View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Dartmouth Digital Commons (Dartmouth College)

Dartmouth College

Dartmouth Digital Commons

Dartmouth Scholarship

Faculty Work

6-29-2012

Prevalence of Streptococci and Increased Polymicrobial Diversity Associated with Cystic Fibrosis Patient Stability

L. M. Filkins
Dartmouth College

T. H. Hampton
Dartmouth College

A. H. Gifford
Dartmouth College

M. J. Gross
Dartmouth College

D. A. Hogan
Dartmouth College

See next page for additional authors

Follow this and additional works at: https://digitalcommons.dartmouth.edu/facoa

Part of the Infectious Disease Commons, Medical Microbiology Commons, and the Respiratory System Commons

Dartmouth Digital Commons Citation

Filkins, L. M.; Hampton, T. H.; Gifford, A. H.; Gross, M. J.; Hogan, D. A.; Sogin, M. L.; Morrison, H. G.; Paster, B. J.; and O'Toole, G. A., "Prevalence of Streptococci and Increased Polymicrobial Diversity Associated with Cystic Fibrosis Patient Stability" (2012). *Dartmouth Scholarship*. 1053.
https://digitalcommons.dartmouth.edu/facoa/1053

This Article is brought to you for free and open access by the Faculty Work at Dartmouth Digital Commons. It has been accepted for inclusion in Dartmouth Scholarship by an authorized administrator of Dartmouth Digital Commons. For more information, please contact dartmouthdigitalcommons@groups.dartmouth.edu.

Authors
L. M. Filkins, T. H. Hampton, A. H. Gifford, M. J. Gross, D. A. Hogan, M. L. Sogin, H. G. Morrison, B. J. Paster, and G. A. O'Toole

This article is available at Dartmouth Digital Commons: https://digitalcommons.dartmouth.edu/facoa/1053

JB Journals.ASM.org

Prevalence of Streptococci and Increased Polymicrobial Diversity Associated with Cystic Fibrosis Patient Stability

L. M. Filkins,[a] T. H. Hampton,[a] A. H. Gifford,[b] M. J. Gross,[a] D. A. Hogan,[a] M. L. Sogin,[c] H. G. Morrison,[c] B. J. Paster,[d,e] and G. A. O'Toole[a]

Geisel School of Medicine at Dartmouth, Hanover, New Hampshire, USA[a]; Dartmouth-Hitchcock Medical Center, Section of Pulmonary and Critical Care Medicine, Lebanon, New Hampshire, USA[b]; Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, Massachusetts, USA[c]; The Forsyth Institute, Department of Molecular Genetics, Cambridge, Massachusetts, USA[d]; and Department of Oral Medicine, Infection and Immunity, Harvard School of Dental Medicine, Boston, Massachusetts, USA[e]

Diverse microbial communities chronically colonize the lungs of cystic fibrosis patients. Pyrosequencing of amplicons for hypervariable regions in the 16S rRNA gene generated taxonomic profiles of bacterial communities for sputum genomic DNA samples from 22 patients during a state of clinical stability (outpatients) and 13 patients during acute exacerbation (inpatients). We employed quantitative PCR (qPCR) to confirm the detection of *Pseudomonas aeruginosa* and *Streptococcus* by the pyrosequencing data and human oral microbe identification microarray (HOMIM) analysis to determine the species of the streptococci identified by pyrosequencing. We show that outpatient sputum samples have significantly higher bacterial diversity than inpatients, but maintenance treatment with tobramycin did not impact overall diversity. Contrary to the current dogma in the field that *Pseudomonas aeruginosa* is the dominant organism in the majority of cystic fibrosis patients, *Pseudomonas* constituted the predominant genera in only half the patient samples analyzed and reported here. The increased fractional representation of *Streptococcus* in the outpatient cohort relative to the inpatient cohort was the strongest predictor of clinically stable lung disease. The most prevalent streptococci included species typically associated with the oral cavity (*Streptococcus salivarius* and *Streptococcus parasanguis*) and the *Streptococcus milleri* group species. These species of *Streptococcus* may play an important role in increasing the diversity of the cystic fibrosis lung environment and promoting patient stability.

Progressive decline in lung function is the primary cause of morbidity and mortality in cystic fibrosis (CF) patients (5). Mutations in the cystic fibrosis transmembrane conductance regulator protein promote dehydration of the airway surface liquid of the lung epithelium, which leads to defective mucociliary clearance and increased bacterial colonization of CF patient lungs (10). The inflammatory response to bacterial colonization causes irreversible lung tissue damage that progressively decreases lung function (2, 15).

Historically, CF bacterial infections were attributed to very few species, predominately *Pseudomonas aeruginosa* in adults (9). However, studies over the past decade demonstrate that the CF lung environment can host a highly diverse polymicrobial community (22–24). The role of diverse community structure and interspecies interactions in patient health remains poorly understood. Previous reports show that increased age in CF patients correlates with decreased lung diversity and decreased lung function (measured as forced expiratory volume in 1 s [FEV_1]) (4). Further, community composition or gene expression, rather than total bacterial load, may dictate patient disease state (30), although all of these factors likely can contribute to patient health and disease.

While *Pseudomonas* remains a significant pathogen in CF, numerous other aerobic and anaerobic organisms contribute to the community complexity of the CF lung, including members of the following genera: *Rothia*, *Prevotella*, *Stenotrophomonas*, *Streptococcus*, and others (11, 31). Recent reports describe the prevalence of viridans streptococci in the analysis of patient sputum, including the salivarius, milleri (or anginosus), and mitis streptococcus groups (17). The *Streptococcus milleri* group (SMG), which includes *S. anginosus*, *S. constellatus*, and *S. intermedius*, may play a significant role in the pathogenesis of the CF lung (18, 28) by influencing microbial community structures during periods of clinical stability or by causing acute exacerbation in diseased patients (27).

Here, we report that increased bacterial community diversity in CF patient sputum and increased relative abundance of the genus *Streptococcus* positively correlated with patient stability. The predominant streptococci in these patient samples included *S. salivarius*, *S. parasanguis*, and SMG species. A complete understanding of the polymicrobial communities of CF patients, as well as characterizing community features that correlate with exacerbation, can guide innovations to current treatment methods in order to personalize and improve patient care.

MATERIALS AND METHODS

Patient cohort and sputum collection. The study enrollment, assessment of inpatients versus outpatients, and a description of the clinical characteristics of this patient cohort were detailed by Gifford et al. (8). Spontaneously expectorated sputum samples were collected during routine visits for outpatient samples. Spontaneously expectorated inpatient sputum samples were collected within 24 h of hospital admittance, with the exception of one sample collected within the first 72 h. For inpatients, intravenous antibiotic exposure length prior to sputum expectoration was lim-

Received 4 April 2012 Accepted 25 June 2012

Published ahead of print 29 June 2012

Address correspondence to G. A. O'Toole, georgeo@dartmouth.edu.

