ORIGINAL ARTICLE

Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities

Christopher J van der Gast[1], Alan W Walker[2], Franziska A Stressmann[3], Geraint B Rogers[3], Paul Scott[2], Thomas W Daniels[4], Mary P Carroll[4], Julian Parkhill[2] and Kenneth D Bruce[3]

[1]*NERC Centre for Ecology and Hydrology, Wallingford, UK;* [2]*Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK;* [3]*Pharmaceutical Science Division, Molecular Microbiology Research Laboratory, King's College London, London, UK and* [4]*Cystic Fibrosis Unit, Southampton University Hospitals NHS Trust, Southampton, UK*

Cystic fibrosis (CF) patients suffer from chronic bacterial lung infections that lead to death in the majority of cases. The need to maintain lung function in these patients means that characterising these infections is vital. Increasingly, culture-independent analyses are expanding the number of bacterial species associated with CF respiratory samples; however, the potential significance of these species is not known. Here, we applied ecological statistical tools to such culture-independent data, in a novel manner, to partition taxa within the metacommunity into core and satellite species. Sputa and clinical data were obtained from 14 clinically stable adult CF patients. Fourteen rRNA gene libraries were constructed with 35 genera and 82 taxa, identified in 2139 bacterial clones. Shannon–Wiener and taxa-richness analyses confirmed no undersampling of bacterial diversity. By decomposing the distribution using the ratio of variance to the mean taxon abundance, we partitioned objectively the species abundance distribution into core and satellite species. The satellite group comprised 67 bacterial taxa from 33 genera and the core group, 15 taxa from 7 genera (including *Pseudomonas* (1 taxon), *Streptococcus* (2), *Neisseria* (2), *Catonella* (1), *Porphyromonas* (1), *Prevotella* (5) and *Veillonella* (3)], the last four being anaerobes). The core group was dominated by *Pseudomonas aeruginosa*. Other recognised CF pathogens were rare. Mantel and partial Mantel tests assessed which clinical factors influenced the composition observed. CF transmembrane conductance regulator genotype and antibiotic treatment correlated with all core taxa. Lung function correlated with richness. The clinical significance of these core and satellite species findings in the CF lung is discussed.
The ISME Journal advance online publication, 9 December 2010; doi:10.1038/ismej.2010.175
Subject Category: microbial population and community ecology
Keywords: commonness and rarity; oral microbiota; polymicrobial infections; pulmonary disease; species abundance distributions

Introduction

Cystic fibrosis (CF) is an autosomal genetic disorder affecting more than 8500 patients in the United Kingdom alone (Anon, 2009). Although this genetic disorder affects many body systems, its impact on the lungs of CF patients is the most clinically significant. Here, the genetic defect results in the absence or abnormal functioning of the CF transmembrane conductance regulator (CFTR) protein, involved in ion transfer across epithelial cell surfaces. In the lower airways, this leads to an increased viscosity of airway secretions and, in turn, an impaired mucociliary clearance of material entering the lungs (Clunes and Boucher, 2007;

Coakley and Boucher, 2007). Together, these factors make the CF lower airways a favourable anatomical niche for bacterial persistence. Exposure to a diverse range of bacterial species, through constant ventilation and close proximity to the communities of the upper respiratory tract and oral cavity, rapidly results in the establishment of a diverse bacterial community (Rogers *et al.*, 2005). The damage that results from the inflammatory response to the presence of bacteria in the lower airways is the primary cause of morbidity and mortality for CF patients (Balfour–Lynn and Elborn, 2007), and despite substantial improvements in treatment, the predicted survival of UK CF patients is 35.2 years (Anon, 2009). Achieving a better understanding of the way in which infective bacterial communities in the CF lower airways are organised is, therefore, of utmost clinical importance. Further, the large CF patient group, along with the ability to sample non-invasive and well-defined treatment conditions, makes CF lung infections an ideal model

system from which one can gain insight into other chronic respiratory infections more generally. The importance of this is emphasised further, given that respiratory illness is one of the leading global causes of human morbidity and mortality (Mathers *et al.*, 2006).

Traditionally, the microbiological analyses of clinical specimens have relied on cultivation before identification, where routine screening of CF sputum samples has typically focused on only a few bacterial species, including *Pseudomonas aeruginosa*, *Burkholderia cepacia* complex, *Haemophilus influenzae*, *Staphylococcus aureus*, *Moraxella catarrhalis* and *Streptococcus pneumoniae* (Anon, 2008). Increasingly though, approaches and methods developed to address environmental microbiological questions are being adopted to study CF respiratory microbiology (Rogers *et al.*, 2003; Sibley *et al.*, 2008). The adoption of culture-independent methodologies has, perhaps unsurprisingly, revealed the presence of much more complex bacterial communities than previously realised (Rogers *et al.*, 2005, 2009). In detail though, such studies have shown a greater diversity of anaerobic and aerobic species as present in CF sputa and, importantly, the probable presence of species from the oral cavity that may represent new candidate pathogens (Rogers *et al.*, 2006; Harris *et al.*, 2007; Sibley *et al.*, 2008). There is, therefore, now a pressing need to distinguish between those newly reported species that are true colonisers of the CF lower airways, and as such of potential clinical importance, and those introduced to the airways that are only present transiently.

Work on CF respiratory microbiology has also led to the proposal of other important concepts. Extending from observations of bacterial diversity in the CF lower airways, these infections have recently been proposed to be driven by communities of interacting organisms (Rogers *et al.*, 2010a). As such, the chronically colonised CF airways represent a previously unrealised complex and diverse ecosystem (Harrison, 2007). More than this though, the advantages in structuring subsequent research within an ecological framework are clear. Two recent studies of CF respiratory microbiology have notably coupled environmental microbiological approaches with community ecological analyses (Kelpac-Ceraj *et al.*, 2010; Rogers *et al.*, 2010b). Based on species–time relationships, Rogers *et al.* (2010b) demonstrated that reliance on a single respiratory sample was insufficient for the detection of recognised CF pathogens. Kelpac-Ceraj *et al.* (2010) studied bacteria sampled from a paediatric CF patient cohort, and suggested that community composition may well be a better predictor of disease progression than the traditional approach of tracking particular recognised CF pathogens.

Both these strands argue strongly for the need to introduce relevant ecological approaches, with biodiversity being the first term to consider. Hubbell (2001) defined biodiversity as being synonymous with species richness and relative species abundance in space and time, where relative species abundance refers to the commonness or rarity of a species in relation to other species in a given community. As such, the question of what makes a species common or rare has long been of empirical and theoretical interest in community ecology (Preston, 1948; Hubbell, 2001). There is also a long tradition in ecology, whereby species in spatially separated sites can be divided into core and satellite groups at smaller ecological scales (Hanski, 1982, 1999; Ulrich and Zalewski, 2006). Here, the former group comprises species that are widely distributed and typically abundant within local sites and the latter group consists mostly of rare species occurring in low abundance at a limited number of sites. This distinction between frequent and occasional species groups has long been discussed in a metapopulation context (Ulrich and Zalewski, 2006). However, the core–satellite species distinction has been more recently extended to metacommunities, in terms of either spatial (site occupancy) or temporal (persistence) distributions (Magurran and Henderson, 2003; Ulrich and Ollik, 2004; Ulrich and Zalewski, 2006).

