Decade-long bacterial community dynamics in cystic fibrosis airways

Jiangchao Zhao[a], Patrick D. Schloss[b], Linda M. Kalikin[a], Lisa A. Carmody[a], Bridget K. Foster[a], Joseph F. Petrosino[c,d], James D. Cavalcoli[e], Donald R. VanDevanter[f], Susan Murray[g], Jun Z. Li[h], Vincent B. Young[b,i], and John J. LiPuma[a,j,1]

Departments of [a]Pediatrics and Communicable Diseases, [b]Microbiology and Immunology, [h]Human Genetics, and [i]Internal Medicine/Infectious Diseases Division and [e]Center for Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI 48109; [c]Alkek Center for Metagenomics and Microbiome Research and [d]Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, TX 77030; [f]Department of Pediatrics, Case Western Reserve University, Cleveland, OH 44106; and Departments of [g]Biostatistics and [j]Epidemiology, University of Michigan School of Public Health, Ann Arbor, MI, 48109

The structure and dynamics of bacterial communities in the airways of persons with cystic fibrosis (CF) remain largely unknown. We characterized the bacterial communities in 126 sputum samples representing serial collections spanning 8–9 y from six age-matched male CF patients. Sputum DNA was analyzed by bar-coded pyrosequencing of the V3–V5 hypervariable region of the 16S rRNA gene, defining 662 operational taxonomic units (OTUs) from >633,000 sequences. Bacterial community diversity decreased significantly over time in patients with typically progressive lung disease but remained relatively stable in patients with a mild lung disease phenotype. Antibiotic use, rather than patient age or lung function, was the primary driver of decreasing diversity. Interpatient variability in community structure exceeded intrapatient variability in serial samples. Antibiotic treatment was associated with pronounced shifts in community structure, but communities showed both short- and long-term resilience after antibiotic perturbation. There was a positive correlation between OTU occurrence and relative abundance, with a small number of persistent OTUs accounting for the greatest abundance. Significant changes in community structure, diversity, or total bacterial density at the time of pulmonary exacerbation were not observed. Despite decreasing community diversity in patients with progressive disease, total bacterial density remained relatively stable over time. These findings show the critical relationship between airway bacterial community structure, disease stage, and clinical state at the time of sample collection. These features are the key parameters with which to assess the complex ecology of the CF airway.

microbial ecology | microbiome | antibiotics

Chronic infection and inflammation of the airways are the leading causes of morbidity and mortality in cystic fibrosis (CF). Therefore it is reasonable to expect that a better understanding of airway infection in CF could provide opportunities for novel strategies to maintain lung health in this population. Recently, the microbiota of the CF airways has been assessed through the application of culture-independent methods that leverage the phylogenetic information of the bacterial 16S rRNA gene. By using a variety of techniques, including Sanger sequencing of clone libraries (1, 2), terminal restriction fragment length polymorphism analysis (3, 4), microarray hybridization (5, 6), or pyrosequencing (7), these studies have revealed the presence of a far more complex bacterial community in the airways of CF patients than previously appreciated with culture-based methods.

Most of these studies have been cross-sectional, analyzing samples collected at a single time point from individual patients. Correlations between microbial communities and disease progression in cross-sectional studies can be confounded by a host of variables, including patient age, sex, lung disease stage, and antibiotic use. Longitudinal studies that analyze serial samples obtained from individual patients over long periods of time allow better assessment of the impact of these potentially confounding variables in constructing tractable models of the relationship

between the dynamics of the lung microbial community and disease progression (4, 8).

In this study, we used a culture-independent approach to investigate intensively the dynamics of the CF airway bacterial communities over the course of 8–9 y in six age- and sex-matched CF patients. The findings reveal a dynamic ecological system that is influenced markedly by a patient's rate of disease progression as well as by disease stage and clinical state at the time of sampling. These features have significant bearing on measures of diversity and community structure, reinforcing the power of longitudinal studies in relating the status of the airway microbial community to CF disease activity.

Results

Patients and Lung Microbiota Measures. Multiple sputum samples were collected from six adult male patients with CF during the course of 8–9 y. Percent predicted forced expiratory volume in 1 s (%FEV$_1$) was used as a measure of lung function. Three patients (S1, S2, and S3) showed a mild phenotype of lung disease, maintaining relatively stable lung function over several years. The other three patients (P1, P2, and P3) had a more typical course of CF lung disease, exhibiting a progressive decline in lung function during the period of observation. Hereafter these two groups of patients are referred to as "stable" and "progressing" patients, respectively (Table 1 and Fig. 1A).