Supplemental material for this article may be found at http://jb.asm.org/.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JB.00566-12

TABLE 1 Primers used in qPCR studies

Primer	Sequence (5′→3′)	Amplification target	Target species	Reference(s)
Universal For	GTGSTGCAYGGYTGTCGTCA	16S rRNA genes	All bacteria	13, 16
Universal Rev	ACGTCRTCCMCACCTTCCTC	16S rRNA genes	All bacteria	13, 16
rplU For	GCAGCACAAAGTCACCGAAGG	*rplU*	*P. aeruginosa*	This study
rplU Rev	CCGTGGGAAACCACTTCAGC	*rplU*	*P. aeruginosa*	This study
Str1	GTACAGTTGCTTCAGGACGTATC	*tuf*	*Streptococcus*[a]	21
Str2	ACGTTCGATTTCATCACGTTG	*tuf*	*Streptococcus*[a]	21

[a] Detects the genus *Streptococcus*.

ited compared to the length of hospital admission (2-week average duration). Internal review board (IRB) approval was obtained from the Center for Protection of Human Subjects at Dartmouth College (CPHS number 21473), and patients provided written informed consent.

Bacterial strains and growth conditions. For the preparation of genomic DNA (gDNA) from bacterial colonies, we grew *Streptococcus* strains aerobically at 37°C, 5% CO_2 overnight; *Fusobacterium nucleatum*, *Prevotella intermedia*, and *Gemella* sp. at 37°C for 48 h anaerobically on tryptic soy agar (TSA)-5% blood agar plates (Northeast Laboratory Services); and *Pseudomonas aeruginosa* in LB broth at 37°C overnight.

gDNA isolation and patient sputum sample preparation. We employed a modification of the Gentra PureGene Yeast/Bact. kit to isolate gDNA. We passed patient sputum samples resuspended and diluted 2-fold to 5-fold in Tris-EDTA (TE)–0.08% dithiothreitol (DTT) successively through syringes with 16-, 20-, and 23-gauge needles until homogenous. Following treatment for 30 min at 37°C with 3 mg/ml of lysozyme (final concentration), we incubated the samples in cell lysis buffer (Gentra) for 15 min at 80°C. The remainder of gDNA isolation followed the manufacturer's protocol. This isolated gDNA was used for deep sequencing, quantitative real-time PCR (qPCR) studies, and human oral microbe identification microarray (HOMIM) analysis. gDNA for qPCR controls was also prepared using the Gentra Puregene Yeast/Bact. kit, according to the manufacturer's instructions for Gram-positive or Gram-negative species, as appropriate.

Deep-sequencing analysis. For pyrotag analyses of the V4V6 rRNA hypervariable regions in patient sputum gDNA samples, we prepared amplicon libraries using fused primers that contained either the A or B 454 Titanium adapter (Roche Diagnostics), a unique 5-nucleotide (nt) multiplex identifier (MID) for each gDNA sample, and either the conserved 16S rRNA 518F oligonucleotide 5′ CCAGCAGCYGCGGTAAN or 1064R, 5′ CGACRRCCATGCANCACCT (*Escherichia coli* 16S rRNA positions). All MIDs differ by at least two bases and contain no homopolymers. A master mix contained 1× Platinum HiFi *Taq* polymerase buffer, 1.6 units Platinum HiFi polymerase (Life Technologies, Carlsbad CA), 3.7 mM $MgSO_4$, 200 μM deoxynucleoside triphosphates (dNTPs) (PurePeak polymerization mix; ThermoFisher, East Providence RI), and 5 to 20 ng of gDNA brought to a final volume of 100 μl. To mitigate influence of early PCR errors, we routinely divided the samples into three replicate reactions and prepared a no-template negative control for each MID. PCR cycling conditions included initial denaturation at 94°C for 3 min; 30 cycles of 94°C for 30 s, 60°C for 45 s, and 72°C for 1 min; and a final extension at 72°C for 2 min. Replicate samples were pooled to contain equal amounts of amplicon libraries, based on PicoGreen quantification. After pooling the three replicates and evaluating the quality of the PCR amplicons and negative control on a LabChip GX (Caliper, Hopkinton MA), we used a 0.75 volume of Ampure beads (Beckman Coulter, Brea, CA) to remove products under 300 bp. Subsequent to pooling as many as 40 amplicons with unique MIDs, we prepared emulsified PCRs and enrichments according to the Roche Titanium amplicon sequencing protocols (Lib-A emPCR reagents, XLR sequencing reagents, two-region PicoTitre plate) to generate 10,000 to 20,000 pyrotags per library.

Bioinformatics processing. A custom bioinformatic pipeline at the Marine Biological Laboratory performed quality filtering to remove low-quality (average quality scores less than 30) reads and sequences lacking exact primer matches or containing ambiguous bases (Ns). The algorithm UChime (7), combining both the *de novo* and reference database (ChimeraSlayer GOLD) modes, removed chimeric reads. The algorithm GAST assigned taxonomy to each unique read (14), and UCLUST (6) identified operational taxonomic units (OTUs) with 97% sequence identity. The website Visualization and Analysis of Microbial Population Structures (http://vamps.mbl.edu) provides access to individual reads, taxon assignments, and descriptions of individual clusters.

qPCR primer set validation. Table 1 specifies the qPCR primer sets, including the universal primer set originally described by Maeda et al. (16) and evaluated by Horz et al. (13) for broad-range amplification of bacterial species. All primer sets were verified for specificity using the BLAST database. The 15 species that served as controls included *Pseudomonas aeruginosa*, *Streptococcus anginosus*, *Streptococcus constellatus*, *Streptococcus intermedius*, *Streptococcus mitis*, *Streptococcus oralis*, *Streptococcus parasanguis*, *Streptococcus pneumoniae*, *Streptococcus salivarius*, *Streptococcus sanguinis*, *Prevotella intermedia*, *Fusobacterium nucleatum*, *Gemella haemolysans*, *Gemella morbillorum* and *Gemella sanguinis*. Detection of these 15 species with the universal primer set yielded comparable amplification, with a detection limit of 10^3 16S rRNA gene copies. The *rplU* For/Rev primer set specifically amplifies *Pseudomonas aeruginosa* with <1.0% nonspecific amplification of all non-*Pseudomonas aeruginosa* control species specified above. For these specificity assays, we assigned 100% amplification for each control species based on total quantification using the universal primer set and then calculated nonspecific amplification by the *rplU* For/Rev primer set accordingly. Similarly, the Str1/2 primer set, specific for streptococci, amplifies all tested streptococcal species with equivalent sensitivity with the exception of *S. intermedius*, which was consistently underrepresented by 5-fold in our optimization assays. The Str1/2 primer set yielded <0.1% nonspecific amplification of nonstreptococcal species, as defined above for *P. aeruginosa*. The sensitivity of both *rplU* For/Rev and Str1/2 was <25 gene copies. Each primer set had an efficiency of >90% for amplification of the intended species.