Using a 21-year time-series data set, Magurran and Henderson (2003) investigated the relative species abundance of fish species from Hinkley Point in the Bristol Channel (UK). They determined that the behaviour of species, in terms of their permanence in the sampling record, affected their signature on the empirical species abundance distribution (SAD) (Magurran and Henderson, 2003; Magurran, 2007). They found that the satellite species were typically rare and in low abundance when they did occur and the core species, fish that were commonly recorded year after year, had higher abundances. The temporal metacommunity was readily divided into two categories, where core species followed a log-normal distribution, whereas the SAD for the satellite species resembled a log-series distribution. Furthermore, they noted that the two groups of species had differing ecological characteristics; core species were typically adapted to life in estuarine habitats, whereas satellite species were better suited to life in other habitats such as deep water (Magurran and Henderson, 2003; Magurran, 2007). Ulrich and Zalewski (2006), in a spatial study of ground beetles on small lake islands in Poland, also found that the beetle metacommunity SAD could be partitioned into core and satellite species. They concluded that the division was not a sampling artefact but reflected life history strategies, as species differed in their spatial distribution and body size ratios. Both studies demonstrated that biological factors underpinned the relative abundance of the core species, whereas random dispersal was more important in structuring the satellite species (Magurran and Henderson, 2003; Ulrich and Zalewski, 2006; Magurran, 2007). These and

other studies (Ulrich and Ollik, 2004; Gray et al., 2005; Dolan et al., 2009) have strengthened the proposal that a community comprises core and satellite species, and that partitioning the two groups from a (spatial or temporal) metacommunity reveals important aspects of SADs, which would otherwise be neglected without such a distinction. Furthermore, both Gray et al. (2005) and Magurran and Henderson (2003) have advocated that a two-group model is the most parsimonious way to deal with empirical SADs.

In the current study, we applied the two-group core–satellite modelling approach to the bacterial communities present in respiratory samples from a cohort of 14 CF adult patients, in order to address the following aims: (1) establish whether the 'traditionally' recognised CF pathogens were part of the core species group or, alternatively, were randomly dispersed throughout the patients sampled; (2) determine how species, regarded as belonging to the human oral cavity, were distributed between the two groups; and (3) ascertain how anaerobic and aerobic species were distributed throughout the empirical SAD. These findings then allowed the determination of which clinical factors correlated with community characteristics, such as richness and composition, as applied for the whole community and the core and satellite species groups derived.

Materials and methods

Spontaneously expectorated sputum samples were collected from 14 adult CF patients attending the Adult CF Clinic at the Southampton General Hospital, under full ethics approval (Southampton and South West Hampshire Research Ethics Committee (06/Q1704/26)). At the time of sampling, all patients were clinically stable and had not received any antibiotic therapy in the 14 days before enrolment. Clinical parameters, including lung function (forced expiratory volume (litres) in 1 s), CF genotype, recent antibiotic history, temperature, patient-reported measures of cough severity, sputum productivity, breathlessness and general well-being, were recorded at the time of samples collection (Table 1).

All 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 before analysis. Before DNA extraction, sputum samples were washed three times in 0.9% phosphate-buffered saline to remove adherent saliva. The DNA extraction was performed as previously described by Rogers et al. (2004). 16S rRNA genes were amplified by PCR using primers, Bact-7F (5′-AGAG TTTGATYMTGGCTCAG-3′) and Bact-1510R (5′-AC GGYTACCTTGTTACGACTT-3′), as previously described (Stecher et al., 2007). The amplification

Table 1 Clinical characteristics for individual patients

| Patient | Age (years) | Gender | CFTR genotype | FEV_1 | BMI | Temp[a] | Diabetes | Antibiotic therapy | | | | | Mucolytic agents | | Corticosteroids | |
								Macrolide	Colomycin[b]	Tobramycin[b]	Colomycin and Tobramycin[b]	Itraconazole	DNAse	Hypertonic saline	Oral	Inhaled
1	30	M	ΔF508/NK	2.51	29.0	37.4	N	N	N	N	N	N	Y	N	N	Y
2	45	F	ΔF508/NK	1.30	18.2	36.5	Y	Y	Y	N	N	N	N	N	N	Y
3	47	M	ΔF508/NK	1.23	19.9	35.9	Y	Y	Y	N	N	N	Y	N	N	Y
4	30	F	ΔF508/711+3A7G	1.30	27.1	37.3	N	Y	Y	N	N	Y	Y	N	N	Y
5	22	F	ΔF508/ΔF508	1.23	18.0	37.8	N	Y	Y	N	N	N	Y	N	N	Y
6	55	M	ΔF508/G58E	1.59	23.9	36.5	N	Y	Y	N	N	N	Y	N	N	Y
7	21	F	ΔF508/ΔF508	1.68	20.3	36.5	Y	Y	Y	N	N	N	N	N	Y	Y
8	40	M	ΔF508/ΔF508	1.91	19.4	37.0	Y	Y	Y	N	N	N	Y	Y	N	Y
9	22	M	ΔF508/ΔF508	0.73	18.4	36.2	N	Y	N	N	Y	N	Y	N	N	Y
10	17	F	ΔF508/ΔF508	2.86	22.5	37.3	N	Y	Y	N	N	N	Y	N	N	Y
11	24	F	ΔF508/G542X	2.17	21.0	36.9	N	Y	N	Y	N	Y	N	N	N	N
12	20	M	ΔF508/ΔF508	1.19	20.4	36.9	N	N	N	N	N	N	Y	N	N	Y
13	20	M	ΔF508/ΔF508	2.10	28.5	37.5	Y	N	Y	N	Y	Y	Y	N	Y	Y
14	23	M	ΔF508/ΔF508	3.21	35.9	35.9	Y	N	N	Y	N	N	N	N	N	Y

Abbreviations: BMI, body mass index (kg m⁻²); CFTR, cystic fibrosis transmembrane conductance regulator; FEV₁, forced expiratory volume (litres) in 1 s; N, no; NK, genotype not known; Y, yes.
[a]Internal body temperature (°C) at the time of sampling.
[b]Antibiotics were delivered by a nebuliser.

products were essentially full-length 16S rRNA genes, ranging from 1202 to 1517 in length, with a median length of 1475 base pairs.

Clone library construction and sequencing were carried out as described previously (Stecher et al., 2007). Colonies were randomly selected for sequencing from agar plates for each patient. Sequences were aligned using the NAST aligner (DeSantis et al., 2006) and these alignments were subject to extensive manual curation using the ARB package (Ludwig et al., 2004) before further analysis. Sequences were tested for chimeras with Mallard (Ashelford et al., 2006), Bellerophon at Greengenes (DeSantis et al., 2006) and Pintail (Ashelford et al., 2005) and any sequence that appeared to be chimeric was removed. After removal of chimeras and other suspect sequences, the remaining sequences (deposited in GenBank under accession numbers FM995625–FM997761) were initially given a broad classification to the phylum level using the Classifier tool at the RDPII website (Wang et al., 2007). To obtain more detailed taxonomic information, the sequences were divided into phylotypes. Distance matrices were then entered into the DOTUR program (Schloss and Handelsman, 2005) set to the furthest neighbour and 99%-similarity setting. The resulting phylotypes were then assigned similarities to nearest neighbours using MegaBLAST.

