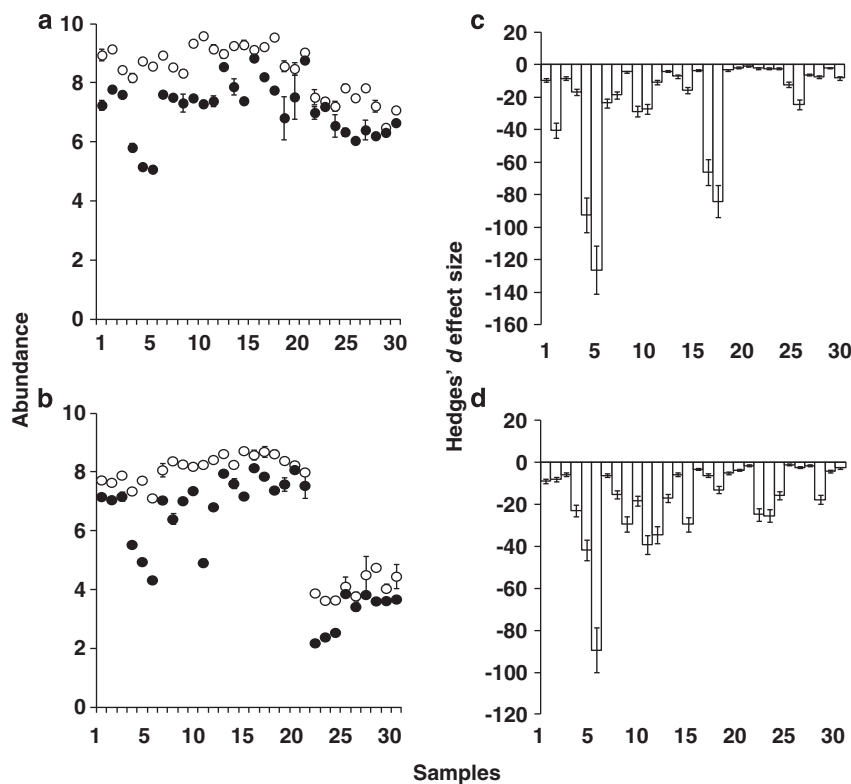
across non-PMA samples was  $9.05 \times 10^8 \pm 1.06 \times 10^9$  CFU ml<sup>-1</sup> equivalents (s.d. of the mean,  $n=30$ ), ranging between  $3.27 \times 10^6$  and  $4.03 \times 10^9$  CFU ml<sup>-1</sup> equivalents. In contrast, bacterial abundance across PMA-treated samples was  $8.91 \times 10^7 \pm 1.79 \times 10^8$  CFU ml<sup>-1</sup> equivalents ( $n=30$ ), ranging between  $1.22 \times 10^5$  and  $7.50 \times 10^8$  CFU ml<sup>-1</sup> equivalents. 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 1b). The mean abundance of *P. aeruginosa* in non-PMA samples was  $1.57 \times 10^8 \pm 1.82 \times 10^8$  CFU ml<sup>-1</sup> equivalents compared with  $2.50 \times 10^7 \pm 1.82 \times 10^7$  CFU ml<sup>-1</sup> equivalents in the PMA-treated samples, ranging from  $4.47 \times 10^3$  to  $6.15 \times 10^8$  and  $1.64 \times 10^2 \pm 1.49 \times 10^8$  CFU ml<sup>-1</sup> equivalents for non-PMA and PMA-treated samples, respectively.

A growing number of studies have observed that bacterial communities are highly variable between CF patients (for example, Rogers *et al.*, 2005, Harris *et al.*, 2007, van der Gast *et al.*, 2011). To cover the large variation among samples, meta-analysis (Borenstein *et al.*, 2009) was used to integrate effect sizes among sputum samples, that is, 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, this was translated into using the replications within the same sample as a study, and so all paired samples as independent studies, hence, providing 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 (Figures 1c and d).