Detection of *Pseudomonas aeruginosa* and *Streptococcus* by qPCR. gDNA purified from control strains was quantified by NanoDrop. We prepared 10-fold dilutions of gDNA, targeting 16S rRNA gene copy numbers of 1×10^6 to 1×10^2, based on genome size and 16S rRNA gene copy number for each species. We assumed a copy number of four for species that lacked complete genome sequences. We detected 16S rRNA genes by qPCR, using the universal primer set and 2× iQ SYBR Green supermix (Bio-Rad) and assigned correction factors for each species' dilution series based on detection with the universal primer set to account for error in NanoDrop quantification, copy number estimation, and dilution preparation. *Pseudomonas aeruginosa* was assigned a correction factor of 1.0, and all other species were assigned accordingly. We prepared a control mix from the individually prepared species dilutions to contain 63% *Pseudomonas aeruginosa*, 24% *Streptococcus*, 8% *Prevotella intermedia*, 3% *Fusobacterium nucleatum*, and 2% *Gemella* 16S rRNA gene copies (after correction factor calculations). The *Streptococcus* fraction contained equivalent 16S rRNA gene quantities from each of the nine streptococcal species listed above, and the *Gemella* fraction contained equal quantities of the three *Gemella* species listed above.

We analyzed patient samples and control mixes in replicates of six for detection with the universal For/Rev, *rplU* For/Rev, and Str1/2 primer

sets. A total of 10 ng of total patient gDNA was analyzed in each well, and the bacterial portion was 0.02 to 0.2 ng, with the remaining DNA derived primarily from the host. We prepared standard curves for each primer set based on quantification cycle (Cq) values for detection of the control mix dilution series and the known quantities of total 16S rRNA genes, *Pseudomonas aeruginosa* 16S rRNA genes, or *Streptococcus* 16S rRNA genes in the control mix. We adjusted the standard curves to account for the fact that *rplU* For/Rev and Str1/2 detect genes that are present in single copy number.

We determined the contribution of 16S rRNA genes from *Pseudomonas aeruginosa* and *Streptococcus* in the patient samples via comparison of replicate Cq detection values to the standard curve for each primer set. The fraction of *Pseudomonas aeruginosa* or *Streptococcus* was further calculated as follows: *Pseudomonas aeruginosa*-specific 16S rRNA gene detection/universal 16S detection or *Streptococcus*-specific 16S rRNA gene detection/universal 16S detection, respectively (see Table 1).

HOMIM. We analyzed purified gDNA from patient sputum samples by using the HOMIM at The Forsyth Institute. A total of 20 to 30 μl gDNA per sample was analyzed, with a target of ~20 ng/μl bacterial DNA, when possible. We prepared gDNA samples and analyzed these samples as described on the HOMIM website (http://mim.forsyth.org/protocol.html). This method was initially published and reviewed by Paster et al. (19, 20).

Statistical analysis. We developed heat maps based on Pearson hierarchical clustering according to two parameters: (i) patient sample and (ii) prevalence of microbial genera or species. We calculated a *P* value for the association of clinical status and microbial community composition by Fisher's exact test.

Box and whisker plots were used throughout the study to compare microbiome data with clinical phenotypes of interest. The bolded middle line of the box represents a median value, and the upper and lower ranges of the box represent the 75% and 25% quartiles, respectively. The whiskers depict 1.5× the interquartile range. Open circles display data points that fall outside 1.5× the interquartile range. We calculated *P* values for all box-and-whisker plots by the nonparametric Mann-Whitney U test.

Comparison of the fractional representation of individual genera in inpatient or outpatient samples was displayed in a transition plot. We calculated significant differences between means by *t* test using Benjamini-Hochberg-corrected *P* values.

We plotted 454 pyrosequencing results and median fraction values for qPCR detection of *Pseudomonas* and *Streptococcus* to show relationships between these data and calculated Pearson correlation lines and *P* values, excluding the indicated outlier points, as described thoroughly in Results.

RESULTS

Characterization of the microbiome of sputum from CF patients. The patient cohort studied here includes 39 CF patients recruited as part of a previously reported cross-sectional study (8). At the time of sputum collection, 25 participants were clinically stable (outpatient status) and 14 participants were in a state of clinical worsening (inpatient status). The average age of the outpatient population was 27.5 years (range, 19 to 52 years) and was 27.5 years (range, 22 to 45 years) for the inpatient group. The outpatient group consisted of 44% females and, similarly, the inpatient group consisted of 43% females. Relatively low FEV_1 was characteristic of patients during exacerbation, which correlated with inpatient status. The inpatient cohort had a median FEV_1 of 27% predicted, while the outpatient population had a median FEV_1 of 56% predicted. Additionally, both the inpatient and outpatient groups included participants who were either on or off maintenance tobramycin treatment at the time of sample collection. Further clinical description and analysis of this patient cohort was previously reported (8).

Genomic DNA (gDNA) was prepared from spontaneously produced patient sputum samples and the 16S rRNA gene profiles characterized by 454 pyrosequencing of amplicon libraries, as reported by Sogin et al. (29) and Huse et al. (14), for 35 of the 39 patients (22 outpatients and 13 inpatients). Genus assignments were resolved for >99% of the total deep-sequencing reads.

A total of 138 genera were assigned for the 35 patient samples. A complete summary of read assignments for each patient is provided in Tables S1 and S2 in the supplemental material. The most dominant genera in this sample set included *Pseudomonas*, *Streptococcus*, *Fusobacterium*, and *Prevotella*, followed by *Rothia*, *Stenotrophomonas*, *Staphylococcus*, *Haemophilus*, *Gemella*, and *Neisseria* (Fig. 1A). These 10 genera accounted for 93% of the reads detected from the patient samples. While not all 10 of the top genera were present in every patient, each genus accounted for at least 1% of the total reads for the patient population as a whole. The presence of these genera in CF patient lungs agrees with the findings reported previously (3, 26, 32).