Two complementary measurements of diversity were used to determine whether the clone sample sizes were large enough to effectively assess the diversity of the bacteria in each of the sputum samples taken, as previously described (Lilley et al., 1996). The indices used were the Shannon–Wiener index (H') and species richness (S^*). H' is a combined figure that reflects the extent of diversity and the evenness of isolate distribution between taxa. It is sensitive to changes in the frequency of common and less common (though not the rarer) species. S^* is simply the number of species in a sample. Differences in H' and S^* were tested using the method of Solow (1993), which employs a randomisation test as follows: the clones derived from two samples (A and B) are listed, the chosen population parameter (for example, H') calculated and the difference in parameters for the two samples noted ($D_{A-B} = H'_A - H'_B$) (Solow, 1993). The clones in the two samples are then amalgamated, randomly mixed and repartitioned to the original sample sizes, the population parameter recalculated and a new difference noted. This random mixing, repartitioning and calculation of the difference are repeated 1000 times and the values ($D_1–D_{1000}$) ranked. Finally, to test whether H'_A and H'_B are significantly different at the 95% confidence level, D_{A-B} is compared with the ranked values. If this difference is greater than or equal to the value of the difference at the 97.5 percentile, or less than or equal to the value of the difference at the 2.5 percentile, then the values for the parameter of the two samples are

considered significantly different. An Excel macro-program was written to apply the Solow's method to the pair-wise comparison of each of the two parameters (S^* and H').

The log-series and log-normal models were fitted to the data, as previously described (Magurran, 2004). Likewise, the χ^2 and Kolmogorov–Smirnov goodness-of-fit tests were performed as described by Magurran (2004). Poisson distribution tests were carried out according to the method described by Krebs (1999). Regression analysis, coefficients of determination (r^2), residuals and significance (P) were calculated using Minitab software (version 14.20, Minitab, University Park, PA, USA). The Bray–Curtis (quantitative) index of similarity and subsequent average linkage clustering of community profiles were performed using the PAST (Palaeontological Statistics, version 1.90) program available from the University of Oslo website link (http://folk.uio.no/ohammer/past) run by Øyvind Hammer. Mantel and partial Mantel tests were performed, as previously described (Green et al., 2004), using the XLSTAT (version 2006, Addinsoft, Paris, France) program.

Results and discussion

A total of 2139 bacterial clones, comprising 35 genera and 82 taxa, were generated from sputum samples taken from 14 adult CF patients (Table 2). To test whether clone sample sizes per patient were sufficiently large to compare enough of the bacterial diversity in CF patient airways, we used a randomised re-sampling method using two indices of diversity (Shannon–Wiener index (H') and taxa richness (S^*)) (Supplementary Figure S1). These analyses provided confirmation that sufficiently large samples had been collected and bacterial diversity was not undersampled.

Positive relationships between mean abundance and distribution (number of sites occupied) have been observed at many spatial scales for taxa when classified into different types of ecological organisation (for example, guild or community) (Guo et al., 2000). Within the current study, we also observed a significant positive distribution–abundance relationship (Figure 1). This indicated that the bacterial taxa that were widely distributed throughout the CF airways sampled were more locally abundant than the taxa with a more restricted distribution. Therefore, as has been noted in studies of animal species, the commonness and rarity of bacterial taxa in the CF airway metacommunity were found to be related to their permanence (Magurran and Henderson, 2003).

Magurran and Henderson (2003) previously partitioned their empirical fish SAD into two groups of persistent/abundant core species and infrequent/less abundant satellite species by searching for a discontinuity in the SAD. Here, we used an alternative method proposed by those authors as a

Table 2 Bacterial taxa sampled from the lungs of the 14 cystic fibrosis patients

Class	Family	Taxon name	Code
Actinobacteria	Actinomycetaceae	*Actinomyces graevenitzii* (98%)	S, Ae, O
		Actinomyces lingnae (98%)	S, Ae, O
		Actinomyces naeslundii	S, Ae, O
		Actinomyces naeslundii/viscosus (98%)	S, Ae, O
	Micrococcineae	*Rothia dentocariosa* (97%)	S, Ae, O
		Rothia mucilaginosa	S, Ae, O
Bacteroides	Bacteroidaceae	*Bacteroides acidofaciens* (88%)	S, An
		Bacteroides eggerthii (88%)	S, An
		Bacteroides uniformis (87%)	S, An
		Bacteroides uniformis (88%)	S, An
		Bacteroides viscericola (86%)	S, An
	Porphyromonadaceae	*Porphyromonas catoniae* (92%)	C, An, O
		Tannerella forsythensis (94%)	S, An, O
	Prevotellaceae	*Prevotella loescheii* (98%)	S, An, O
		Prevotella melaninogenica	C, An, O
		Prevotella nanceiensis	S, An
		Prevotella nigrescens	S, An, O
		Prevotella oris (98%)	C, An, O
		Prevotella oulora	S, An, O
		Prevotella pallens	S, An, O
		Prevotella pleuritidis	S, An
		Prevotella salivae	C, An, O
		Prevotella tannerae	C, An, O
		Prevotella veroralis	C, An, O
	Rikenellaceae	*Alistipes shahii* (82%)	S, An, O
Flavobacteria	Flavobacteriaceae	*Capnocytophaga gingivalis*	S, Ae, O
		Capnocytophaga granulosa (98%)	S, Ae, O
		Capnocytophaga leadbetteri (98%)	S, Ae, O
		Capnocytophaga sputigena	S, Ae, O
Bacilli	Staphylococcaceae	*Gemella haemolysans*	S, Ae, O
		Staphylococcus aureus	S, Ae
	Aerococcaceae	*Abiotrophia defectiva*	S, Ae, O
	Carnobacteriaceae	*Granulicatella adiacens*	S, Ae, O
		Granulicatella para-adiacens	S, Ae, O
	Streptococcaceae	*Streptococcus infantis*	S, Ae, O
		Streptococcus intermedius/constellatus	S, Ae, O
		Streptococcus mitis	S, Ae, O
		Streptococcus mitis/pneumoniae	C, Ae, O
		Streptococcus oralis	S, Ae, O
		Streptococcus parasanguis	C, Ae, O
		Streptococcus parasanguis/oralis/mitis	S, Ae, O
		Streptococcus pneumoniae	S, Ae
		Streptococcus salivarius	S, Ae, O
		Streptococcus salivarius/vestibularis/thermophilus	S, Ae, O
		Streptococcus sanguinis	S, Ae, O
Clostridia	Clostridiaceae	*Clostridium metallolevans* (98%)	S, Ae
	Eubacteriaceae	*Mogibacterium neglectum*	S, Ae, O
	Lachnospiraceae	*Catonella morbi* (98%)	C, An, O
		Dorea longicatena	S, An
		Eubacterium sp. 'Smarlab BioMol-2301166'	S, An
		Eubacterium sp. 'Smarlab BioMol-2301231'	S, An
		Oribacterium sinus (98%)	S, An
		Butyrate-producer SR1/1 (97%)	S, An
	Veillonellaceae	*Veillonella atypica*	C, An, O
		Veillonella dispar	C, An, O
		Veillonella parvula	C, An, O
		Veillonella ratti (96%)	S, An, O
		Dialister invisus	S, An, O
		Megasphaera micronuciformis	S, An, O
		Selenomonas infelix	S, An, O
Fusobacteria	Fusobacteriaceae	*Fusobacterium naviforme/nucleatum*	S, An, O
		Fusobacterium nucleatum	S, An, O
		Leptotrichia trevisanii (97%)	S, An, O
Betaproteobacteria	Alcaligenaceae	*Achromobacter xylosoxidans/insolitus*	S, Ae
	Neisseriales	*Eikenella corrodens*	S, Ae, O
		Neisseria cinerea	C, Ae, O
		Neisseria elongata	S, Ae, O
		Neisseria flava/sicca/mucosa/pharyngis	C, Ae, O
		Neisseria flavescens (98%)	S, Ae, O