The bacterial communities in 126 sputum samples obtained from these six patients were profiled using pyrosequencing of the 16S rRNA V3–V5 hypervariable region. A total of 633,746 high-quality reads were generated from these samples with an average of 5,030 reads per sample. From these sequences, 662 operational taxonomic units (OTUs) (based on 97% DNA sequence similarity) were identified. On average, 36 OTUs per sample were detected; however, the range was large (5–114 OTUs per sample). After normalization to the sample with the smallest number of reads (1,152), 259 OTUs were included in further analyses. Among these, 47 OTUs each had a relative abundance of more than 1% in at least one sputum sample (Table S1).

Community Diversity Relative to Disease Progression and Clinical State. The diversity of the bacterial community, based on calculation of

Table 1. Study subjects

	Stable patients			Progressing patients		
	S1	S2	S3	P1	P2	P3
CFTR genotype	Unknown*	ΔF508/1717–1G > A	ΔF508/ΔF508	ΔF508/ΔF508	ΔF508/ΔF508	ΔF508/ΔF508
Age range	18.9–28.0	21.2–30.1	19.8–28.1	21.1–29.2	18.7–27.3	22.0–30.1
Observation period (y)[†]	9.1	8.9	8.3	8.1	8.6	8.1
%FEV$_1$ range	81–83	83–88	91–82	68–30	65–41	77–25
Avg. %FEV$_1$ change[‡]	0.11	0.56	−1.11	−4.69	−2.79	−6.42
No. specimens	23	14	13	20	30	26

*No mutations identified from the 70 mutation sequencing panel, which includes ΔF508. CF diagnosis based on positive sweat chloride test.

[†]Years between collection dates of first and last sputum samples.

[‡]Calculated as the difference between %FEV$_1$ values when first and last sputum samples were collected divided by the observation period.

the Shannon index, fluctuated in serial samples from all patients but remained relatively constrained in the three stable patients (S1, S2 and S3). In contrast, serial samples from the three progressing patients (P1, P2 and P3) exhibited statistically significant decreases in community diversity during the course of several years (Fig. 1B).

CF is characterized by periods of relatively stable pulmonary function punctuated by episodic exacerbations of respiratory symptoms and decreased lung function which usually warrant antibiotic treatment. To examine the impact of these changes in *clinical state* on bacterial diversity, each of the 126 samples was assigned to one of four "BETR" categories: baseline (B); pulmonary exacerbation before initiation of episodic antibiotic therapy (E); antibiotic treatment of pulmonary exacerbation (T); and recovering from pulmonary exacerbation (R) (Table 2). Eight samples with insufficient medical record information were labeled as "unknown" (U). Not unexpectedly, a comparison of the diversity of bacterial communities revealed that treatment samples, collected when patients were receiving antibiotics for exacerbation, were significantly less diverse than baseline, exacerbation, or recovering samples (Fig. 1C). No significant differences in diversity levels were observed between baseline and exacerbation samples, whereas diversity levels from recovering samples were significantly lower than those from exacerbation samples.

Impacts of Age, Antibiotic Therapy, and Lung Function on Diversity. Having demonstrated that clinical state at the time of sample collection influences measures of community diversity, we next sought to assess the relative impacts of patient age, lung function, and antibiotic therapy on diversity. Antibiotic load was quantified as the sum of the number of days a patient received antibiotics (prescribed for maintenance therapy as well as for acute treatment of pulmonary exacerbations) during the 30 d before specimen collection (see *SI Materials and Methods* for details). In the three stable patients, the diversity of the community did not decrease significantly with increasing age and mildly decreasing lung function (Fig. 1B); similarly, no significant decrease in diversity was observed after adjustment for antibiotic load. In contrast, in the three progressing patients, where increasing age and decreasing lung function were each associated with decreasing diversity, adjustment for antibiotic load eliminated or diminished the statistical significance of these trends (Fig. S1).

Samples from all six patients were combined to explore further the effects of age, lung function, and antibiotic load on diversity in both univariate and multivariate linear mixed models (Table S2). In the univariate models, all three variables were correlated individually with decreasing diversity. In the multivariate model, however, antibiotic use was the strongest predictor of decrease in diversity, overshadowing the contributions of age and lung function to contribute to a significant trend ($P < 0.001$ for antibiotic

load; $P > 0.10$ for age and lung function; see *SI Results* for details). Collectively, these results indicated that antibiotic use, rather than patient age or lung function, was the primary driver of the decreasing bacterial diversity observed in patients with typically progressive CF lung disease.

