The ISME Journal (2011) 5, 20–29

www.nature.com/ismej

ORIGINAL ARTICLE

Phylogenetic and metabolic diversity of bacteria associated with cystic fibrosis

Adam M Guss[1,3,4], Guus Roeselers[1,3], Irene LG Newton[1,5], C Robert Young[1], Vanja Klepac-Ceraj[2,6], Stephen Lory[2] and Colleen M Cavanaugh[1]
[1]*Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA and*
[2]*Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, MA, USA*

In patients afflicted with cystic fibrosis (CF), morbidity and mortality are primarily associated with the adverse consequences of chronic microbial bronchial infections, which are thought to be caused by a few opportunistic pathogens. However, recent evidence suggests the presence of other microorganisms, which may significantly affect the course and outcome of the infection. Using a combination of 16S rRNA gene clone libraries, bacterial culturing and pyrosequencing of barcoded 16S rRNA amplicons, the microbial communities present in CF patient sputum samples were examined. In addition to previously recognized CF pathogens such as *Pseudomonas aeruginosa* and *Staphylococcus aureus*, >60 phylogenetically diverse bacterial genera that are not typically associated with CF pathogenesis were also detected. A surprisingly large number of fermenting facultative and obligate anaerobes from multiple bacterial phyla was present in each sample. Many of the bacteria and sequences found were normal residents of the oropharyngeal microflora and with many containing opportunistic pathogens. Our data suggest that these undersampled organisms within the CF lung are part of a much more complex microbial ecosystem than is normally presumed. Characterization of these communities is the first step in elucidating potential roles of diverse bacteria in disease progression and to ultimately facilitate advances in CF therapy.
The ISME Journal (2011) **5,** 20–29; doi:10.1038/ismej.2010.88; published online 15 July 2010
Subject Category: microbial population and community ecology
Keywords: CF; lung; microbial diversity; polymicrobial disease; *Pseudomonas aeruginosa*; pyrosequencing

Introduction

Polymicrobial infections are increasingly being recognized as clinically important in diseases such as inflammatory bowel disease, otitis media, vaginosis, periodontitis and cystic fibrosis (CF) (Brogden *et al.*, 2005). Understanding the diversity of microorganisms and their interactions during infection is essential to effectively treating these conditions, as infectious microorganisms may participate in a complex ecological community, generating niches used by others, nutritional dependency, or other synergistic or antagonistic relationships (Pedron and Sansonetti, 2008). CF is a human autosomal recessive disease that in one of its many manifestations results in decreased clearance of mucus from the lungs with concomitant bacterial lung infections (Smith *et al.*, 1996; Gomez and Prince, 2007). An ensuing inflammatory response causes progressive lung damage and is the primary cause of CF morbidity and mortality (Lyczak *et al.*, 2002). Common CF bronchial pathogens, as identified by the traditional method of culturing bacteria from expectorated mucus, include *Pseudomonas aeruginosa*, *Staphylococcus aureus*, *Burkholderia* species and the *Streptococcus milleri* group (Lyczak *et al.*, 2002; Gibson *et al.*, 2003; Sibley *et al.*, 2008). Pathogens isolated from the CF lung are traditionally thought to occur as monocultures or as consortia of a small number of species (Wahab *et al.*, 2004; Harrison, 2007).

However, the CF lung is a stable, warm, humid, organic compound-rich environment with limited microbial clearance, providing favorable conditions for the proliferation of many organisms. Hence, it would be surprising if the presence of only one or a few microorganisms were the rule, rather than the exception. Indeed, recent research showed that the microbial communities present in the CF lung may be more complex. For instance, by using culture media that are not selective for the traditional CF pathogens, several atypical microorganisms have been isolated, including various *Bordetella*,

Correspondence: CM Cavanaugh, Department of Organismic and Evolutionary Biology, Harvard University, Biological Laboratory 4081, 16 Divinity Avenue, Cambridge, MA, USA.
E-mail: cavanaug@fas.harvard.edu
[3]These authors contributed equally to this work.
[4]Current address: Thayer School of Engineering, Dartmouth College, Hanover, NH, USA.
[5]Current address: Department of Biological Sciences, Wellesley College, Wellesley, MA, USA.
[6]Current address: Forsyth Institute, Boston, MA, USA.
Received 1 March 2010; revised 20 May 2010; accepted 21 May 2010; published online 15 July 2010

Ralstonia, *Acinetobacter* and *Moraxella* species (Coenye *et al.*, 2002). Initial culture-independent studies using terminal restriction fragment length polymorphism (T-RFLP) suggested the presence of a greater bacterial diversity than typically cultured from CF patients (Rogers *et al.*, 2006; Sibley *et al.*, 2008). Although T-RFLP profiles can give a first approximation of the level of diversity, they do not allow unambiguous classification, leaving the microorganisms present unidentified. Recently, 16S rRNA clone library studies of CF lungs described variable community compositions—some dominated by a single known CF pathogen, whereas others contained multiple organisms or organisms not routinely identified as CF pathogens (Harris *et al.*, 2007; Bittar *et al.*, 2008). Two recent studies used barcoded pyrosequencing methods, each on a single CF sputum sample as a PCR template, but these results were based on Roche GS 20 chemistry (Roche Applied Sciences, Branford, CT, USA) (produces shorter reads; ~100 bp) and not analyzed beyond the clustering of sequences by community profile (Hamady *et al.*, 2008; Armougom *et al.*, 2009).

Microbial diversity assessment is a relatively new application for pyrosequencing and has yielded amazing insights into the community composition in a variety of environments (Sogin *et al.*, 2006; Al Masalma *et al.*, 2009; Price *et al.*, 2009). In this study, we examine the microbial diversity found within CF patient sputum samples using a polyphasic approach. Phylotypes present in sputum samples were analyzed using 16S rRNA clone libraries and pyrosequencing to generate large numbers of 16S rRNA gene sequences. In addition, sequence analysis directed subsequent culturing efforts.

The results presented here indicate a tremendously expanded spectrum of bacterial phylotypes, including opportunistic pathogens, associated with the CF lung and also point to possible metabolic interactions involving anaerobic bacteria.

Materials and methods

Sputum acquisition and handling

Four sputum samples were collected from four different randomly selected anonymous patients with CF at the Children's Hospital in Boston, MA USA under an IRB-approved protocol and were immediately put on ice until examination, which occurred within 1 h of collection. Each sample was washed three times in 20 ml sterile phosphate-buffered saline pH 7.2, which has been reported to adequately remove contaminating saliva (Rogers *et al.*, 2006), liquefied by incubation with an equal volume of sterile 1 mM dithiothreitol in phosphate-buffered saline on ice for 1 h with occasional vortexing, and used for DNA isolation and culture study as described below. The remainder of each liquefied sample was preserved for targeted culture study by mixing with 1/2 volume sterile 50% glycerol and storage at −80 °C.