Table 2 (Continued)

Class	Family	Taxon name	Code
		Neisseria meningitidis (95%)	S, Ae, O
		Neisseria meningitidis (98%)	S, Ae, O
Epsilonproteobacteria	Campylobacteraceae	*Campylobacter concisus*	S, Ae, O
		Campylobacter gracilis	S, Ae, O
Gammaproteobacteria	Altermonadaceae	*Alishewanella fetalis*	S, Ae
	Pastuerellaceae	*Haemophilus haemolyticus* (98%)	S, Ae, O
		Haemophilus influenzae	S, Ae, O
		Haemophilus parahaemolyticus	S, Ae
		Haemophilus parainfluenzae (97%)	S, Ae
		Terrahaemophilus aromaticivorans	S, An
	Moraxellaceae	*Acinetobacter lwoffii*	S, Ae
	Pseudomonadaceae	*Pseudomonas aeruginosa*	C, Ae
	Xanthomonadaceae	*Stenotrophomonas maltophilia*	S, Ae

Abbreviations: C, core group taxon; S, satellite group; Ae, aerobe; An, anaerobe; O, taxa associated with the oral microbiota.
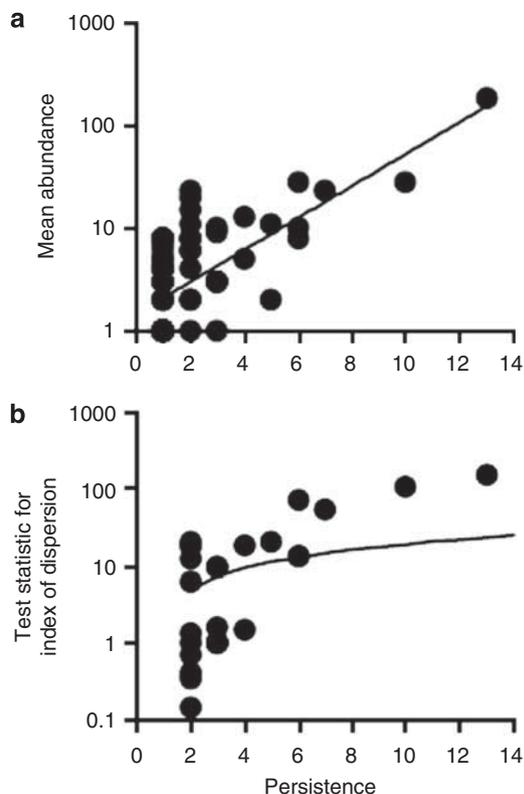
Figure 1 Distribution and dispersal of bacterial taxa across patients. (**a**) The number of patients for whom each bacterial taxon was observed, plotted against the mean abundance ($\log_{10}$ scale) across all patients ($r^2 = 0.44$; $F_{1,80} = 62.2$; $P < 0.0001$). (**b**) Random and non-random dispersal 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 14 patients sampled. 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.

way of objectively dividing empirical SAD into the two species groups. Specifically, we chose to first decompose the overall distribution using the ratio of

variance to the mean abundance for each bacterial taxon; as such, this approach provides a tractable and objective solution to partitioning a SAD into two groups of core and satellite species (Magurran and Henderson, 2003). The variance to mean ratio, or index of dispersion, is an index used to model whether the species follows a Poisson distribution, falling between the 2.5 and 97.5% confidence limits of the χ^2 distribution (Krebs, 1999). Plotting the indices of dispersion against persistence throughout the CF airways sampled, 14 bacterial taxa were randomly distributed through space, that is, those taxa that fall below the 2.5% confidence limit line (Figure 1). Bacterial taxa that occurred only in a single patient were excluded from this analysis, as their dispersion in space would have no variance. However, following Fisher's logic, 53 taxa would be included in the log-series distribution (Fisher *et al.*, 1943). For the purpose of the current study, those 67 taxa that were randomly distributed were classified as satellite species and the remaining 15 non-randomly distributed taxa were classified as core taxa, in the CF bacterial metacommunity (see Table 2 for core and satellite taxa identification).

Following categorisation of the CF bacterial taxa into the two groups, the satellite taxa data were fitted with the log-series distribution model, whereas the core group was fitted with the log-normal distribution model (Figure 2). Both the χ^2 and Kolmogorov–Smirnov goodness-of-fit tests were employed to evaluate whether the observed distributions were not significantly different from the expected distributions; for example, if $P > 0.05$ for both tests, then a fit could be assumed. For fitting of the log-series model to the satellite taxa, the Kolmogorov–Smirnov test statistic was $D = 0.072$, whereas the approximate critical value at the $P < 0.05$ level was greater, $D_{0.05} = 0.109$. Similarly, for fitting the log normal to the core group, the values obtained were $D = 0.152$ and $D_{0.05} = 0.230$. In addition, for both the log-series and -normal model fits, the χ^2 values were not significantly different at
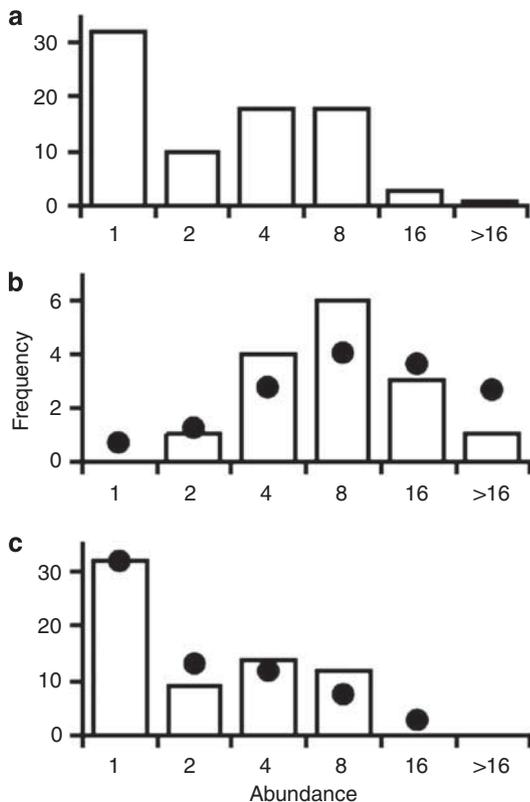
Figure 2 The abundance distributions for the bacterial metacommunity for (**a**) all taxa, (**b**) core taxa best predicted by the log-normal model ($\chi^2(4) = 2.71$; $P = 0.607$) and (**c**) satellite taxa best predicted by the log-series model ($\chi^2(4) = 4.61$; $P = 0.330$). The frequency of each $\log_2$ abundance class predicted by the log-normal and -series models is shown as a dot.