Figure 1B shows patient samples clustered according to deep-sequencing profiles of the overall most abundant four genera, which accounted for 86% of the total reads. These top-four genera are also highly prevalent in the patient population, meaning that they are present in most patients. Remaining genera are highly abundant (when present), highly prevalent (but not highly abundant), or neither highly abundant nor prevalent (see Fig. S1 in the supplemental material). While some low-prevalence and/or low-abundance organisms may have significant biological impact, for initial profiling of this patient cohort, only the top-four most abundant and prevalent organisms were analyzed. Focusing on only the top-four genera of the data set, we have identified four exploratory community profiles in which the patient samples cluster: (i) high *Pseudomonas*, low *Streptococcus*; (ii) medium *Pseudomonas*, medium *Streptococcus*; (iii) low *Pseudomonas*, high *Streptococcus*; and (iv) low *Pseudomonas*, low *Streptococcus*, but high other (predominantly *Fusobacterium* or *Prevotella*).

Overall, the association between clinical status and these four exploratory clusters is greater than one would expect by chance ($P = 0.01$, Fisher's exact test), indicating that most clusters are significantly enriched in samples derived from uniquely inpatients or outpatients. However, predominance of *Pseudomonas* in a patient sample is not predictive of patient status (as profile 1 is composed of both inpatient samples and outpatient samples). Figure 1B also displays that *Pseudomonas* constitutes the single predominant genus in only about half of the patient samples from this cohort, whereas the remaining samples are predominated by a combination of *Pseudomonas* and *Streptococcus* or other prevalent genera.

Increased diversity correlates with outpatient status. Our comparisons of microbial profiles with patient clinical information (8) revealed several significant correlations. (i) Outpatient samples had significantly higher diversity than inpatient samples (Fig. 2A). Differences in the deep-sequencing effort for each group do not explain this correlation since the outpatient and inpatient samples had a comparable number of reads (see Fig. S2 in the supplemental material). (ii) The state of being on or off maintenance treatment with tobramycin at the time of sputum collection (irrespective of clinical status) did not significantly affect microbial diversity (Fig. 2B), although we observed a trend toward higher diversity for those patients off tobramycin. Similarly, Zhao and colleagues recently reported that maintenance antibiotic administration had a minimal impact on the microbial diversity in sputum samples from a longitudinal study of six patients. How-

FIG 1 Characterization of the polymicrobial communities of sputum samples from cystic fibrosis inpatients and outpatients. (A) Fraction of 454 pyrosequencing reads assigned to each of the top-10 genera detected in the patient sample set as a whole is shown for sputum samples analyzed from inpatients (INPT; $n = 13$) and outpatients (OUTPT; $n = 22$) from a cross-sectional study. The legend indicates the color assigned to each indicated genus. (B) Heat map of patient samples based on Pearson hierarchical clustering of relative bacterial abundance (by deep sequencing) for the most prevalent four genera, which account for ~86% of the total pyrosequencing reads. Patients can be described as one of four profiles: (i) high *Pseudomonas*, low *Streptococcus*; (ii) medium *Pseudomonas*, medium *Streptococcus*; (iii) low *Pseudomonas*, high *Streptococcus*; (iv) low *Pseudomonas*, low *Streptococcus*, but high other (primarily *Fusobacterium* or *Prevotella*). The group number is indicated in the bar at the bottom of the panel, and the inset key indicates the color associated with the fraction of reads for each patient and genera assigned.

ever, episodic antibiotic treatment of pulmonary exacerbations significantly decreased microbial diversity during their longitudinal study (32). (iii) Maintenance treatment with tobramycin did not correlate with the fraction of *Pseudomonas* or *Streptococcus* in the sputum samples (see Fig. S3 in the supplemental material). (iv) Patient age did not correlate with diversity for any comparisons (data not shown). This finding is also consistent with the results described in Zhao et al., where no significant correlation between microbial community diversity and age was observed (32). (v) The fraction of *Pseudomonas* did not correlate with any clinical parameters collected for this patient cohort. Of particular note, no significant correlation was observed between the fraction of *Pseudomonas* and inpatient or outpatient status (Fig. 3A).

***Streptococcus* abundance correlates with clinically stable lung disease.** Based on the finding that outpatients have increased bacterial diversity in their sputum, we explored correlations between patient status (i.e., inpatients versus outpatients) and prevalence of individual genera. The deep-sequencing data showed a significantly higher *Streptococcus* fraction in outpatients than in inpatients (Fig. 3B). The analysis shown in Fig. 1B confirms this observation by showing that inpatient samples clustered to the low proportion of streptococcus profiles, whereas more than 50% of the outpatients had a midlevel to high proportion of streptococcus. This correlation between increased *Streptococcus* fraction and patient stability is independent of the *Pseudomonas* fraction (Fig. 3A). While there is a trend toward increased *Pseudomonas* fraction in inpatients compared to that in outpatients, this difference is not statistically significant. Similarly, no significant correlation with patient status was detected for *Fusobacterium* or *Prevotella* (data not shown).

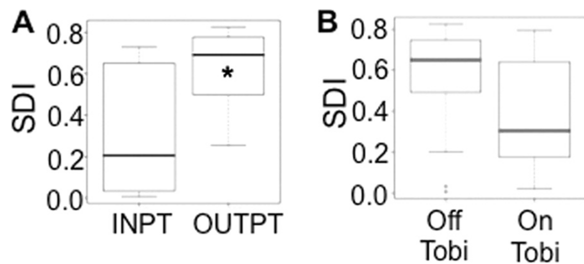

FIG 2 Increased diversity correlates with outpatient status and is not impacted by tobramycin treatment. Box-and-whisker plots of the Simpson diversity index based on the complete deep-sequencing profile of each sputum sample (each read assigned to a single genus). Box parameters, the bold line represents median diversity, while the upper and lower ranges of the box represent the 75% and 25% quartiles, respectively; whisker parameters, 1.5× the interquartile range; open circles, data points that fall outside 1.5× the interquartile range. *P* values were calculated using the nonparametric Mann-Whitney U test. (A) Inpatient (INPT; $n = 13$) samples compared to outpatient (OUTPT; $n = 22$) samples. The asterisk indicates *P* values <0.05. (B) Comparison of sputum samples collected from patients off tobramycin antibiotic treatment (Off Tobi; $n = 24$) compared to on tobramycin (On Tobi; $n = 11$), irrespective of patient status. No significant difference in diversity was detected.

Less predominant genera may also have a significant impact on microbial community dynamics and clinical outcome. Fractional representation of all genera assigned in this data set was compared for outpatients and inpatients (Fig. 3C). Most strikingly, the fractional representation of the genus *Gemella* (which accounts for ~2% of the total reads for this sample set) is greater than 30-fold higher in outpatients than in inpatients. Interestingly, *Haemophilus* (similarly, ~2% of the total reads) is also enriched in outpatient samples compared to in inpatient samples. The biological significance of the increased fractional representation of these organisms in the outpatient samples compared to in the inpatient samples is not understood at this time.

qPCR analysis confirms the genus population profiling determined by 454 pyrosequencing. Deep-sequencing analysis of polymicrobial samples by 454 pyrosequencing has gained widespread acceptance as a method for profiling clinical specimens. However, few reports independently verify the results of deep-sequencing data. To address this issue, we employed a qPCR method to confirm our deep-sequencing results. For these studies, we used a combination of species-specific and group-specific primer sets from previous publications or newly developed primers (Table 1). We validated all primers for their accurate detection of and specificity toward control gDNA from representative species of the top genera in the patient sputum (see Materials and Methods for details).