16S rRNA gene amplification, library construction and sequencing

DNA was isolated from the liquefied sputum using the Qiagen DNeasy genomic DNA extraction kit protocol D (Qiagen, Valencia, CA, USA), which included a lysozyme step for improved DNA extraction from diverse organisms. The 16S rRNA gene was amplified using bacteria-specific universal primers 27F and 1492R (Lane, 1991) with a proof-reading polymerase (Finnzyme Phusion) according to manufacturers' instructions with the following modifications: for each sample, 400 µl PCR reaction mix using 1500 ng template DNA was divided among 8 tubes and run for 18 cycles to minimize PCR bias (Polz and Cavanaugh, 1998). PCR products were concentrated to 25 µl, gel purified and used as template for three more PCR cycles using fresh reagents to minimize the formation of PCR-induced chimera sequences (Thompson *et al.*, 2002). PCR products were cloned into *Escherichia coli* using the pCR4-TOPO plasmid vector (Invitrogen, Carlsbad, CA, USA). Randomly chosen colonies from each library were grown in liquid culture at 37 °C overnight. The 16S rRNA genes were PCR-amplified from each clone using 1 µl culture as template, and PCR primers M13-F and M13-R. PCR products were sequenced unidirectionally using primer 27F.

Taxonomic classification and phylogenetic analysis of 16S rRNA gene sequences

The 16S rRNA gene sequences generated from clone libraries and cultured isolates (see below) were manually edited and trimmed to the same length (equivalent to *E. coli* base positions 105–815) in Sequencher 4.7 (Gene Codes Corp, Ann Arbor, MI, USA). Sequences were clustered into operational taxonomic units (OTUs) by ⩾99% sequence identity, and sequences with the minimum sum of the square of distances between sequences within each 99% cluster were used in the taxonomic and phylogenetic analyses that followed. BLAST (NCBI's blastall version 2.2.15) was used to search each sequence against the GenBank database (the search excluded uncultured and environmental sample sequences). Subsequently, sequences were taxonomically classified using the Ribosomal Database Project (RDP) Naïve Bayesian Classifier tool (Wang *et al.*, 2007) using an 80% confidence threshold.

Sequences were aligned using the SINA aligner (http://www.arb-silva.de), and manually evaluated in MacClade 4.06 (Maddison and Maddison, 2000). Sequences were then imported into ARB (Kumar *et al.*, 2006) in which the majority of sequences were assigned to phyla based on their position after

parsimony insertion into the ARB database dendogram, omitting hypervariable portions of the rRNA gene using a filter based on the Lane mask (Lane, 1991).

Maximum likelihood trees were constructed with novel and reference sequences selected from the ARB-SILVA database (Version 100; 2009), using RAxML-VI-HPC v2.2 under the GTRCAT model of evolution (Stamatakis *et al.*, 2008). *Thermatoga maritima* (M21774), used as an outgroup, was subsequently pruned from the ML trees. Bootstrap resampling (1000 replicates) was used to test the robustness of inferred topologies; values $\geqslant 50\%$ are shown at the tree nodes.

Pyrosequencing of barcoded 16S rRNA gene amplicons

For greater sample depth, sputum samples 2, 3 and 4 (for which adequate amounts of DNA were available) were analyzed by massively parallel pyrosequencing of barcoded amplicons. A fragment of the 16S rRNA gene (~ 330 bp), spanning the V1 and V2 hypervariable regions, was PCR amplified from three sputum samples. Empirical and *in silico* analyses indicate that the V1–V2 region outperforms other variable regions in reproducing full-length 16S rRNA gene-based taxonomic classifications (Wang *et al.*, 2007; Hamp *et al.*, 2009). Universal Bacteria primers 27F and 338RII (Lane, 1991; Daims *et al.*, 1999) were modified by adding ligation adaptors and/or MID barcodes (that is, sample identification sequences) to the 5′-ends (Supplementary Table S1).

PCR was performed using a high-fidelity polymerase (Phusion Hot Start, Finnzymes, Espoo, Finland), at 50 °C annealing temperature, using 1500 ng template in 400 µl volume (split between 8 tubes) and 20 cycles. Amplicons, purified and concentrated to 50 µl using the Promega PCR purification kit (Promega, Madison, WI, USA), quantified spectrophotometrically (NanoDrop 1000, Thermo Scientific) and standardized to 100 ng µl^{-1}, were used as templates for emulsion PCR using the emPCR kit II (Roche Applied Sciences). DNA was sequenced using a Genome Sequencer FLX and the GS-LR70 kit (Roche Applied Sciences) by the Environmental Genomics Core Facility (University of South Carolina) on LR70 plates following Roche standard protocols.

FASTA-formatted sequences and corresponding quality scores (QC) were extracted from the SFF data file using the GS Amplicon software package (Roche Applied Sciences). Sequences were binned by sample of origin using the unique barcodes, which were removed before downstream analyses. Length and average quality score were evaluated for each read; sequences were culled if the length was <210 bp and >280 bp, the SFF quality score was <30, they contained any ambiguous base calls (Ns), or did not match the primer or one of the used tag sequences.

Sequences were aligned using the Infernal secondary structure based aligner (Nawrocki and Eddy, 2007) implemented in the RDP pyrosequencing pipeline (http://wildpigeon.cme.msu.edu/pyro/index.jsp). Aligned sequences were clustered into OTUs defined by 97% similarity using the complete-linkage clustering tool implemented in the RDP pyrosequencing pipeline. Rarefaction curves were calculated using EstimateS (Version 7.5, RK Colwell, http://purl.oclc.org/estimates).

Shannon–Weaver and Chao1 biodiversity indices were calculated for each sputum sample using EstimateS. The Shannon–Weaver index (Shannon and Weaver, 1963) is a nonparametric diversity index that combines estimates of richness (the total number of OTUs) and evenness (the relative abundance of OTUs). For example, communities with one dominant species have a low index, whereas communities with a more even distribution have a higher index. Chao1 (Chao, 1987) is a nonparametric estimator of the minimum richness (number of OTUs) and is based on the number of rare OTUs (singletons and doublets) within a sample.

Sequences were taxonomically classified by the RDP-II Naïve Bayesian Classifier (Wang *et al.*, 2007; Liu *et al.*, 2008) using an 80% confidence threshold. Sequences that could not be classified to at least kingdom level were excluded from subsequent diversity analyses.