the $P < 0.05$ level (Figure 2). Therefore, as the observed distributions were not significantly different from the expected, the log-series and -normal models that had been fitted, respectively, to the satellite and core groups could be accepted. Conversely, no good fit was obtained for the overall empirical SAD by either of the two models ($P < 0.05$ in each case for both types of goodness-of-fit tests). It has been generally accepted that empirical SADs for large communities of animal or plant species tend towards the log normal (Preston, 1948; Gaston and Blackburn, 2000; Magurran and Henderson, 2003). However, this view has begun to change, as increasing lines of evidence have demonstrated that large communities are typically characterised by an excess of rare species, resulting in a negatively skewed distribution (Gaston and Blackburn, 2000; Hubbell, 2001; Magurran, 2004). Despite the CF metacommunity investigated here having only relatively low bacterial diversity compared with other systems such as soil, when the distribution for all CF bacterial taxa was plotted (Figure 2), or rather when the core and satellite group distributions were superimposed, a negatively skewed distribution characterised by rare taxa was the result.

The satellite group comprised 67 different bacterial taxa from 33 genera and accounted for only 11.1% of the total abundance (Table 2). From the satellite taxa, only 53 taxa were sampled from one individual patient. In contrast, the core group comprised 15 taxa from 7 genera (including *Catonella* (1 taxon), *Neisseria* (2), *Porphyromonas* (1), *Prevotella* (5), *Pseudomonas* (1), *Streptococcus* (2) and *Veillonella* (3)), accounting for 88.9% of the total abundance (Table 2). The core group was dominated by one taxon, *P. aeruginosa*, considered previously to be the predominant CF respiratory pathogen. This accounted for 70.6% of the total abundance and was found in all patients (except patient 1 (P1)) (Figure 3). With regard to the distribution of other recognised and candidate CF pathogens, the five recorded pathogens were randomly distributed throughout the patients and were therefore assigned to the satellite taxa group. Each of these CF pathogens was found only in low abundance and in single patients (*Achromobacter xylosoxidans* (P12, 5 clones), *H. influenzae* (P14, 5 clones), *Staphylococcus aureus* (P1, 1 clone), *Stenotrophomonas maltophilia* (P1, 4 clones) and *Streptococcus intermedius/constellatus* part of the *Streptococcus milleri* group (P10, 1 clone)). In total, these represented less than 0.75% of the total abundance. No taxa from the *B. cepacia* complex were recorded in any individual from the patient cohort.

In the current study, only strict anaerobes were classed as anaerobes, whereas aerobes, facultative anaerobes and microaerophiles were classed as aerobes (Table 2). In agreement with previous studies (for example, Rogers *et al.*, 2004; Tunney *et al.*, 2008), we also found strict anaerobic taxa to be diverse and abundant within the CF airways (Figure 3). That is, of the 82 taxa observed in the CF bacterial metacommunity, 38 were strict anaerobes. These anaerobic taxa were widely distributed throughout both the core and satellite groups, accounting for 15.2 and 5.4% of the total abundance, respectively. In terms of richness, the anaerobes within the core group were represented by 4 genera and 10 taxa (*Catonella*, *Porphyromonas*, *Prevotella* and *Veillonella*), and within the satellite group by 16 genera and 28 taxa (Figure 3 and Table 2). The colonisation and persistence of anaerobic taxa could be explained by the recent observations that oxygen gradients occur in CF lung mucus and proliferation of *P. aeruginosa* within the mucus can generate anaerobic conditions (Yoon *et al.*, 2002; Tunney *et al.*, 2008). Although anaerobic taxa have been implicated in the CF lung and their presence within the CF airways can be explained, their role in infection and inflammation needs to be further elucidated (Rogers *et al.*, 2004).

It has been suggested that the oral cavity can serve as a pathogen reservoir for a variety of respiratory infections (Scannapieco, 1999). Previously, a number of species, normally associated with the oral cavity
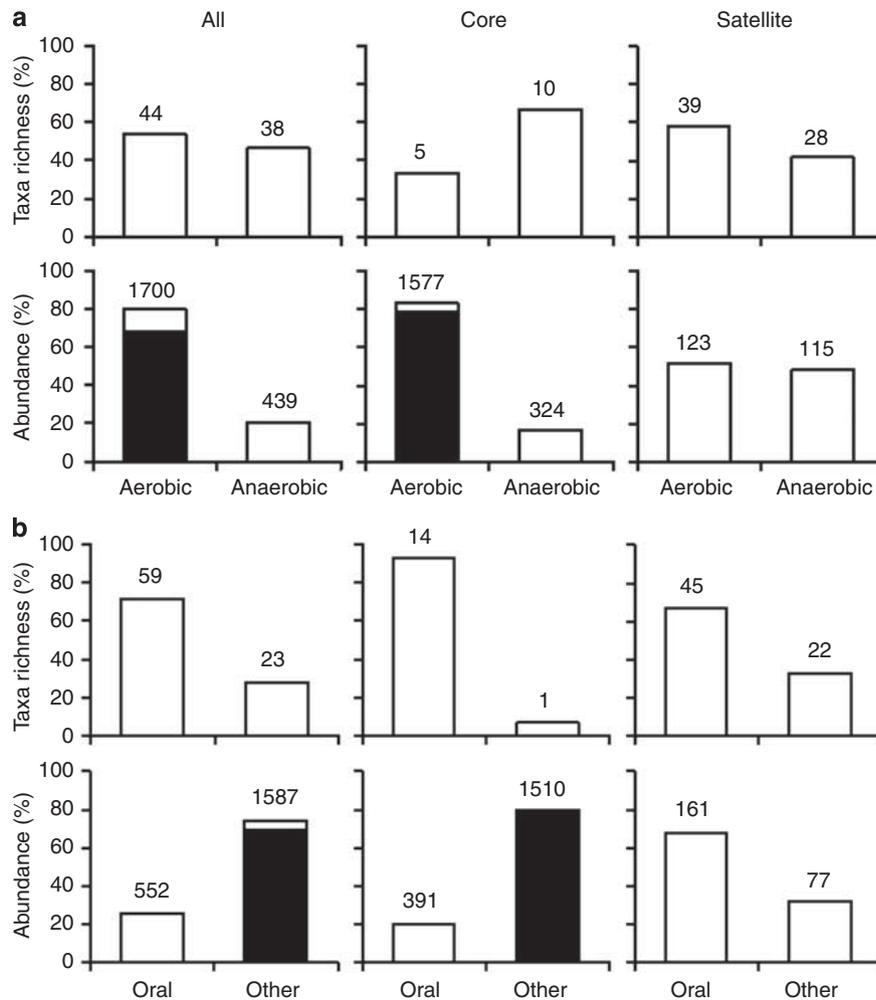
Figure 3 The richness and abundance of (**a**) aerobic and anaerobic bacteria and (**b**) oral microbiota within the whole metacommunity, the core group and the satellite group. Also given are the actual values of taxa richness and abundance (number of clones). Shaded areas represent percentage abundance of *Pseudomonas aeruginosa* within the whole metacommunity and the core group of bacterial taxa.