### *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 386 213 bacterial sequence reads (mean =  $6436.9 \pm 634.2$  ( $n=60$ )), identifying 154 genera and 267 taxa, was 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



**Figure 1** The effect of PMA treatment on total bacteria and *P. aeruginosa* abundance. Given are (a) Bacterial and (b) *P. aeruginosa* abundance (log<sub>10</sub> scale CFU ml<sup>-1</sup> equivalents) measured by Q-PCR for paired samples that have been PMA-treated (closed circles) and not treated (open circles). Error bars represent the s.d. of the mean ( $n=3$ ). Also, given is the meta-analysis of (c) bacterial and (d) *P. aeruginosa* abundance using Hedges'  $d$  effect size measure between PMA-treated and non-treated paired samples. Columns represent effect size and error bars represent the s.e. ( $v_i$ ) 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.

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 undersampled was obtained.

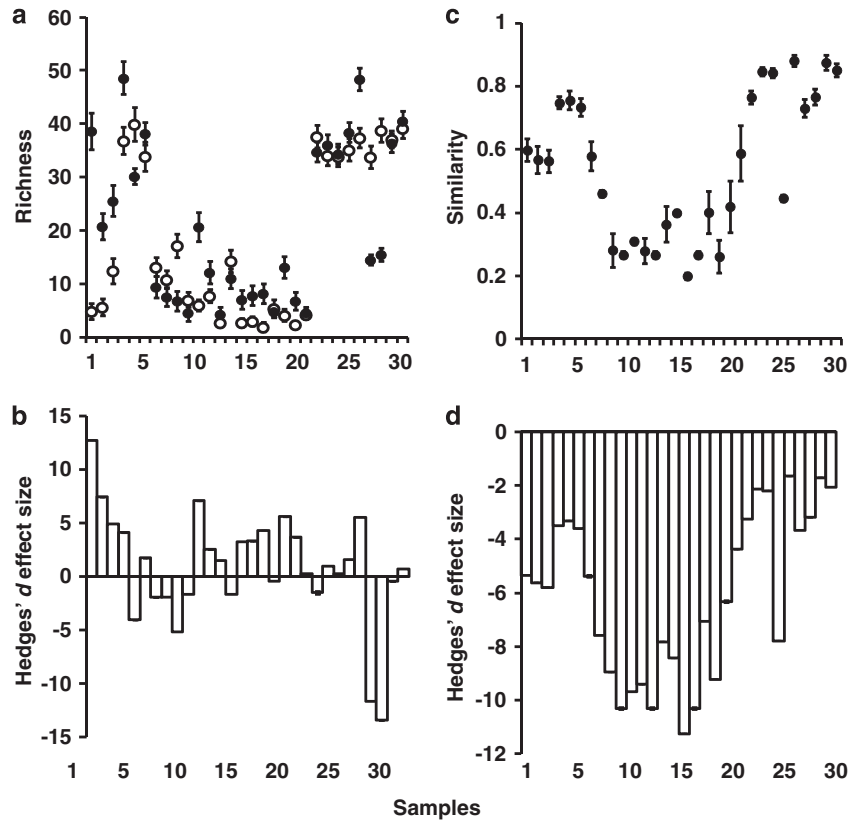
To avoid potential biases in comparisons of diversity between local communities due to varying number of sequences per sample (Gihring *et al.*, 2012), a randomised re-sampling method (with a uniform sub-sample size of  $n=1637$ ) using three indices of diversity (taxa richness ( $S^*$ ), Shannon–Wiener 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. Richness across individual samples, regardless of treatment, was highly variable, with mean values of  $21.0 \pm 15.1$  and  $18.8 \pm 15.3$ , respectively. Meta-analysis of individual samples revealed both significant negative and positive effects on richness by PMA treatment (Figure 2b).

Changes in community composition because of 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  $0.46 \pm 0.23$  in PMA-treated samples when compared with paired non-PMA-treated samples (Figure 2c).