Bacterial isolation and culturing

Sequence data analyses, suggesting the presence of microorganisms not typically detected in a clinical setting, directed subsequent culturing efforts from CF sputum. Samples were plated on different types of solid culture media under aerobic and anaerobic conditions. Liquefied sputum was diluted in 10-fold increments in phosphate-buffered saline, and 100 µl of each dilution was plated on standard Tryptic Soy Broth-blood agar plates aerobically for 'nonselective' isolation of bacteria.

Dilutions of the glycerol-preserved sputum samples were plated on three additional solid media. To imitate the ionic composition of CF sputum, medium AG1 consisted of the salt portion of Synthetic CF Medium (SCFM) (Palmer *et al.*, 2007) supplemented with 0.5% submaxillary porcine mucin (Sigma-Aldrich, St Louis, MO, USA) as the sole carbon and energy source. Medium AG2 included the SCFM salts above, supplemented with 5 g tryptone, 2 g yeast extract, 2 g glucose, 1 g *N*-acetylglucosamine and 5% sheep blood per liter. Medium AG3, identical to medium AG2, was further supplemented with 20 mM sodium lactate. Each medium was incubated under both aerobic and anaerobic conditions. Anaerobic conditions were created by addition of 0.5 g cysteine–HCl and 1 ml 0.1% resazurin per liter and incubation using the GasPak Anaerobic System (Becton Dickinson, Franklin Lanes, NJ, USA). All plates were incubated

at 37 °C and each isolate was single-colony purified at least once.

Purified colonies from each isolate were suspended in 100 µl H_2O. The 16S rRNA genes were PCR-amplified using 1 µl cell suspension as template with primers 27F and 1492R and sequenced unidirectionally using primer 27F. Sequence data obtained from isolates and clone libraries have been submitted to NCBI GenBank under accession numbers: GQ900752-GQ900888.

Results

Diverse microbial communities were detected in all analyses of CF sputum samples with clone library, barcoded pyrosequencing, and culture analyses revealing a deep reservoir of bacterial diversity. These results indicate that the conventional model based on culturing, which assumes diseased lungs are characterized by bacterial monocultures or limited cocultures, do not accurately represent the genetic and metabolic diversity of the CF microbial assemblage.

16S rRNA libraries

To examine the diversity of bacteria within the CF lung without the bias associated with culture, 16S rRNA gene libraries were constructed from four sputum samples. Between 90 and 220 clones (~800 bp) were sequenced from each (Table 1). A total of 38 OTUs (≥99% sequence identity) were present in the combined libraries. These 38 OTUs represent 22 genera from 6 bacterial phyla (Figure 1 and Supplementary Figure S1), 11 of which were unique to a given sample (Supplementary Figure S2B).

The majority of the clone library diversity occurred within the Bacteroidetes, Fusobacteria and Firmicutes phyla (Figure 1 and Supplementary Figure S1), thus resembling other human microbiomes (Pei *et al.*, 2004; Eckburg *et al.*, 2005; Bik *et al.*, 2006). *P. aeruginosa* was the only Proteobacteria species consistently detected across all libraries (based on >99% similarity by BLASTN analysis and ML phylogeny) (Supplementary Figure S1A). While sequences from typical cultured CF pathogens, for example, *P. aeruginosa* and *S. aureus*, were identified, many phylotypes corresponding to organisms not routinely cultured from CF samples were also detected (Figure 1 and Supplementary Figure S1). For example, anaerobic organism phylotypes, including those typically found in the oral cavity (Aas *et al.*, 2005; Keijser *et al.*, 2008), were detected at relatively high frequency (Figure 1 and Supplementary Figure S1). These include the genera *Streptococcus* (Firmicutes) and *Rothia* (Actinobacteria), which are fermentative, facultative anaerobes that produce lactate as a primary metabolic end product, as well as diverse obligate anaerobes in the genera *Bacteroides*, *Prevotella*, *Porphyromonas* (Bacteroidetes) and *Veillonella* (Firmicutes).

In addition to these anaerobes, never-before cultured phylotypes were also detected. Clones were found in two sputum libraries with high sequence identity (99.3–100%) to members of the candidate phylum TM7 (Figure 1 and Supplementary Figure S1D), an uncultured bacterial clade detected in a variety of environments, ranging from hydrothermal vents to the human mouth (Hugenholtz *et al.*, 2001; Paster *et al.*, 2001; Tringe *et al.*, 2008). Other clones for which no corresponding cultured species are known (defined as <97% rRNA gene sequence identity) included sequences most similar to the genera *Bacteroides*, *Bergeyella*, *Butyrivibrio*, *Lachnospira*, *Prevotella*, *Porphyromonas* and *Terrimonas* (Supplementary Figure S1).

Table 1 Sampling depth and biodiversity found by 16S rRNA gene clone library analysis of sputum samples from CF patients 1, 2, 3 and 4

	CF sputum sample			
	1	*2*	*3*	*4*
Sampling depth				
Total 16S rRNA gene clones	91	215	220	93
OTUs (99% ID)	17	20	19	8
Phyla	4	6	5	3
Genera	8	13	7	5

Abbreviations: CF, cystic fibrosis; ID, DNA identity; OTUs, operational taxonomic units; RDP, Ribosomal Database Project.
The numbers of phyla and genera are based on taxonomic classification by the RDP Naïve Bayesian Classifier.

Pyrosequencing

The application of massively parallel pyrosequencing to sputum samples from the CF lung allowed for a much greater depth of sampling of the microbial diversity than either clone libraries or culture; this method allows for the detection and identification of 16S rRNA sequences that comprise a smaller fraction of the PCR amplicon pool, including potential pathogens. Between ~34 000–68 000 high-quality pyrosequences were analyzed per sputum sample with more than 1000 distinct sequences (100% identity over ~230 nucleotides) detected in each of the three samples (Table 2). When sequences were grouped based on 97% identity, to account for possible sequence errors (Kunin *et al.*, 2009), ~200–300 sequence clusters were identified per sample.

Diversity estimators were computed for each sample using the OTUs (97%) (Table 2). Shannon–Weaver diversity indices, which denote the relative OTU abundance, for samples 2 (1.15) and 3 (1.39) were similar but the Shannon–Weaver index of sample 4 (2.64) was higher (Table 2), indicating that this community is more diverse despite the lower number of observed OTUs (lower richness).

The Chao1 minimum richness estimates indicate that sputum samples 2 and 3 did not reach the plateau with the current sequencing effort (Table 2). The Chao1 estimate for sample 4 (~243) is in close agreement with the observed number of OTUs (231). Indeed, rarefaction curves (Figure 2a) of samples 2 and 3 did not become asymptotic by ~35 000 reads, whereas sample 4 approached saturation, indicating that the OTU diversity was almost completely covered.