We used *Pseudomonas aeruginosa*-specific, *Streptococcus*-specific, and universal primer sets in qPCR assays of the 19 (out of the 35) patient sputum samples that contained sufficient gDNA for both pyrosequencing and qPCR analysis. The fraction of *Pseudomonas aeruginosa* and *Streptococcus* present in each patient sample is calculated based on comparison to standard curves developed from a known mixture of bacterial gDNA (see Materials and Methods for details).

qPCR measurement of the fraction of *Pseudomonas aeruginosa* in inpatient and outpatient samples positively correlated with the fraction of 454 pyrosequencing reads assigned to the *Pseudomonas* genus for each sample (Fig. 4A). The Pearson correlation of the

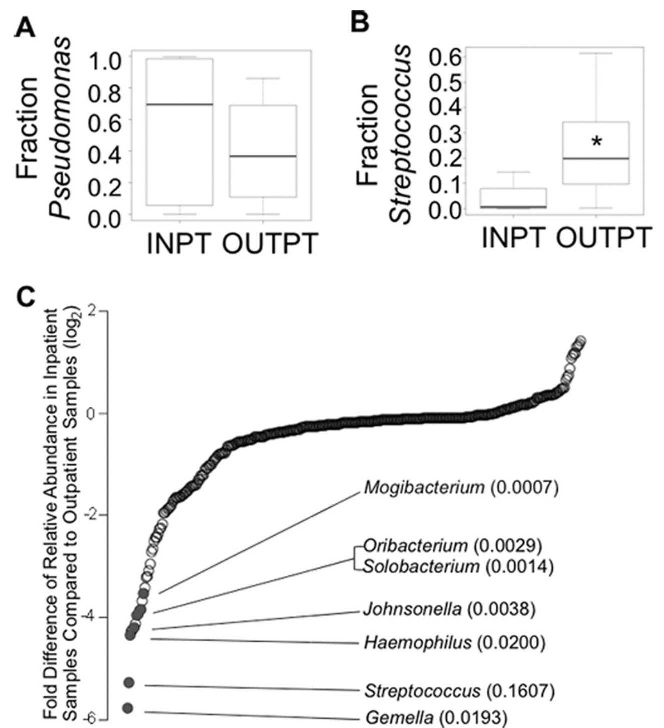

FIG 3 Increased *Streptococcus* fraction correlates with clinical patient stability. Box-and-whisker plots comparing the fraction of 454 pyrosequencing reads assigned to a single genus for inpatient and outpatient samples. Median fraction *Streptococcus* or *Pseudomonas* is represented by the bold middle line of the box. The extremities of the box represent the 75% and 25% quartiles. The whiskers of the plot indicate 1.5× the interquartile range. *P* values were calculated using the nonparametric Mann-Whitney U test. (A) Comparison of *Pseudomonas* fraction; (B) comparison of *Streptococcus* fraction. The asterisk indicates *P* values <0.05. (C) Transition plot comparing the fractional representation of taxonomic read assignments in inpatient samples to outpatient samples, based on 454 pyrosequencing results. Plotted circles correspond to the $\log_2$-fold difference in fractional representation of an individual taxon between the inpatient cohort and the outpatient cohort. *t* test comparisons of relative abundance in inpatient versus outpatient samples for all genera detected in the sample set were performed, and the reported *P* values are Benjamini-Hochberg corrected for multiple hypothesis testing. Key: open circles, no significant difference in fractional representation, $P > 0.05$; gray circles, significantly higher fractional representation in outpatient samples, $P \leq 0.05$. No detected genera were significantly enriched in inpatient samples compared to in outpatient samples after Benjamini-Hochberg correction. Genera are labeled in the order they are located on the transition plot (most significantly different at the bottom). Relative fractional representation in the total sample set is specified for each labeled genus.

Pseudomonas plot has a value of 0.968. Further, the regression line has a slope of 1.08, indicating that qPCR yielded marginally higher values than deep sequencing. qPCR results confirm that *Pseudomonas aeruginosa* accounts for the vast majority of *Pseudomonas* present in CF patient sputum. Similarly, we observed a direct correlation between the fraction of *Streptococcus* species detected by qPCR and by deep sequencing (Fig. 4B). The deep-sequencing–qPCR regression line for *Streptococcus* has a slope of 1.03 and a Pearson correlation value of 0.941. The residuals are uniformly distributed for both plots, indicating that the errors are unbiased.

Two outlying samples were detected in our analysis. These samples display differential measurements by qPCR and deep sequencing of *Pseudomonas aeruginosa* in INPT 12 and *Streptococcus* in OUTPT 11. We did not include these samples in the analysis of

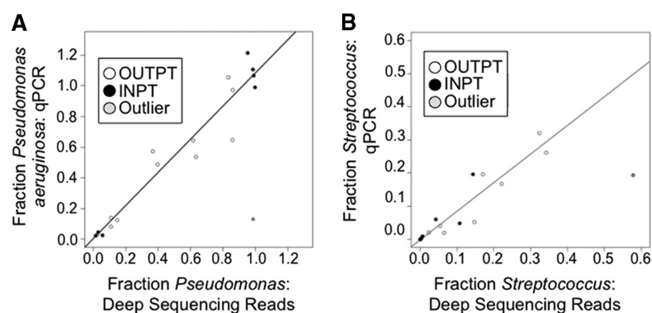

FIG 4 qPCR analysis independently verifies 454 pyrosequencing detection of the most prevalent bacteria in cystic fibrosis patient sputum samples. (A) The fraction of *Pseudomonas aeruginosa* determined by qPCR (*rplU* detection/universal detection) correlates to the fraction of deep-sequencing reads assigned to the *Pseudomonas* genus. Pearson correlation = 0.968. (B) The fraction of *Streptococcus* determined by qPCR (*tuf* detection/universal detection) correlates to the fraction of deep-sequencing reads assigned to the *Streptococcus* genus. Pearson correlation = 0.941. All samples were analyzed in replicates of six for each primer set, with the exception of OUTPT 11 (n = 3). Median fraction values for each sample are plotted for qPCR detection and were used for the correlation analysis. Pearson correlation lines were calculated, excluding the outlier points. Legend: open circle, outpatient (OUTPT); black closed circle, inpatient (INPT); gray closed circle, outlier.

regression lines. *Pseudomonas* pyrotags accounted for 98% of the deep-sequencing reads for INPT 12, while qPCR measured only 13% *Pseudomonas aeruginosa* in that same sample using the *rplU* primer set. For this sample, a second *Pseudomonas aeruginosa*-specific primer set, targeted to an alternative gene, *oprD*, also underreported *Pseudomonas aeruginosa* compared to 454 pyrosequencing (not shown). These results may indicate that a non-*aeruginosa* species of the genus *Pseudomonas* predominates in this sample.