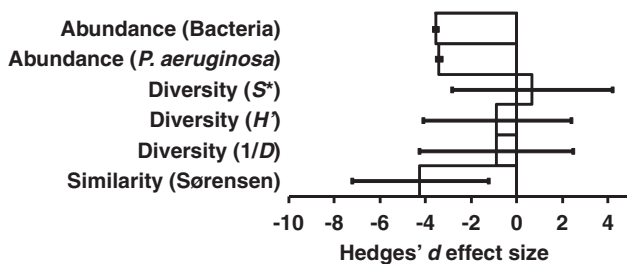
Meta-analysis showed significant negative effects (decreases in similarity) for all individual paired samples (Figure 2d).

#### Overall effects of PMA treatment on abundance, diversity and composition

When the overall effects of PMA treatment on different community characteristics were examined, a significant reduction was observed for both bacterial and *P. aeruginosa* 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 be otherwise incorporated into abundance measures using Q-PCR without prior PMA treatment. Although no overall significant effect on richness was observed, and hence diversity (confirmed using  $H'$  and  $1/D$  measures of diversity (Figure 3)), this indicated that diversity can be either under- or overestimated in local communities without prior PMA treatment (Figures 1 and 3). An overall significant negative effect on community composition was observed using the Sørensen index of similarity (Figure 3). This would indicate that skewed observations of local community composition are reached without prior PMA treatment.



**Figure 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 with paired non-treated samples (assuming a similarity of one 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 s.d. 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 s.e. ( $v_i$ ) of the effect size ( $n = 2000$ ). The s.e. bars that cross zero indicate no significant effect by PMA treatment.



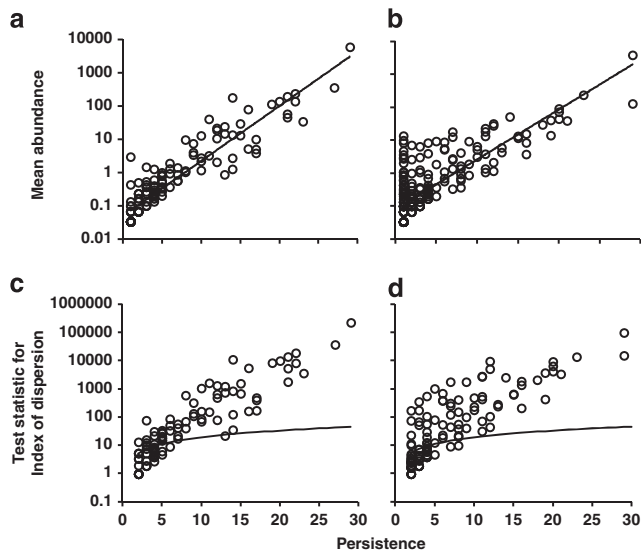
**Figure 3** Overall meta-analysis for measures of abundance, diversity and community similarity using Hedges'  $d$  effect size measure between all PMA-treated versus non-treated samples. Columns represent effect size and error bars represent the s.e. ( $v_i$ ) of the effect size ( $n = 30$ ). The s.e.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

#### The effects of PMA treatment at the metacommunity level

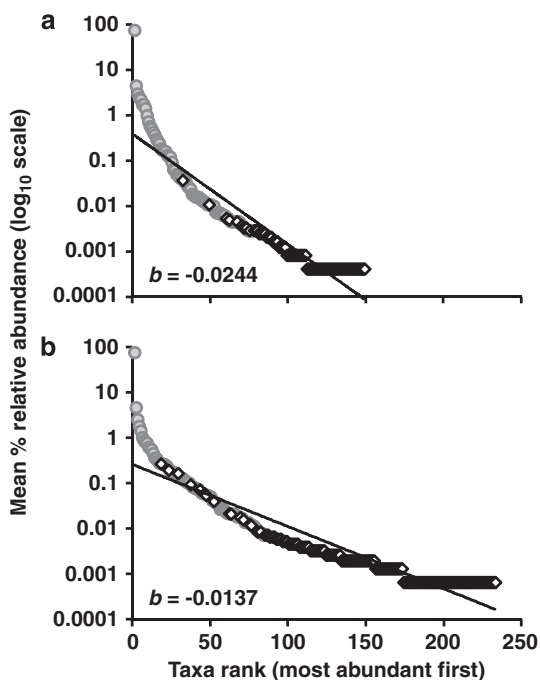
At the metacommunity level, it has been established that a dichotomy exists within metacommunity species abundance distributions (SADs), consisting of 'regionally' common and locally abundant core species and infrequent rare satellite species groups (for example, Magurran and Henderson, 2003;