Sequence clusters (97% similarity) were assigned to eight bacterial phyla (Figures 1 and 2d), two of which (Tenericutes and Spirochetes) were not found by 16S clone library analysis or culturing (Figure 1).

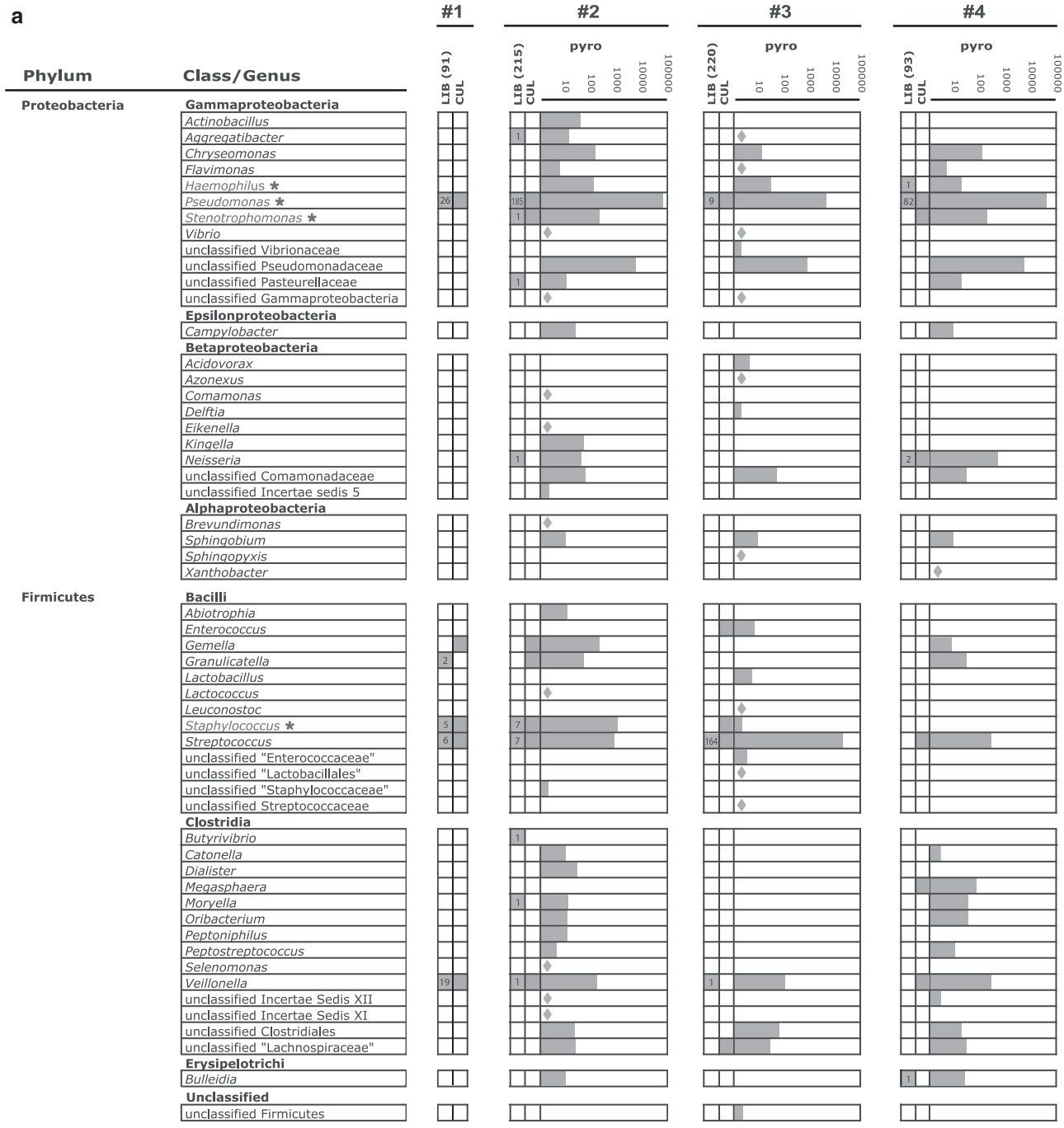

Figure 1 (**a**, **b**) Bacterial genera detected in CF sputum samples by 16S rRNA gene clone libraries, bacterial culturing and pyrosequencing. The abundance of pyrosequencing reads attributed to different genera within samples 2, 3 and 4 is shown in bar graphs (log scale). Sample 1 did not generate adequate amounts of DNA for pyrosequencing analysis. In panel **a**, diamond (◆) indicates a singleton. Genus designation was determined using the RDP Naïve Bayesian Classifier tool. Traditional CF pathogens are in red and indicated by an asterisk (*). The color reproduction of the figure is available on the html full text version of the paper.