Additionally, a discrepancy in the measurement of the fraction of the *Streptococcus* genus was seen between the two methods for OUTPT 11. In this sample, the qPCR method detected 19% *Streptococcus*, while 454 pyrosequencing detected 58% *Streptococcus.* While we do not fully understand the reason for this discrepancy, we note that the GAST algorithm used here resolved the *Streptococcus* reads for this patient sample to the species level and they were assigned to *Streptococcus pneumoniae* (not shown). Therefore, this sample was analyzed using a *Streptococcus pneumoniae*-specific primer set targeted to *psaA*. Similar to the *Streptococcus* genus analysis, qPCR measured 13% *Streptococcus pneumoniae* in OUTPT 11, an underrepresentation compared to the deep-sequencing results. Technical or biological factors related to OUTPT 11 are not germane to the central thesis of this report and will be explored at a later date.

Overall, the qPCR assay presented and validated here provides an alternate, high-throughput method for analyzing complex, polymicrobial patient samples. This assay can be used to broadly characterize the sputum microbiome, as well as verify trends determined in deep-sequencing experiments.

Oral streptococci and the SMG species are the predominant streptococci detected in the patient sputum samples. The correlation observed between the fractional representation of *Streptococcus* and clinically stable disease (i.e., outpatients) prompted us to further characterize the streptococcal species in the patient samples. A selection of 13 samples (8 outpatients and 5 inpatients for which sufficient sputum-derived gDNA was available) was analyzed on the HOMIM, which contains probes for the detection of ~300 oral species, including the majority of *Streptococcus* species previously identified in CF patient sputum samples (17). Each analyzed sputum gDNA sample is scored based on relative intensity of hybridization for each probe on the microarray and is assigned a semiquantitative value of 0 to 5, where 0 is no hybridization above background and 5 is intense hybridization to probe. Table S3 in the supplemental material displays a complete list of hybridization intensity score assignments to all probes for the 13 samples analyzed by HOMIM.

As a control, HOMIM analysis confirmed the presence of *Pseudomonas aeruginosa* in the 13 samples analyzed. Further, the overall bacterial community profiles determined by HOMIM correlated with the pyrosequencing results (see Table S1 in the supplemental material). Ahn et al. previously reported the correlation of these two methods (1).

The HOMIM results for the 13 selected samples showed that the patient samples hybridized most strongly to probes targeted to *Streptococcus salivarius*, *Streptococcus parasanguis*, and SMG species (Fig. 5). As can be seen from the heat map derived from the HOMIM data, the outpatient samples cluster separately from the inpatient cohort. The majority of strong hybridization to *Streptococcus* sp. probes was detected in the outpatient samples, with minimal hybridization occurring with the inpatient cluster. This finding is consistent with the higher fraction of *Streptococcus* detected in the outpatient group than the inpatients, as validated via 454 pyrosequencing (Fig. 3B) and qPCR (Fig. 4B).

DISCUSSION

Microbiome analysis of this cohort of CF patients showed a correlation between inpatient status and decreased sputum bacterial diversity. Inpatient status for this cohort was previously shown to correlate with increased prevalence of cystic fibrosis-related diabetes and low FEV_1, low serum iron, and high sputum iron levels, a phenotype of more severe disease (8). However, the impact of individual species or interspecies interactions for each of these phenotypes remained unexplored. Importantly, clinical observations may suggest testable hypotheses about how underlying polymicrobial community composition, diversity, and relative abundance may alter the disease process in CF.

The current dogma in the field of CF dictates that *Pseudomonas aeruginosa* is the predominant organism in the majority of CF patients. Culture-independent methods of analyzing CF patient lower respiratory samples are slowly remodeling our understanding of the complexity of these microbial communities. The molecular profiling of 35 sputum samples reported here shows that the current dogma held true in about half the patient samples (Fig. 1B). For the remaining half, each patient sample was dominated by a non-*Pseudomonas* genus or a combination of *Pseudomonas* and *Streptococcus.* Further, communities dominated by a non-*Pseudomonas* genus were identified in both outpatients and inpatients. The prevalence of these alternative genera and their potential role in acute exacerbation is an essential factor that should be considered in future studies.

While 454 pyrosequencing, Illumina sequencing, and Ion Torrent technologies have gained popularity for microbiome analysis of clinical samples, this investigation provides independent verification of deep-sequencing findings. To this end, the qPCR assays of the sputum DNA samples largely confirmed the *Pseudomonas* and *Streptococcus* profiles as determined by 454 pyrosequencing.