Ulrich and Zalewski, 2006; Dolan *et al.*, 2009; 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). It is important to note that here we define 'core' and satellite species group set within a long established and defined ecological context (for example, Hanski, 1982; Magurran and Henderson, 2003), which is specific to the particular metacommunity being studied (comprised of samples from distinct local communities). Therefore, it is possible that core and satellite groups can differ between defined metacommunities and studies.

In the current study, the same approach was employed 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, the relationship between mean abundance of taxa and local community occupancy/persistence was plotted (Figures 4a and b). The expectation for a coherent metacommunity would be a 'regional'



**Figure 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)  $r^2 = 0.86$ ,  $F_{1,147} = 679.4$ ,  $P < 0.0001$  and (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  $\chi^2$  distribution. The 97.5% confidence limit was not plotted, as no taxon fell below that line.



**Figure 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)  $r^2 = 0.88$ ,  $F_{1,231} = 1669.1$ ,  $P < 0.0001$ .

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 (that is, by dispersal from outside of the metacommunity) (Prosser *et al.*, 2007), resulting in a positive relationship between distribution among sites and abundance. Here significant distribution–abundance relationships were observed for both the PMA- and non-PMA-treated metacommunity (Figures 4a and 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, the treated and non-treated metacommunity SADs were objectively divided 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  $\chi^2$  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 (that is, those taxa that fall  $< 2.5\%$  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 were classified as core group taxa, in the non-PMA-treated metacommunity (see Supplementary Table S3 for core and satellite identification). For the PMA-treated metacommunity, the satellite group was comprised of 163 taxa (with 112 taxa, each occurring in only one sample) and 70 core group taxa (Figure 4d).

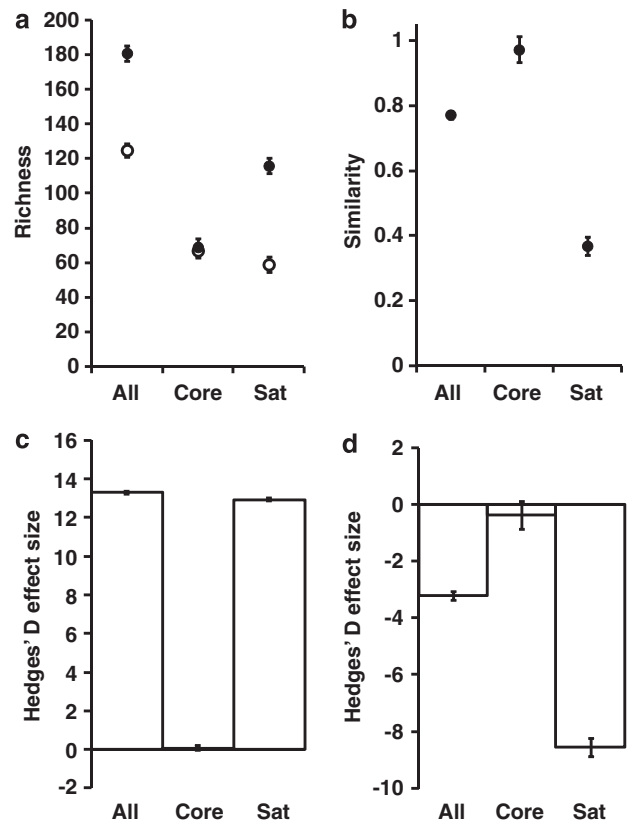
Following the categorisation of the CF bacterial taxa into the two groups, the metacommunity SADs with and without prior treatment were visualised as rank-abundance plots to examine differences in diversity and metacommunity structure (Figure 5). The slope values for each plot were used as a descriptive statistic of evenness, where a value of zero represents perfect evenness (Ager *et al.*, 2010). A change in community structure was observed in the PMA-treated metacommunity ( $b = -0.0137$ ) when compared with the non-treated metacommunity ( $b = -0.0244$ ). Using the *t*-distribution test, Fowler *et al.* (1998) demonstrated that the slopes were significantly different ( $t = 12.86$ ;  $d.f. = 1378$ ;  $P < 0.0001$ ), indicating an increase in community structure evenness following PMA treatment.