Community Structure Dynamics. We compared community structures across patients to determine if interindividual differences in structure exceeded the changes we observed between serial samples from the same patient. Pairwise ecologic distances were calculated for all samples using the Bray–Curtis (BC) distance metric, which takes into account both community membership and relative abundance. These distances then were visualized by principal coordinate analysis (PCoA), which showed considerable overlap between communities from the two sets of stable and progressing patients (Fig. 2A). However, in comparing patient pairs, PCoA revealed appreciably distinct communities in individual patients (Fig. S2). Likewise, analysis of molecular variance (AMOVA) (9) showed significant differences in communities between all patient pairs, except between patients P2 and P3.

BC-based PCoA plots of communities from individual patients also showed the impact of clinical state on the dynamics of community structure (Fig. 2B). The majority of sputum samples from the three stable patients were collected during baseline states, and communities clustered in limited areas of PCoA plots. In contrast, the three progressing patients experienced more exacerbations and more courses of antibiotic treatment. Baseline and exacerbation samples from these patients clustered in the left half of the PCoA plots, whereas samples collected during antibiotic treatment drifted into other areas, particularly in patients P1 and P3. PCoA using the Jaccard distance metric, which takes into account only community membership, revealed similar results (Fig. S3).

The significance of the relationships between the structure of the bacterial community (measured as BC distance), patient age, %FEV$_1$, and antibiotic load was assessed using the Mantel test. The positive correlation between BC distance and antibiotic load was greater (Mantel $r = 0.32$, $P < 0.01$) than that between BC distance and age (Mantel $r = 0.20$, $P < 0.01$) or between BC distance and %FEV$_1$ (Mantel $r = 0.22$, $P < 0.01$). When antibiotic load was used as a control in a partial Mantel test, lesser correlations between BC distance and age (Mantel $r = 0.11$, $P = 0.02$) and between BC distance and %FEV$_1$ (Mantel $r = 0.12$, $P < 0.01$) were observed.

Communities in samples obtained after recovery from treatment of exacerbation were closer to the baseline and exacerbation (preantibiotic) communities in the ordination than were samples obtained during antibiotic treatment, suggesting resiliency of communities after antibiotic perturbation. For example, sample 10 from patient P1 (Fig. 2B) was obtained at baseline, and sample
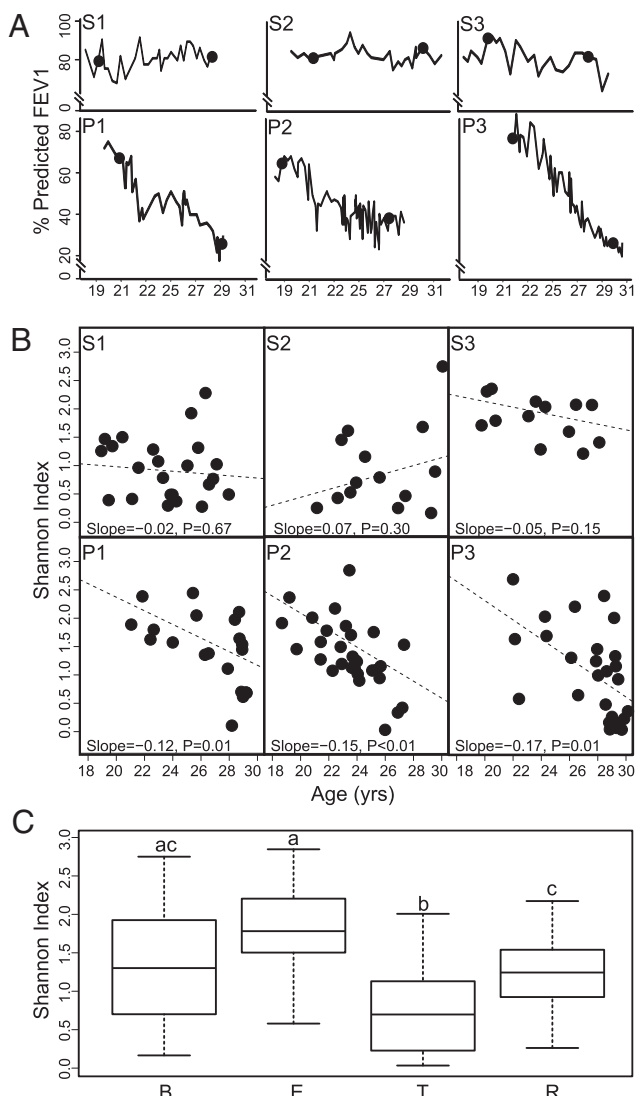
Fig. 1. Changes in lung function and community diversity. (*A*) Percent predicted FEV_1 values (*y* axis) are plotted against patient age (*x* axis). Patients S1, S2, and S3 have relatively stable disease; patients P1, P2, and P3 have progressive disease. Solid circles indicate the ages when the first and last sputum samples were obtained. (*B*) Temporal changes in community diversity in stable (*Upper Row*) and progressing (*Lower Row*) patients. In each plot, community diversity measured by Shannon index is plotted against patient age when sputum was collected. (*C*) Boxplot comparison of microbial diversity in sputum samples obtained during different BETR clinical states. The top and bottom boundaries of each box indicate 75th and 25th quartile values, respectively, and black lines inside each box represent 50th quartile (median) values. Ends of the whiskers mark the lowest and highest diversity index in each BETR category. The letters above the boxes indicate comparison results in group means; groups with the same letter do not have significant differences in means, but those indicated by different letters are significantly different. (See *SI Results* for details.)