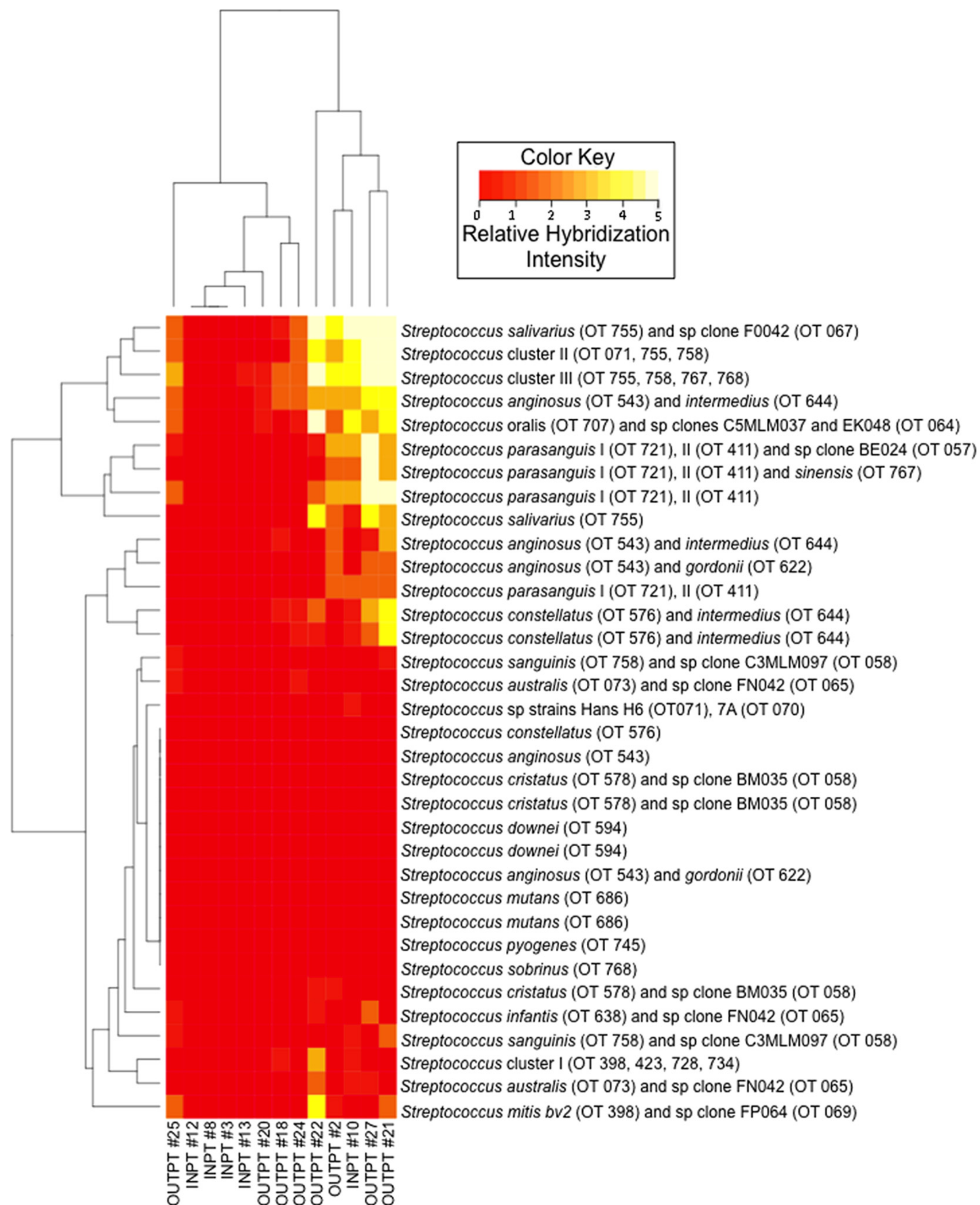

FIG 5 Oral streptococci and streptococcal species are prevalent in CF sputum samples. Shown is the assignment of relative abundance (score of 0 to 5) for each sample based on intensity of hybridization to each *Streptococcus*-specific 16S rRNA gene probe. Probes are species specific, group specific (hybridize 2 to 3 oral taxon), or cluster specific (hybridize multiple streptococcal oral taxon). Pearson hierarchical clustering was performed based on hybridization intensity in two dimensions: (i) probe and (ii) patient sample.

Measurement of the fractional representation of these two predominant genera correlated for the two methods in 18 out of 19 samples. However, a single outlier in the detection of *Pseudomonas* or *Streptococcus* occurred in two distinct samples, described in detail in Results (see Fig. 4). For the *Pseudomonas* outlier (INPT 12), our data are consistent with the conclusion that a non-*aeruginosa* strain may predominate in this patient. Although we do not understand the discrepancy in the fraction of *Streptococcus* measured by qPCR and 454 pyrosequencing for this single outlier (OUTPT 11), it may be due to PCR conditions or primer specificity or possibly represent an interesting aspect of the microbiology of this patient warranting further investigation. While each individual method largely results in consistent profiling of the dominant genera of CF patient sputum, only through the combination of the qPCR and deep-sequencing methods were we able to identify these unique outlier samples.

We further profiled the species of streptococci present in our patient samples using the HOMIM. In this analysis, species tar-

geted by probes for oral cavity-associated streptococci and the SMG species were hybridized most strongly. The predominance of *S. salivarius* and *S. parasanguis* in our outpatient samples is consistent with previous reports (17). The prevalence of SMG species in our CF patients is also consistent with previous findings by Sibley et al. (28). However, in their patient cohort, Sibley et al. showed that the SMG species were associated with acute exacerbations, while in our cross-sectional patient cohort, SMG species were most abundant in the clinically stable patients. We suggest that modest levels of *S. salivarius*, *S. parasanguis*, and SMG species may increase the diversity of the CF lung and contribute to patient health, while excessive levels, particularly of SMG species, may lead to increased pathogenicity and clinical decline.

The samples analyzed here are complex due to a multitude of clinical treatment, host, and microbial factors that should be considered when interpreting the results of this sample set. Intravenous antibiotics may have initial impacts on the microbial community of exacerbating patients, including altering overall diversity, as early as within the first 24 h of hospital admittance, even though no consistent clinical improvement was observed this early during admittance. Of note, inpatients were treated with combinatorial antibiotic therapy that does not target any single microbial population. Additionally, gDNA of killed bacterial cells during early admittance may persist in the sputum. However, we are unaware of any studies specifying the persistence and rates of decay of gDNA in patient sputum for organisms prevalent in CF.

Furthermore, due to the passage of sputum through the oral cavity prior to sample collection, the effect of potential salivary contamination should be considered. There is strong evidence that the *Streptococcus* sp. and other oral cavity-associated species in sputum samples derive from the lung rather than from contamination with saliva. Rogers et al. showed that oral mouthwash samples compared to sputum samples from a single patient contain distinct polymicrobial communities (25). Additionally, Harris et al. analyzed bronchoalveolar lavage (BAL) fluid from children with cystic fibrosis and detected various oral cavity-associated organisms, including *Streptococcus* sp., *Prevotella* sp., *Fusobacterium* sp., and others. Collection of BAL fluid samples bypasses the oral cavity, verifying prevalence of these organisms in patient lungs (12). Further, our work presented here shows that the 454 pyrosequencing, qPCR, and HOMIM all display the same significant difference in the relative abundance of *Streptococcus* in outpatient samples compared to that in inpatient samples. Given that both inpatients and outpatients provided sputum samples by spontaneous expectoration, any potential contamination of samples with saliva would equally affect both samples sets. While variability in the volume of expectorated sputum samples existed between patients, we did not detect any correlation between the volume of sputum expectoration and fractional representation of *Streptococcus* by deep sequencing (not shown). Therefore, we conclude that the *Streptococcus* in outpatient sputum samples specifically reflects their presence in the CF lung.

It is formally possible that less prevalent organisms, such as *Gemella* or *Haemophilus*, drive microbial community dynamics in CF patient lungs. Here, we focus on *Streptococcus* due to its significant correlation with outpatient clinical status, its high abundance, and its high prevalence in this patient cohort. Overall, the increased fractional representation of numerous low-abundance genera in outpatient samples may increase the diversity of the corresponding lung communities, perhaps promoting clinical stability.