Furthermore, the changes in evenness were driven by changes in richness, where 149 taxa were observed in the non-PMA-treated metacommunity compared with 233 taxa in the PMA-treated metacommunity (Figure 5). To address the issue of whether previously undetected extant taxa, without prior PMA treatment, were either core (regionally common and locally abundant) or satellite (infrequent and rare) members of the metacommunity. A total of 118 taxa were previously undetected, of which 106 were categorised as satellite group members and 12 taxa as core group members. Furthermore, a total of 34 taxa, comprised of 31 satellite group and only 3 core group members, was not detected after prior PMA treatment. Despite the inclusion of a small number of core taxa group members, it could be concluded that the majority of taxa, both newly detected and undetected after PMA treatment, were predominately infrequent rare members of the metacommunity.

Subsequently, to what extent this drove overall patterns of diversity and composition of the metacommunity was examined, by comparing mean richness and changes in composition for the whole metacommunity, and the core and satellite groups between treatments (Figure 6a). Again to avoid potential biases in comparisons of diversity and composition between samples because of varying mean numbers of sequences, a randomised re-sampling method was used (with a uniform subsample size, using the lowest number of sequences in each pairwise comparison for the metacommunity ( $n=8172$ ), core ( $n=6370$ ) and satellite ( $n=214$ ) taxa groups). Richness was significantly higher in the PMA-treated metacommunity when compared with the non-treated metacommunity, confirmed using meta-analysis (Figure 6c). When partitioned into groups, PMA treatment did not lead to a significant change in core group richness, but a significant change in richness was observed in the satellite group, again demonstrating that the majority of previously undetected taxa belonged to the satellite group, which drove the changes in diversity of the overall metacommunity (Figures 6a and 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  $0.23 \pm 0.01$ ) and the satellite group ( $0.63 \pm 0.03$ ). Conversely, the composition of the core group was not significantly affected by PMA treatment ( $0.03 \pm 0.04$ ) (Figures 6b and 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.



**Figure 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 Figures 1 and 2. Also, given is the meta-analysis of (c) richness and (d) similarity using Hedges' *d* effect size measure. Columns represent effect size and error bars represent the s.e. ( $v_i$ ) of the effect size ( $n=2000$ ). The s.e. bars that cross zero indicate no significant effect by 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) of 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 reads) with PMA. In contrast, the abundance and distribution of *S. pneumoniae* decreased from 14 (408 reads) to 7 (36 reads). *Staphylococcus aureus* (two samples) and the candidate pathogen *Streptococcus intermedius* (part of the *Streptococcus milleri* group (three 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 three and one local communities, respectively. Members



of the genus *Mycobacterium*, some of which are regarded as pathogens in CF, were also found to belong to the satellite group, being detected in one local community. Conversely, members of the *Burkholderia* genus that were found to belong to the core taxa group (seven 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. To assess the effectiveness of such interventions, it is important 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 important 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 NVCE treatment on community characteristics was therefore of key importance in the present study. Our data indicate that contribution of DNA from non-viable sources prevents the accurate characterisation of CF bacterial communities by pyrosequencing.

We demonstrate that suppressing the contribution of DNA from non-viable sources is an important consideration 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.

## Acknowledgements

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

## 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, McKeen 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). *Practical 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 satellite 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 techniques. *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 TW, 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.

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)