11 was obtained 15 wk later during treatment for an exacerbation. Sample 12 was collected after return to baseline 11 wk thereafter and resided close to sample 10 on the PCoA plot. Resiliency over the course of several years also was suggested by PCoA. In patient P2, for example, the community in the last collected sample, obtained during recovery from treatment for exacerbation, clustered close to the first sample, collected during baseline 8.5 y earlier. Similarly, the last samples obtained during periods without antibiotic treatment in patient P1 (sample 13, collected 7.5 y after the first sample) and in patient P3 (sample 12, collected 6.5 y

after the first sample) also resided close to their respective first samples (Fig. 2*B*).

OTU Occurrence and Relative Abundance. The occurrence and relative abundance of OTUs were dynamic in serial samples from both stable and progressing patients (Fig. 3, Fig. S4, and Table S1), but differences between these groups were noted. The most pronounced difference was with *Staphylococcus,* which occurred in 67% of samples from progressing patients but in only 6% of samples from stable patients.

Within each patient's set of serial samples, certain OTUs were characterized as persistent, being defined as detected with a relative abundance of >1% in at least 50% of the samples obtained from that individual throughout the 8–9 y of observation. *Pseudomonas* was detected in at least 92% of samples in each patient (Table S1). Species in the *S. milleri* group were found in at least 58% of samples from each patient, and species in the *S. mitis* group were detected in at least 65% of samples from all patients except patient P3. A positive relationship was found between OTU occurrence and relative abundance. Although only 10 (3.9%) of the 259 total OTUs detected in the 126 sputum samples were defined as persistent (i.e., were found in more than 50% of samples in at least one of the six patients), these taxa accounted for 83–90% of the total relative abundance in each patient (Table S3).

Clinical state, particularly treatment of exacerbation, affected the occurrence and relative abundance of OTUs. In patient P3, for example, *S. mitis* group species were detected in 46% of all samples; however, they were detected in only 15% of samples obtained during treatment for exacerbation and in 86% of samples obtained at exacerbation before initiation of antibiotic therapy. In all three progressing patients, more OTUs were characterized as persistent when only samples obtained at baseline or exacerbation were considered (Table S3). In other words, in the absence of antibiotic treatment or recovery from treatment, more OTUs (primarily anaerobic species) met the criteria for being considered persistent in these patients.

The effects of repeated courses of antibiotic therapy for exacerbation were especially evident in later stages of lung disease, in which communities were heavily dominated by one or two OTUs (Fig. 3 and Fig. S4). In patient P1, communities late in the disease course were heavily dominated by *Achromobacter* and *Pseudomonas,* with these two OTUs accounting for up to 99.9%

Table 2. BETR categories

Clinical state	Description
Baseline (B)	Well or mild increase in pulmonary symptoms
	Not a doctor defined pulmonary exacerbation
	Not hospitalized for increase in pulmonary symptoms
	Not on episodic antibiotics for >30 d
	May or may not be on maintenance antibiotics
Exacerbation (E)	Doctor defined pulmonary exacerbation or increased pulmonary symptoms fitting definition of exacerbation*
	Before start of episodic IV or oral antibiotics
	Not on episodic antibiotics for >30 d
	May or may not be on maintenance antibiotics
Treatment (T)	On IV or oral episodic antibiotics for treatment of doctor defined exacerbation or increased pulmonary symptoms fitting definition of exacerbation*
Recovering (R)	Off episodic antibiotics ≤30 d
	May or may not be on maintenance antibiotics
	May or may not be back to baseline clinical state