environment, have been identified in CF sputum samples (Rogers *et al.*, 2004, 2006). Through their presence in well-controlled studies, it is clear that these oral-associated species are not simply contaminants. In addition, Rogers *et al.* (2006) concluded that the oral cavity acts as both a reservoir and a 'stepping stone' for bacterial immigration into the CF lung. Our findings add support to that study, as bacterial taxa associated with the oral cavity were widely distributed across the empirical SAD and accounted for over 72% of the overall bacterial diversity observed (Figure 3; Table 2). This included 24 anaerobic and 35 aerobic bacterial taxa, and accounted for 18 and 7.9% of the total abundance, respectively. Within the core group, 'oral' bacterial taxa were dominant in terms of richness (6 genera and 14 taxa (Table 2)) but comprised only 20.6% of the abundance within that group (Figure 3). Again, for the satellite group, 'oral' bacterial taxa comprised over 67% of the richness (20 genera, 45 taxa) and over 75% of the abundance within that group. But being randomly distributed

and rare within the metacommunity, each taxon was detected only in less than two different patients (mean = 1.4 patients ± 0.8). What is evident is that bacteria can immigrate from the oral environment, and colonise and persist within the lower airways of CF patients. Although this is an important finding, from a clinical viewpoint, much still remains to be determined about their role in respiratory disease within CF patients.

A growing number of studies have observed that bacterial community composition is highly variable between CF patients (Rogers *et al.*, 2005; Harris *et al.*, 2007; Kelpac-Ceraj *et al.*, 2010). Here, similarities and differences in the composition of CF bacterial communities were assayed using the Bray–Curtis quantitative index of similarity (S_{BC}); dendrograms were generated for an average linkage cluster analysis of profiles from all taxa, and the core and satellite groups (Figure 4). The resulting cluster analyses for all taxa also revealed that community composition was highly variable between patients. The mean similarity of the bacterial communities
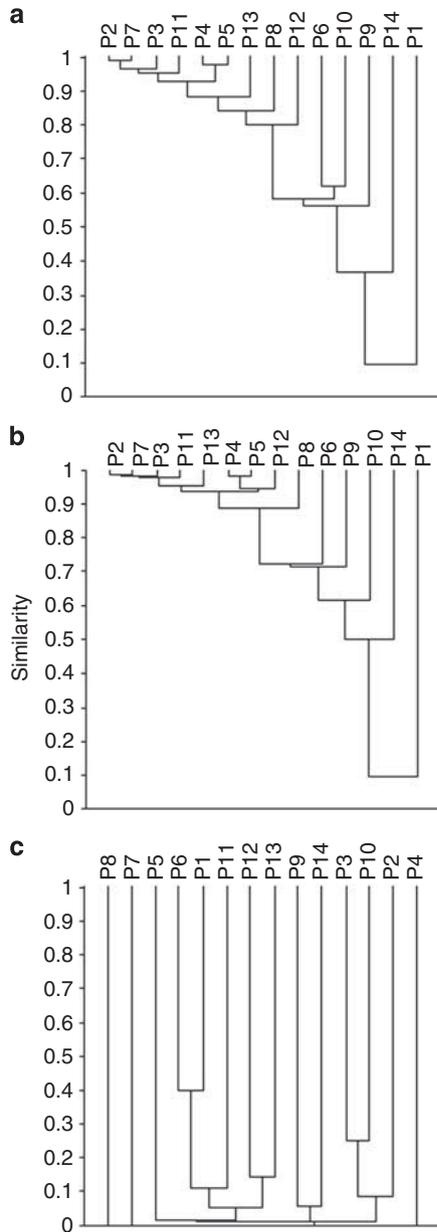
Figure 4 Dendrograms of bacterial community composition in the 14 patients for (**a**) all taxa, (**b**) the core and (**c**) satellite taxa groups. Patient taxa profiles were compared using the Bray–Curtis quantitative index of similarity and average linkage clustering.

taken pair-wise was 0.56, with a s.d. of ± 0.24 ($n = 91$ pair-wise comparisons) (Figure 4). When we examined compositional similarities and differences for members of core and satellite groups between patients, we found that similarity was more conserved for the core group (mean $S_{BC} = 0.61 \pm 0.25$) than in the satellite group, which comprised rarer spatially random distributed taxa ($S_{BC} = 0.02 \pm 0.06$). The presence and abundance of *P. aeruginosa* also affected the similarity between patients. When P1, in whom this taxon was not present, was removed from the analyses, the mean similarity of core group members between patients increased to $S_{BC} = 0.70 \pm 0.14$ ($n = 78$ pair-wise comparisons).

In the salivary microbiota, it has been established that there is a high bacterial diversity, within and between individuals, whereby each person typically has a unique set of normal oral bacterial species (Nasidze *et al.*, 2009). In the current study, we found that members of the oral cavity environment dominated the diversity of the CF metacommunity (Figure 3). Therefore, we re-assessed similarity in community composition between patients with and without the oral cavity taxa. When the oral microbiota members were removed from the analyses, the mean similarity of bacterial communities between patients increased to $S_{BC} = 0.80 \pm 0.03$. Conversely, the composition of the oral microbiota between patients was highly variable, with a mean similarity of $S_{BC} = 0.16 \pm 0.17$. This would suggest that colonisation of the CF lungs by bacterial taxa from the oral cavity strongly influences the high variability observed between CF lung bacterial communities.