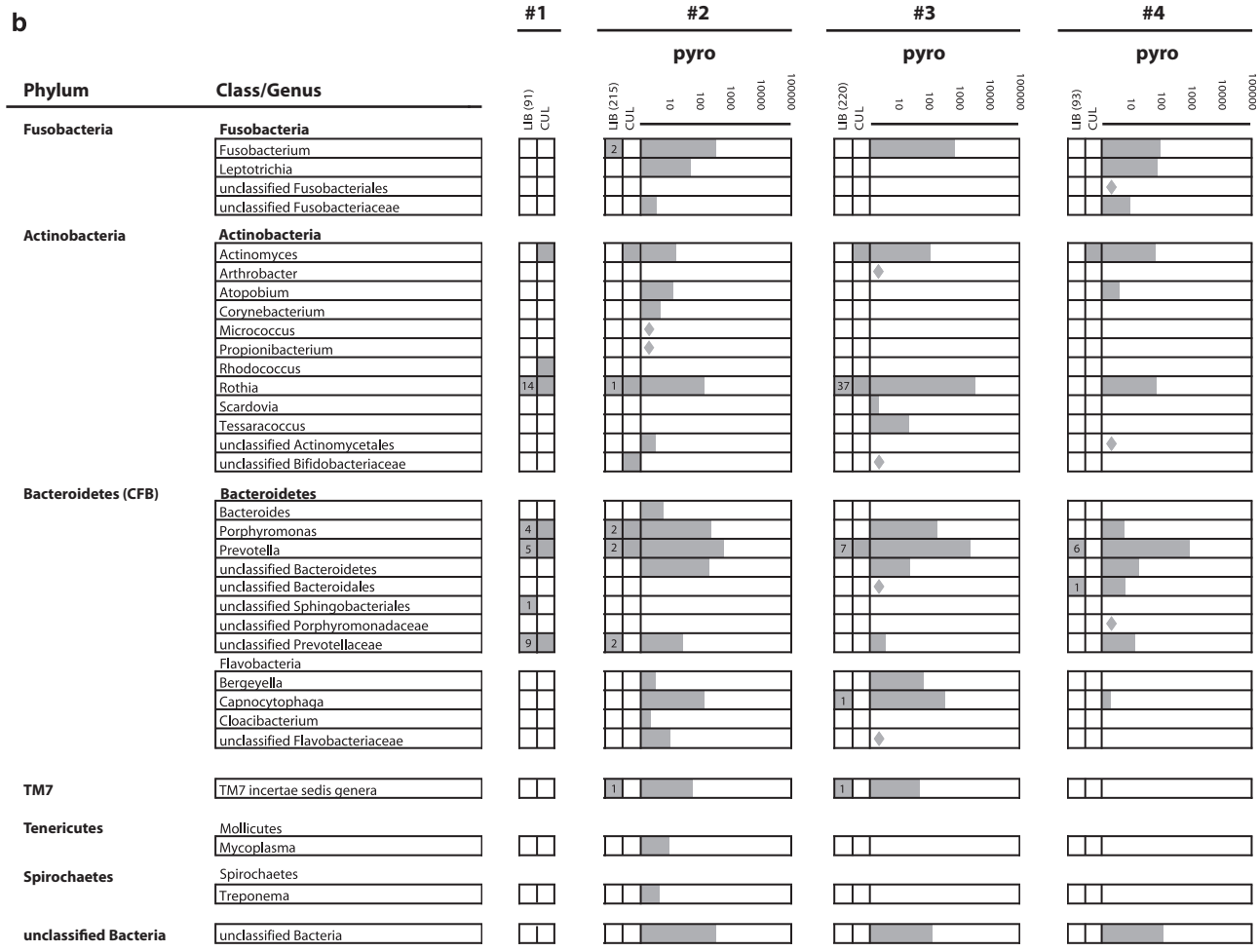

Figure 1 Continued

Table 2 Sampling depth and biodiversity found by barcoded pyrosequencing of sputum samples from CF patients 2, 3 and 4

	CF sputum sample		
	2	*3*	*4*
Sampling depth			
Total pyrosequences	53851	40791	72707
High-quality pyrosequences	47484	34640	68083
Unique sequences (100% ID)	1272	1174	1155
OTUs (97% ID)	346	261	238
Phyla	9	7	5
Genera	47	28	27
Diversity indices			
Shannon–Weaver (97%)	1.39	1.15	2.64
Chao1 (97%)	461	341	242

Abbreviations: CF, cystic fibrosis; OTUs, operational taxonomic units; RDP, Ribosomal Database Project.
Numbers of phyla and genera are based on taxonomic classification by the RDP Naïve Bayesian Classifier. The Shannon–Weaver index combines estimates of richness (total number of OTUs) and evenness (relative abundance). Chao1 is an estimator of the minimum richness and is based on the number of rare OTUs (singletons and doublets) within a sample.

The eight phyla comprised >60 bacterial genera (Table 2 and Figure 1). Nineteen genera were shared between all three sputum samples analyzed by pyrosequencing (Figures 1 and 2c), eight of which were also found using 16S rRNA clone libraries and culture techniques. Thus, these genera may represent a core of organisms common within the CF lung. A surprising number of reads were attributed to genera known to include opportunistic pathogens (Supplementary Table S2).

As in the 16S clone library analysis, the most common bacterial genus detected was *Pseudomonas*; 101 480 pyrosequences were assigned to this genus (Figure 1) with all but three sequences being most similar to *P. aeruginosa* (determined by BLASTN; ⩾99% sequence identity). *P. aeruginosa*-like sequences comprised, on average across all three samples, 67% of the entire PCR amplicon pool (range = 16–95%). The next most common bacterial genera were *Streptococcus* (21% average frequency), *Rothia* (3%) and *Prevotella* (3%). The rest of the genera found within one or more of the samples comprised <1% of the total amplicon population.

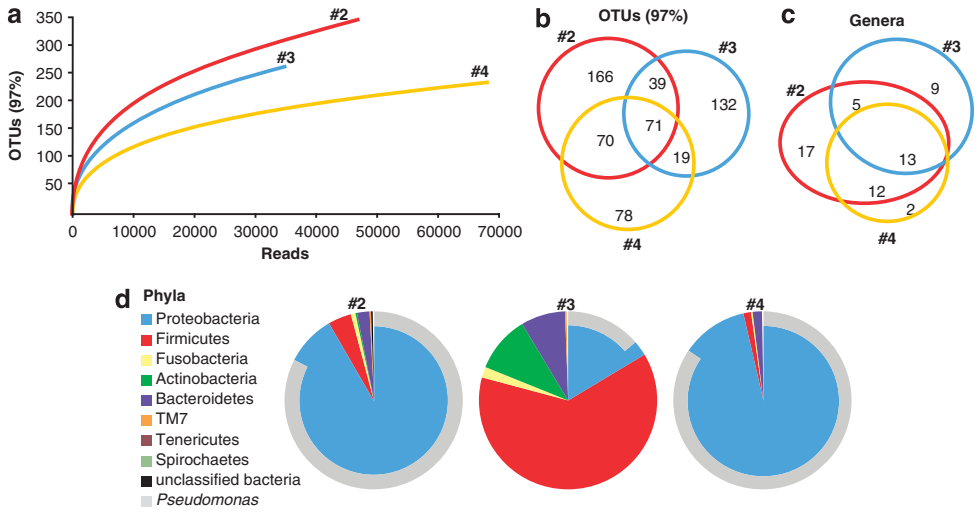

Figure 2 Sequence diversity and abundance detected by pyrosequencing in CF sputum samples 2, 3 and 4. (**a**) Rarefaction curves for total bacterial communities in the three samples at a 99% sequence similarity level. Curvature toward the horizontal indicates the increased sequencing effort required to observe novel OTUs when only rare OTUs remain to be discovered. (**b**) Venn diagram showing unique and shared OTUs (97%) in each sputum sample. (**c**) Venn diagram showing numbers of unique and shared genera detected in CF sputum samples by pyrosequencing. Genus designation was determined using the RDP Naïve Bayesian Classifier tool. Sequences that could not be classified to genera are not included in this figure. (**d**) Pie charts showing the taxonomic breakdown (phylum level) of pyrosequences (grey bars indicate sequences affiliated to the genus *Pseudomonas*).