Further analyses of the community structure and interspecies interactions in the lungs of CF patients may reveal markers for the development of acute exacerbation as well as build our understanding of healthy microbial communities that promote patient stability. Furthermore, the microbial community associations detected from the sputum samples of this patient cohort will be used to guide our future *in vitro* studies exploring the underlying mechanisms and microbial interactions of the CF lung environment, which will facilitate personalized patient treatment and the development of novel therapeutics.

ACKNOWLEDGMENTS

We thank Alexis Kokaras of The Forsyth Institute for performing HOMIM assays and the Translational Research Core (TRC) for collecting and processing clinical samples.

The TRC is supported by grants from the National Center for Research Resources (5P20RR018787) and the National Institute of General Medical Sciences (8 P20 GM103413) from the National Institutes of Health to B.A.S. This work was also supported by a pilot grant from the Cystic Fibrosis Foundation Research Development Program (STANTO011RO), a grant from the Hitchcock Foundation to G.A.O., grants from the National Institutes of Health to M.L.S. (4UH3DK083993-02) and to B.A.S. (R01-HL074175-09), and a grant from the National Institute of Dental and Craniofacial Research (DE11443) to B.J.P. The project described was supported by award number T32GM008704 from the National Institute of General Medical Sciences.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

REFERENCES

1. **Ahn J, et al.** 2011. Oral microbiome profiles: 16S rRNA pyrosequencing and microarray assay comparison. PLoS One **6**:e22788. doi:10.1371/journal.pone.0022788.
2. **Bals R, Weiner DJ, Wilson JM.** 1999. The innate immune system in cystic fibrosis lung disease. J. Clin. Invest. **103**:303–307.
3. **Bittar F, Rolain JM.** 2010. Detection and accurate identification of new or emerging bacteria in cystic fibrosis patients. Clin. Microbiol. Infect. **16**: 809–820.
4. **Cox MJ, et al.** 2010. Airway microbiota and pathogen abundance in age-stratified cystic fibrosis patients. PLoS One **5**:e11044. doi:10.1371/journal.pone.0011044.
5. **Dodge JA, Lewis PA, Stanton M, Wilsher J.** 2007. Cystic fibrosis mortality and survival in the UK: 1947-2003. Eur. Respir. J. **29**:522–526.
6. **Edgar RC.** 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics **26**:2460–2461.
7. **Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R.** 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics **27**: 2194–2200.
8. **Gifford AH, et al.** 2011. Iron and CF-related anemia: expanding clinical and biochemical relationships. Pediatr. Pulmonol. **46**:160–165.
9. **Gilligan PH.** 1991. Microbiology of airway disease in patients with cystic fibrosis. Clin. Microbiol. Rev. **4**:35–51.
10. **Gomez MI, Prince A.** 2007. Opportunistic infections in lung disease: *Pseudomonas* infections in cystic fibrosis. Curr. Opin. Pharmacol. **7**:244–251.
11. **Guss AM, et al.** 2011. Phylogenetic and metabolic diversity of bacteria associated with cystic fibrosis. ISME J. **5**:20–29.
12. **Harris JK, et al.** 2007. Molecular identification of bacteria in bronchoalveolar lavage fluid from children with cystic fibrosis. Proc. Natl. Acad. Sci. U. S. A. **104**:20529–20533.
13. **Horz HP, Vianna ME, Gomes BP, Conrads G.** 2005. Evaluation of universal probes and primer sets for assessing total bacterial load in clinical samples: general implications and practical use in endodontic antimicrobial therapy. J. Clin. Microbiol. **43**:5332–5337.

14. **Huse SM, et al.** 2008. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. PLoS Genet. **4**:e1000255. doi: 10.1371/journal.pgen.1000255.
15. **Lyczak JB, Cannon CL, Pier GB.** 2002. Lung infections associated with cystic fibrosis. Clin. Microbiol. Rev. **15**:194–222.
16. **Maeda H, et al.** 2003. Quantitative real-time PCR using TaqMan and SYBR Green for *Actinobacillus actinomycetemcomitans*, *Porphyromonas gingivalis*, *Prevotella intermedia*, *tetQ* gene and total bacteria. FEMS Immunol. Med. Microbiol. **39**:81–86.
17. **Maeda Y, et al.** 2011. Population structure and characterization of viridans group streptococci (VGS) including *Streptococcus pneumoniae* isolated from adult patients with cystic fibrosis (CF). J. Cyst. Fibros. **10**:133–139.
18. **Parkins MD, Sibley CD, Surette MG, Rabin HR.** 2008. The *Streptococcus milleri* group—an unrecognized cause of disease in cystic fibrosis: a case series and literature review. Pediatr. Pulmonol. **43**:490–497.
19. **Paster BJ, Dewhirst FE.** 2009. Molecular microbial diagnosis. Periodontol. 2000 **51**:38–44.
20. **Paster BJ, Olsen I, Aas JA, Dewhirst FE.** 2006. The breadth of bacterial diversity in the human periodontal pocket and other oral sites. Periodontol. 2000 **42**:80–87.
21. **Picard FJ, et al.** 2004. Use of *tuf* sequences for genus-specific PCR detection and phylogenetic analysis of 28 streptococcal species. J. Clin. Microbiol. **42**:3686–3695.
22. **Rogers GB, et al.** 2004. Characterization of bacterial community diversity in cystic fibrosis lung infections by use of 16S ribosomal DNA terminal restriction fragment length polymorphism profiling. J. Clin. Microbiol. **42**:5176–5183.
23. **Rogers GB, et al.** 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.
24. **Rogers GB, et al.** 2005. Bacterial activity in cystic fibrosis lung infections. Respir. Res. **6**:49.
25. **Rogers GB, 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.
26. **Sibley CD, Rabin H, Surette MG.** 2006. Cystic fibrosis: a polymicrobial infectious disease. Future Microbiol. **1**:53–61.
27. **Sibley CD, Parkins MD, Rabin HR, Surette MG.** 2009. The relevance of the polymicrobial nature of airway infection in the acute and chronic management of patients with cystic fibrosis. Curr. Opin. Investig. Drugs **10**:787–794.
28. **Sibley CD, et al.** 2008. A polymicrobial perspective of pulmonary infections exposes an enigmatic pathogen in cystic fibrosis patients. Proc. Natl. Acad. Sci. U. S. A. **105**:15070–15075.
29. **Sogin ML, et al.** 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere." Proc. Natl. Acad. Sci. U. S. A. **103**:12115–12120.
30. **Stressmann FA, et al.** 2011. Does bacterial density in cystic fibrosis sputum increase prior to pulmonary exacerbation? J. Cyst. Fibros. **10**:357–365.
31. **Tunney MM, 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.
32. **Zhao J, et al.** 26 March 2012. Decade-long bacterial community dynamics in cystic fibrosis airways. Proc. Natl. Acad. Sci. U. S. A. [Epub ahead of print.] doi:10.1073/pnas.1120577109.