To examine which clinical factors influenced the composition of all taxa, and the core and satellite group members between patients, we employed Mantel and partial Mantel tests. Both CFTR genotype and recent antibiotic treatment were found to correlate with the composition of all taxa and the core group members across patients, but not with the satellite group (Table 3). No other clinical factors such as similarities in lung function (forced expiratory volume in 1 s), patient age, gender or steroid treatment correlated with similarities in composition (Mantel tests $P > 0.05$ in all cases). Finding no significant correlations between clinical factors and

Table 3 Mantel and partial Mantel test summary statistics

	Mantel				Partial Mantel			
	r(SA)	P	r(SG)	P	r(SA.G)	P	r(SG.A)	P
Whole community	0.516	0.0001*	−0.294	0.003*	0.509	0.0001*	−0.278	0.008*
Core group	0.504	0.0001*	−0.350	0.001*	0.499	0.0001*	−0.343	0.002*
Satellite group	−0.255	0.994	0.077	0.763	−0.249	0.992	0.050	0.687

The partial Mantel statistic r(SA.G) estimates the correlation between S (community similarity) and A (similarity in antibiotic treatments), while controlling for the effect of G (differences in CF genotype). Conversely, r(SG.A) estimates the correlation between S and G, while controlling for A. Mantel test statistics to examine the correlation between S and A or G, respectively, are given. The P-values have been calculated using the distribution of the partial Mantel or Mantel test statistic estimated from 9999 permutations. Asterisks denote those relationships that were significant at the $P < 0.05$ level.

satellite group composition between patients was not surprising, as the satellite species are a result of random dispersal into the CF airways of a given patient. Using partial Mantel tests, we found that both CFTR genotype and antibiotic treatment were still significantly correlated with the composition of all taxa and the core group members between patients when controlling for the effects of the other respective clinical factor (Table 3). Importantly, antibiotic treatment and CFTR genotype were not auto-correlated when the Mantel test was performed ($r(AG) = -0.111$, $P = 0.304$). This could suggest that both similarities in CFTR genotype and antibiotic treatment between the adult patients sampled select for similar bacterial communities and specifically for members of the core taxa group.

We also investigated which clinical factors correlated with bacterial taxa richness across the patient cohort (Figure 5). We found that only lung function measured by forced expiratory volume in 1 s significantly correlated with richness, demonstrating significant positive linear relationships for total taxa richness, and the core and satellite group member richness (Figure 5), whereby taxa richness decreased with a reduction in lung function. More research is needed to be able to explain this observation, particularly given the fact that forced expiratory volume in 1 s is the single best predictor of mortality and is an important determinant of the timing of transplantation (Kerem *et al.*, 1992; Doershuk and Stern, 1999; Oikonomou *et al.*, 2002). In addition, however, our findings agree, in part, with those of the study by Kelpac-Ceraj *et al.* (2010) on paediatric CF patients, in that we also found that similarities in community composition correlated with similarities in CFTR genotype and antibiotic treatments. However, we did not find a relationship between patient age and richness. We postulate that this may be a result of bacterial community composition and richness being highly dynamic in younger patients, and that these community dynamics may stabilise with increase in patients' age.

In conclusion, this work suggests that for adult patients, a division can be made in terms of the species detected into those that are core or satellite to the CF lung. This has important implications for subsequent studies, and particularly those involving deep sequencing. Through this study, however, certain species were identified as core. Although any species present in the CF airways may be important, the identification of such core species should direct at least the initial focus of research. Of the species identified as core, only *P. aeruginosa* has been regarded traditionally as a key CF pathogen. It is worth noting that the other species that were regarded as traditional CF pathogens were found here to be transient. However, this study also identified 10 species within 4 genera associated previously with the oral cavity as important; all known pathogens in other infectious contexts. The importance of this was emphasised by the impact
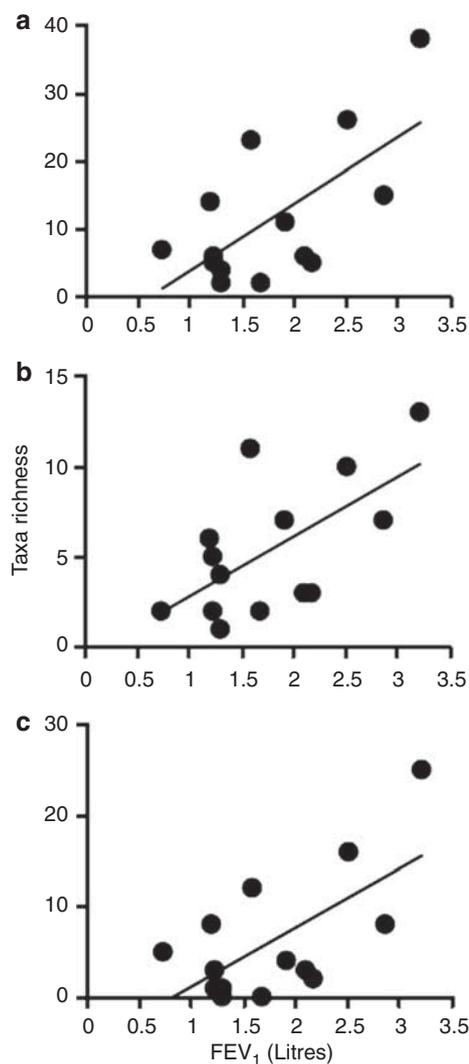

Figure 5 The relationships between bacterial taxa richness and lung function (FEV$_1$ in litres) for (**a**) the whole metacommunity, (**b**) the core group and (**c**) the satellite group. In each case linear regression lines have been fitted. For the whole metacommunity, $r^2 = 0.43$, $F_{1,12} = 9.2$, $P < 0.01$; core group, $r^2 = 0.40$, $F_{1,12} = 7.82$, $P < 0.02$; and satellite group, $r^2 = 0.42$, $F_{1,12} = 8.67$, $P < 0.01$.

of removing the satellite species on the overall levels of similarity observed. In addition, factors such as CFTR genotype and antibiotic use correlated across patients with the composition of the core group members but not with the satellite group. Combined, this reinforces the need to understand the behaviour of these species, both singly and in combination, in relation to CF lung disease. Clearly, more research is needed to understand the clinical relevance of these findings. Using the approach employed here, a large multi-centre study has been initiated, including patients representing a broader range of CF respiratory disease. As such, the present study should act as a focus for debate over the best ways to interpret clinically relevant information using ecological tools. Already, this work has generated fresh insight into the bacterial communities within the CF lung, and contributes to the efforts to further improve

therapy. This work also strongly supports the notion that the application of ecological approaches may similarly provide fresh insights into many other clinical scenarios.

Acknowledgements

Data deposition footnote
GenBank accession numbers: FM995625–FM997761.

References

Anon. (2008). *Investigation of Bronchoalveolar Lavage, Sputum and Associated Specimens*. BSOP 57. Health Protection Agency: London, UK.

Anon. (2009). *UK CF Registry Annual Data Report 2007*. Cystic Fibrosis Trust: Bromley, UK.

Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ. (2005). At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Appl Environ Microbiol* **71**: 7724–7736.

Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ. (2006). New screening software shows that most recent large 16S rRNA gene clone libraries contain chimeras. *Appl Environ Microbiol* **72**: 5734–5741.

Balfour-Lynn IA, Elborn JS. (2007). Respiratory disease: infection. In: Hodson M, Bush A, Geddes DM (eds). *Cystic Fibrosis*. Hodder Arnold: London, pp 137–158.