Bacterial isolation and culturing

A culture-based approach with subsequent 16S rRNA gene sequencing was used using both standard and nonstandard media under aerobic and anaerobic conditions (Figure 1 and Supplementary Figure S1). As expected, *P. aeruginosa*, *S. aureus* and *Stenotrophomonas maltophilia*, detected in the 16S clone libraries and classically cultured from CF samples, were isolated from the samples on standard, 'nonselective' media (aerobic Tryptic Soy Broth-blood agar). Atypical CF species and genera, detected in the 16S clone and pyrosequencing libraries, were also isolated using these media, including *Streptococcus mitis* strains, *Rothia dentocariosa* and *Rothia mucilaginosa*. Anaerobic bacteria that constituted major components of the clone and pyrosequencing libraries were specifically targeted for isolation by inclusion of cysteine in the culture medium and incubation without O_2. Both obligate and facultative anaerobic bacteria were cultured with little difficulty under these conditions, including *Prevotella, Porphyromonas*, and *Gemella* species (Figure 1 and Supplementary Figure S1).

Discussion

Microbial phylogenetic diversity

The CF lung harbors a remarkably diverse population of microorganisms within eight bacterial phyla, comprising >60 genera, including facultative and obligate anaerobes, oral bacteria and opportunistic pathogens, many of which have never before been found in the CF lung (Figure 1). The generation of this novel result is due to the use of barcoded amplicon pyrosequencing in addition to 16S rRNA library sequencing, which subsequently directed culture attempts. These results indicate that traditional culture- and cloning-based methods are insufficient to describe the microbial populations present in the CF lung (Figure 1 and Supplementary Figure S2A), consistent with recent molecular studies (Rogers *et al.*, 2006; Harris *et al.*, 2007; Bittar *et al.*, 2008; Armougom *et al.*, 2009).

The microbial diversity detected is not surprising given that the CF lung environment is warm, humid and organic-rich, thus creating favorable conditions for proliferation of many organisms. CF lung conditions also limit the activities of several arms of the innate defense system, in particular mechanical removal of organisms by mucociliary clearance (Bals *et al.*, 1999). The majority of the genera detected both by culture and molecular methods have been also been reported in the human oral and nasal cavities, including *Capnocytophaga*, *Fusobacterium*, *Neisseria*, *Porphyromonas*, *Prevotella* and *Veillonella* strains, as well as sequences related to the Lachnospiraceae, Spirochetes and TM7 clones (Paster *et al.*, 1998, 2006; Pei *et al.*, 2004; Aas *et al.*, 2005). Although the possibility that these organisms were due to oral contamination cannot be excluded, this is unlikely as each sputum sample was washed thoroughly following conventional protocols to remove saliva. As expectoration of CF patient sputum certainly contaminates saliva on a routine basis, organisms common to both environments could be explained by (1) saliva contamination of sputum; (2) sputum contamination of saliva; or (3) similar organisms present in both at the time of sample collection. Microorganisms associated with the oropharyngeal cavity may be repeatedly introduced

into the lungs of CF patients by normal breathing or by aspiration of saliva during exacerbation. With the diminished clearance of mucus, these microorganisms may be able to colonize and persist. Alternatively, a recent metagenomic examination of indoor, recirculated air detected the presence of many putative human-associated microorganisms, revealing another possible inoculum for the CF lungs (Tringe *et al.*, 2008). As 'normal' human lungs are generally not considered to be colonized (Flanagan *et al.*, 2007), samples from healthy patients would not be informative controls.

Facultative and obligate anaerobes associated with CF

A few of the bacteria detected by culture and sequence analyses (for example, *Veillonella*, *Prevotella* and *Porphyromonas*) found in this study are obligate anaerobes. Although it may initially be surprising to think of the human lung as having anaerobic niches, mucus is often primarily anoxic (Worlitzsch *et al.*, 2002). Diffusion of O_2 through thick mucus is slow, and bacterial respiration is fast (Matsui *et al.*, 2006). Therefore, microaerophilic or anaerobic pockets in the CF lung may be the rule rather than the exception *in vivo*, suggesting that anaerobes and microaerophiles could be significant members of the CF lung microbial populations (Alvarez-Ortega and Harwood, 2007). Indeed, diverse anaerobes were recently cultured from a variety of CF samples, providing additional support that such bacteria may be a common component in the CF lung (Tunney *et al.*, 2008).

Further evidence in support of the importance of anaerobes comes from the recent detection of lactate in CF sputum (Palmer *et al.*, 2007). Many of the genera detected in our study are known to produce lactate as a fermentation end product (for example, *Staphylococcus*, *Streptococcus* and *Rothia* strains). Further, lung epithelial cells can secrete lactate under both physiological and pathophysiological conditions (De Backer *et al.*, 1997). Notably, *Veillonella* is characterized by the ability to consume lactate anaerobically for growth (Ng and Hamilton, 1971; Seeliger *et al.*, 2002), suggesting that these organisms could interact metabolically with lactate-producing bacteria in the CF lung microbial community. Given that strict anaerobes require a reducing agent in the growth medium, they are likely to be missed in most clinical culture-based tests using standard aerobic (selective) media.

Opportunistic pathogens

In addition to the two most abundant genera (*Pseudomonas* and *Streptococcus*) in the clone and pyrosequencing libraries (Figure 1), more than 60 genera were detected, although some at low (<1%) frequency. These less common organisms might be clinically relevant as many are related to opportunistic pathogens not previously identified in CF samples (Supplementary Table S2). Although CF does not cause immunodeficiency, CF patients' innate defenses appear to be impaired and they succumb to opportunistic pathogen infections in much the same way as immunocompromised patients (Govan and Deretic, 1996; Tummler and Kiewitz, 1999).

A few of the obligate anaerobes found through alternative culture techniques (for example, *Veillonella*, *Prevotella* and *Porphyromonas*) are also opportunistic pathogens. Given that potential pathogens remain undetected by traditional culture-based analysis, culturing conditions need to change and diagnostic tools that are independent of culture need to be incorporated into clinical screening of CF patients.

Microbial interactions

CF patient deterioration is largely an indirect effect of infection; lung damage is predominantly the result of a neutrophil-dominated chronic inflammatory response to the bacteria (Page and Kornman, 1997). Thus, the diverse bacteria detected, including uncultured organisms, may contribute to the inflammatory response. Consistent with this hypothesis is the observation that the first signs of inflammation in infants generally predate the presence of culturable organisms (Balough *et al.*, 1995; Khan *et al.*, 1995). Although these data have been interpreted to mean that CF pathology contains an additional immune component, an alternative explanation is that uncultured bacteria infect the CF lung early on in life. The clinical importance of uncultured organisms such as the TM7 and unclassified Clostridiales and Bacteroidetes merits investigation.

Although possible interactions among bacterial species in CF bronchial infections have been largely ignored, several studies have begun to elucidate the ways in which microorganisms might interact within the CF lung. For example, some strains from oropharyngeal sputum samples of CF patients modulate gene expression of *P. aeruginosa*, causing increased virulence of *P. aeruginosa* in a rat lung infection model (Duan *et al.*, 2003). *P. aeruginosa* cells lyse *S. aureus* and can also inhibit the growth of *Staphylococcus epidermis* through membrane vesicles (Mashburn and Whiteley, 2005, Mashburn *et al.*, 2005). *In vitro* studies on interactions between the fungus *Candida albicans*, a known colonizer of the CF lung, and *P. aeruginosa* reveal an antagonistic effect of *P. aeruginosa* cells on the filamentous form of *C. albicans* (Hogan and Kolter, 2002). All these studies involved isolating microorganisms first and then investigating pairwise interactions *in vitro* and may represent only 'the tip of the iceberg' of a dynamic and interactive microbial community in the CF lung. As such, viewing the CF lung as an environmental habitat, with niches ranging from oxic to anoxic, and complex microbial communities and interactions rather than a mono- or coculture, could lead to a better understanding of the clinical

repercussions various bacteria have in disease progression, and may facilitate advances in the treatment of CF and other polymicrobial diseases.

Acknowledgements

We thank the Children's Hospital Boston Pulmonary Fellows for sample collection, Joe Jones for technical advice on pyrosequencing and Kristina Fontanez for assistance with the pyrosequencing analysis. This work was supported by a Rubicon grant by the Netherlands Organisation for Scientific Research (NWO) to GR, a Harvard University Microbial Sciences Initiative (MSI) postdoctoral fellowship to AMG, a Howard Hughes Predoctoral Fellowship to ILGN, an internal grant from the University of South Carolina Office of Research and Health Sciences to CRY, grants from NIH (GM068516) to SL, and NSF (OCE-0453901) and Harvard University Center for the Environment (HUCE) to CMC.

References

Aas JA, Paster BJ, Stokes LN, Olsen I, Dewhirst FE. (2005). Defining the normal bacterial flora of the oral cavity. *J Clin Microbiol* **43**: 5721–5732.

Al Masalma M, Armougom F, Scheld WM, Dufour H, Roche PH, Drancourt M *et al.* (2009). The expansion of the microbiological spectrum of brain abscesses with use of multiple 16S ribosomal DNA sequencing. *Clin Infect Dis* **48**: 1169–1178.

Alvarez-Ortega C, Harwood CS. (2007). Responses of *Pseudomonas aeruginosa* to low oxygen indicate that growth in the cystic fibrosis lung is by aerobic respiration. *Mol Microbiol* **65**: 153–165.

Armougom F, Bittar F, Stremler N, Rolain JM, Robert C, Dubus JC *et al.* (2009). Microbial diversity in the sputum of a cystic fibrosis patient studied with 16S rDNA pyrosequencing. *Eur J Clin Microbiol Infect Dis* **28**: 1151–1154.

Balough K, McCubbin M, Weinberger M, Smits W, Ahrens R, Fick R. (1995). The relationship between infection and inflammation in the early stages of lung disease from cystic fibrosis. *Pediatr Pulmonol* **20**: 63–70.

Bals R, Weiner DJ, Wilson JM. (1999). The innate immune system in cystic fibrosis lung disease. *J Clin Invest* **103**: 303–307.

Bik EM, Eckburg PB, Gill SR, Nelson KE, Purdom EA, Francois F *et al.* (2006). Molecular analysis of the bacterial microbiota in the human stomach. *Proc Natl Acad Sci USA* **103**: 732–737.

Bittar F, Richet H, Dubus JC, Reynaud-Gaubert M, Stremler N, Sarles J *et al.* (2008). Molecular detection of multiple emerging pathogens in sputa from cystic fibrosis patients. *PloS ONE* **3**: e2908.

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

Chao A. (1987). Estimating the population-size for capture recapture data with unequal catchability. *Biometrics* **43**: 783–791.

Coenye T, Goris J, Spilker T, Vandamme P, LiPuma JJ. (2002). Characterization of unusual bacteria isolated from respiratory secretions of cystic fibrosis patients and description of *Inquilinus limosus* gen. nov, sp. nov. *J Clin Microbiol* **40**: 2062–2069.

Daims H, Bruhl A, Amann R, Schleifer KH, Wagner M. (1999). The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: development and evaluation of a more comprehensive probe set. *Syst Appl Microbiol* **22**: 434–444.

De Backer D, Creteur J, Zhang H, Norrenberg M, Vincent JL. (1997). Lactate production by the lungs in acute lung injury. *Am J Respir Crit Care Med* **156**: 1099–1104.

Duan K, Dammel C, Stein J, Rabin H, Surette MG. (2003). Modulation of *Pseudomonas aeruginosa* gene expression by host microflora through interspecies communication. *Mol Microbiol* **50**: 1477–1491.

Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M *et al.* (2005). Diversity of the human intestinal microbial flora. *Science* **308**: 1635–1638.

Flanagan JL, Brodie EL, Weng L, Lynch SV, Garcia O, Brown R *et al.* (2007). Loss of bacterial diversity during antibiotic treatment of intubated patients colonized with *Pseudomonas aeruginosa*. *J Clin Microbiol* **45**: 1954–1962.

Gibson RL, Burns JL, Ramsey BW. (2003). Pathophysiology and management of pulmonary infections in cystic fibrosis. *Am J Respir Crit Care Med* **168**: 918–951.

Gomez MI, Prince A. (2007). Opportunistic infections in lung disease: *Pseudomonas* infections in cystic fibrosis. *Curr Opin Pharmacol* **7**: 244–251.

Govan JR, Deretic V. (1996). Microbial pathogenesis in cystic fibrosis: mucoid *Pseudomonas aeruginosa* and *Burkholderia cepacia*. *Microbiol Rev* **60**: 539–574.

Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R. (2008). Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat Methods* **5**: 235–237.

Hamp TJ, Jones WJ, Fodor AA. (2009). Effects of experimental choices and analysis noise on surveys of the 'rare biosphere'. *Appl Environ Microbiol* **75**: 3263–3270.

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.

Hogan DA, Kolter R. (2002). *Pseudomonas-Candida* interactions: an ecological role for virulence factors. *Science* **296**: 2229–2232.

Hugenholtz P, Tyson GW, Webb RI, Wagner AM, Blackall LL. (2001). Investigation of candidate division TM7: a recently recognized major lineage of the domain Bacteria with no known pure-culture representatives. *Appl Environ Microbiol* **67**: 411–419.