Clunes MT, Boucher RC. (2007). Cystic fibrosis: the mechanisms of pathogenesis of an inherited lung disorder. *Drug Discov Today Dis Mech* **4**: 63–72.

Coakley RD, Boucher RC. (2007). Pathophysiology: epithelial cell biology and ion channel function in the lung, sweat gland and pancreas. In: Hodson M, Bush A, Geddes DM (eds). *Cystic Fibrosis*. Hodder Arnold: London, pp 59–68.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K *et al.* (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**: 5069–5072.

Doershuk CF, Stern RC. (1999). Timing of referral for lung transplantation for cystic fibrosis: overemphasis on FEV_1 may adversely affect overall survival. *Chest* **115**: 782–787.

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 Biogeogr* **36**: 887–895.

Fisher RA, Corbet AS, Williams CB. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *J Anim Ecol* **12**: 42–58.

Gaston KJ, Blackburn TM. (2000). *Patterns and Process in Macroecology*. Blackwell Publishing: Oxford, UK.

Green JL, Holmes AJ, Westoby M, Oliver I, Briscoe D, Dangerfield M *et al.* (2004). Spatial scaling of microbial eukaryote diversity. *Nature* **432**: 747–750.

Grey JS, Bjorgesaeter A, Ugland KI. (2005). The impact of rare species on natural assemblages. *J Anim Ecol* **74**: 1131–1139.

Guo Q, Brown JH, Valone TJ. (2000). Abundance and distribution of desert annuals: are spatial and temporal patterns related? *J Ecol* **88**: 551–560.

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

Hanski I. (1999). *Metapopulation Ecology*. Oxford University Press: Oxford, UK.

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.

Harrison F. (2007). Microbial ecology of the cystic fibrosis lung. *Microbiology* **153**: 917–923.

Hubbell SP. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press: Princeton, NJ.

Kelpac-Ceraj V, Lemon KP, Martin TR, Allgaier M, Kembel SW, Knapp AA *et al.* (2010). Relationship between cystic fibrosis respiratory tract bacterial communities and age, genotype, antibiotics and *Pseudomonas aeruginosa*. *Environ Microbiol* **12**: 1293–1303.

Kerem E, Reisman J, Corey M, Canny GJ, Levison H. (1992). Prediction of mortality in patients with cystic fibrosis. *N Engl J Med* **326**: 1187–1191.

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

Lilley AK, Fry JC, Bailey MJ, Day MJ. (1996). Comparison of aerobic heterotropic taxa isolated from four root domains of mature sugar beet (*Beta vulgaris*). *FEMS Microbiol Ecol* **21**: 231–242.

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar *et al.* (2004). ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363–1371.

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

Magurran AE. (2004). *Measuring Biological Diversity*. Blackwell Publishing: Oxford, UK.

Magurran AE. (2007). Species abundance distributions over time. *Ecol Lett* **10**: 347–354.

Mathers CD, Lopez AD, Murray CJL. (2006). The burden of disease and mortality by condition: data, methods and results for 2001. In: Lopez AD, Mathers CD, Ezzati M, Murray CJL, Jamison DT (eds). *Global Burden of Disease and Risk Factors*. Oxford University Press: New York, NY, pp 45–240.

Nasidze I, Li J, Quinque D, Tang K, Stoneking M. (2009). Global diversity in the human salivary microbiome. *Genome Res* **19**: 636–643.

Oikonomou A, Manavis J, Karagianni P, Tsanakas J, Wells AU, Hansell DM *et al.* (2002). Loss of FEV_1 in cystic fibrosis: correlation with HRCT features. *Eur Radiol* **12**: 2229–2235.

Preston FW. (1948). The commonness, and rarity, of species. *Ecology* **29**: 254–283.

Rogers GB, Carroll MP, Serisier DJ, Hockey PM, Jones G, Bruce KD. (2004). Characterisation of bacterial community diversity in cystic fibrosis lung infections by use of 16S ribosomal DNA terminal restriction

12

fragment length polymorphism profiling. *J Clin Microbiol* **42**: 5176–5183.

Rogers GB, Carroll MP, Serisier DJ, Hockey PM, Jones G, Kehagia V et al. (2006). Use of 16S rRNA gene profiling by terminal restriction fragment length polymorphism analysis to compare bacterial communities in sputum and mouthwash samples from patients with cystic fibrosis. *J Clin Microbiol* **44**: 2601–2604.

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, Daniels TWV, Tuck A, Carroll MP, Connett GJ, David GJ et al. (2009). Studying bacteria in respiratory specimens by using conventional and molecular microbiological approaches. *BMC Pulm Med* **9**: 14–24.

Rogers GB, Hart CA, Mason JR, Hughes M, Walshaw MJ, Bruce KD. (2003). Bacterial diversity in cases of lung infection in cystic fibrosis patients: 16S ribosomal DNA (rDNA) length heterogeneity PCR and 16S rDNA terminal restriction fragment length polymorphism profiling. *J Clin Microbiol* **41**: 3548–3558.

Rogers GB, Stressmann A, Walker AW, Caroll MP, Bruce KD. (2010a). Lung infections in cystic fibrosis: deriving clinical insight from microbial complexity. *Expert Rev Mol Diagn* **10**: 187–196.

Rogers GB, Skelton S, Serisier DJ, van der Gast CJ, Bruce KD. (2010b). Determining cystic fibrosis-affected lung microbiology: comparison of spontaneous and serially induced sputum samples by use of terminal restriction fragment length polymorphism profiling. *J Clin Microbiol* **48**: 78–86.

Scannapieco FA. (1999). Role of oral bacteria in respiratory infection. *J Periodontol* **70**: 793–802.

Schloss PD, Handelsman J. (2005). Introducing DOTUR, a computer programme for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* **71**: 1501–1506.

Sibley CD, Parkins MD, Rabin HR, Duan K, Norgaard JC, Surette MG. (2008). A polymicrobial perspective of pulmonary infections exposes an enigmatic pathogen in cystic fibrosis patients. *Proc Natl Acad Sci USA* **105**: 15070–15075.

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

Stecher B, Robbiani R, Walker AW, Westendorf AM, Barthel M, Kremer M et al. (2007). *Salmonella enterica* serovar *typhimurium* exploits inflammation to compete with the intestinal microbiota. *PloS Biol* **5**: 2177–2189.

Tunney MM, Field TR, Moriarty TF, Patrick S, Doering G, Muhlebach MS et al. (2008). Detection of anaerobic bacteria in high numbers in sputum from patients with cystic fibrosis. *Am J Respir Crit Care Med* **177**: 995–1001.

Ulrich W, Ollik M. (2004). Frequent and occasional species and the shape of relative-abundance distributions. *Div Distr* **10**: 263–269.

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.

Wang Q, Garrity GM, Tiedje JM, Cole JR. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**: 5261–5267.

Yoon SS, Hennigan RF, Hilliard GM, Ochsner UA, Parvatiyar K, Kamani MC et al. (2002). *Pseudomonas aeruginosa* anaerobic respiration in biofilms: Relationships to cystic fibrosis pathogenesis. *Dev Cell* **3**: 593–603.