Keijser BJF, Zaura E, Huse SM, van der Vossen JMBM, Schuren FHJ, Montijn RC *et al.* (2008). Pyrosequencing analysis of the oral microflora of healthy adults. *J Dent Res* **87**: 1016–1020.

Khan TZ, Wagener JS, Bost T, Martinez J, Accurso FJ, Riches DW. (1995). Early pulmonary inflammation in infants with cystic fibrosis. *Am J Respir Crit Care Med* **151**: 1075–1082.

Kumar Y, Westram R, Kipfer P, Meier H, Ludwig W. (2006). Evaluation of sequence alignments and oligonucleotide probes with respect to three-dimensional structure of ribosomal RNA using ARB software package. *BMC Bioinformatics* **7**: 240.

Kunin V, Engelbrektson A, Ochman H, Hugenholtz P. (2009). Wrinkles in the rare biosphere: pyrosequencing errors lead to artificial inflation of diversity estimates. *Environ Microbiol* **12**: 118–123.

Lane DJ. (1991). 16S/23S rRNA sequencing. In: Stackebrandt, E (ed). *Nucleic Acid Techniques in Bacterial Systematics*. John Wiley & Sons: New York, pp. 115–175.

Liu Z, DeSantis TZ, Andersen GL, Knight R. (2008). Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic Acids Res* **36**: e120.

Lyczak JB, Cannon CL, Pier GB. (2002). Lung infections associated with cystic fibrosis. *Clin Microbiol Rev* **15**: 194–222.

Maddison D, Maddison W. (2000). *MacClade 4: Analysis Phylogeny and Character Evolution*. Sinauer Associates: Sunderland, MA, USA.

Mashburn LM, Jett AM, Akins DR, Whiteley M. (2005). *Staphylococcus aureus* serves as an iron source for *Pseudomonas aeruginosa* during *in vivo* coculture. *J Bacteriol* **187**: 554–566.

Mashburn LM, Whiteley M. (2005). Membrane vesicles traffic signals and facilitate group activities in a prokaryote. *Nature* **437**: 422–425.

Matsui H, Wagner VE, Hill DB, Schwab UE, Rogers TD, Button B *et al.* (2006). A physical linkage between cystic fibrosis airway surface dehydration and *Pseudomonas aeruginosa* biofilms. *Proc Natl Acad Sci USA* **103**: 18131–18136.

Nawrocki EP, Eddy SR. (2007). Query-dependent banding (QDB) for faster RNA similarity searches. *PLoS Comput Biol* **3**: e56.

Ng SK, Hamilton IR. (1971). Lactate metabolism by *Veillonella parvula*. *J Bacteriol* **105**: 999–1005.

Page RC, Kornman KS. (1997). The pathogenesis of human periodontitis: an introduction. *Periodontol* **14**: 9–11.

Palmer KL, Aye LM, Whiteley M. (2007). Nutritional cues control *Pseudomonas aeruginosa* multicellular behavior in cystic fibrosis sputum. *J Bacteriol* **189**: 8079–8087.

Paster BJ, Boches SK, Galvin JL, Ericson RE, Lau CN, Levanos VA *et al.* (2001). Bacterial diversity in human subgingival plaque. *J Bacteriol* **183**: 3770–3783.

Paster BJ, Dewhirst FE, Coleman BC, Lau CN, Ericson RL. (1998). Phylogenetic analysis of cultivable oral treponemes from the Smibert collection. *Int J Syst Bacteriol* **48**(Part 3): 713–722.

Paster BJ, Olsen I, Aas JA, Dewhirst FE. (2006). The breadth of bacterial diversity in the human periodontal pocket and other oral sites. *Periodontol* **42**: 80–87.

Pedron T, Sansonetti P. (2008). Commensals bacterial pathogens and intestinal inflammation: an intriguing menage a trois. *Cell Host Microbe* **3**: 344–347.

Pei ZH, Bini EJ, Yang LY, Zhou MS, Francois F, Blaser MJ. (2004). Bacterial biota in the human distal esophagus. *Proc Natl Acad Sci USA* **101**: 4250–4255.

Polz MF, Cavanaugh CM. (1998). Bias in template-to-product ratios in multitemplate PCR. *Appl Environ Microbiol* **64**: 3724–3730.

Price LB, Liu CM, Melendez JH, Frankel YM, Engelthaler D, Aziz M *et al.* (2009). Community analysis of chronic wound bacteria using 16S rRNA gene-based pyrosequencing:impact of diabetes and antibiotics on chronic wound microbiota. *PloS ONE* **4**: e6462.

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.

Seeliger S, Janssen PH, Schink B. (2002). Energetics and kinetics of lactate fermentation to acetate and propionate via methylmalonyl-CoA or acrylyl-CoA. *Fems Microbiol Lett* **211**: 65–70.

Shannon CE, Weaver W. (1963). *The Mathematical Theory of Communication*. University of Illinois Press: Urbana, IL, USA.

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.

Smith JJ, Travis SM, Greenberg EP, Welsh MJ. (1996). Cystic fibrosis airway epithelia fail to kill bacteria because of abnormal airway surface fluid. *Cell* **85**: 229–236.

Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR *et al.* (2006). Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proc Natl Acad Sci USA* **103**: 12115–12120.

Stamatakis A, Hoover P, Rougemont J. (2008). A Rapid Bootstrap Algorithm for the RAxML Web Servers. *Syst Biol* **57**: 758–771.

Thompson JR, Marcelino LA, Polz MF. (2002). Heteroduplexes in mixed-template amplifications: formation consequence and elimination by 'reconditioning PCR'. *Nucleic Acids Res* **30**: 2083–2088.

Tringe SG, Zhang T, Liu X, Yu Y, Lee WH, Yap J *et al.* (2008). The airborne metagenome in an indoor urban environment. *PLoS ONE* **3**: e1862.

Tummler B, Kiewitz C. (1999). Cystic fibrosis: an inherited susceptibility to bacterial respiratory infections. *Mol Med Today* **5**: 351–358.

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.

Wahab AA, Janahi IA, Marafia MM, El-Shafie S. (2004). Microbiological identification in cystic fibrosis patients with CFTR I1234V mutation. *J Trop Pediatr* **50**: 229–233.

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

Worlitzsch D, Tarran R, Ulrich M, Schwab U, Cekici A, Meyer KC *et al.* (2002). Effects of reduced mucus oxygen concentration in airway *Pseudomonas* infections of cystic fibrosis patients. *J Clin Invest* **109**: 317–325.

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