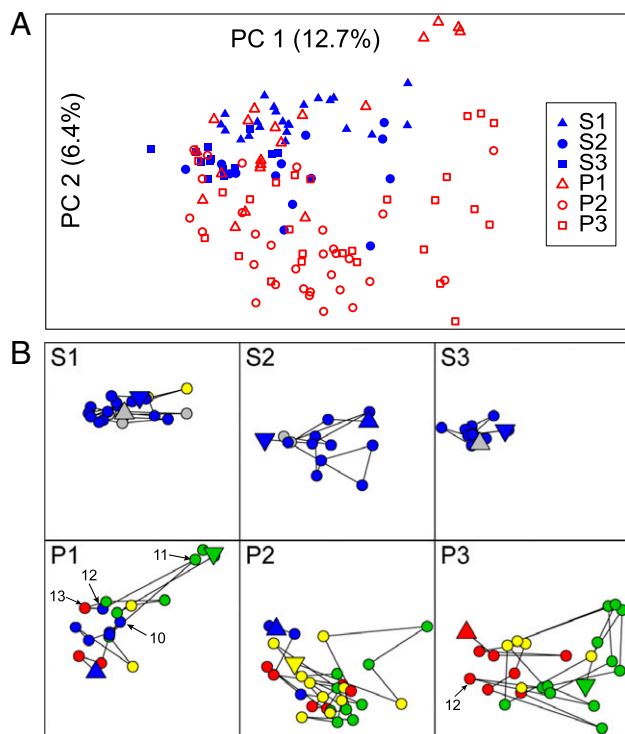*See *SI Materials and Methods* for details.

Fig. 2. PCoA of bacterial community structures using BC distances. (*A*) Community structures of samples from three stable patients (blue) and three progressing patients (red). Distances between symbols on the ordination plot reflect relative similarities in community structures. PC1, principal coordinate 1; PC2, principal coordinate 2. (*B*) Community structure dynamics in serial samples highlighting the patient's clinical state on day of collection. The scale and percent of variation explained by PC1 and PC2 for each plot are the same as shown in *A*. Upper row shows stable patients (S1, S2, S3); lower row shows progressing patients (P1, P2, P3). Circles represent individual sputum samples and are color coded to reflect BETR category (B, blue; E, red; T, green; R, yellow; U, gray); lines connect serial samples. Upright and inverted triangles indicate first and last samples, respectively, in each serial set. Samples 10–13 from patient P1 and sample 12 from patient P3 are indicated by black arrows and are described in the text.

Fig. 3. Relative abundance of OTUs in serial sputum samples from patient P2. Relative abundances of OTUs, each accounting for >1% of the total bacterial community, are shown in each sample. Relative abundances of all remaining OTUs, each representing ≤1% of the community, are shown in gray at the top of each bar. Top line indicates patient age at time of sample collection. Letters beneath each bar indicate BETR category (Table 2). Heat map indicates antibiotic load (*Materials and Methods*) at the time of sputum sample collection. Circles indicate total bacterial density (16S rRNA copies/mL sputum) based on quantitative PCR. Similar graphs for patients S1, S2, S3, P1, and P3 are shown in Fig. S4. *Pseudomonas*-1 represents *Pseudomonas aeruginosa*. *Streptococcus*-1 represents species within the *Streptococcus mitis* group. *Streptococcus*-2 represents species within the *Streptococcus milleri* group.

of the relative abundance in some samples. In patients P2 and P3, communities in later samples were dominated by *Pseudomonas*, nearly to the exclusion of other OTUs.

Quantitation of Bacterial Load. Bacterial densities in samples were calculated by quantitative PCR analysis using a universal 16s rRNA primer-probe assay. No significant differences in total bacterial densities were detected in sputum samples from individual patients or from the three stable patients compared with the three progressing patients (Fig. S5*A*; *P* > 0.05, linear mixed model). Similarly, bacterial density did not differ significantly between samples based on clinical BETR state (Fig. S5*B*; *P* > 0.05, linear mixed model).

Discussion

Previous cross-sectional studies have reported that bacterial community diversity in CF is correlated positively with pulmonary function and inversely with patient age, suggesting a role for community composition in the pathophysiology of CF airway disease (5, 6). In our longitudinal study, we, too, found a positive correlation between diversity and lung function; however, at the individual patient level, decreasing community diversity was not necessarily correlated with increasing age. In the three patients with a mild lung disease phenotype, community diversity remained quite stable throughout time spans of 8–9 y. Furthermore,
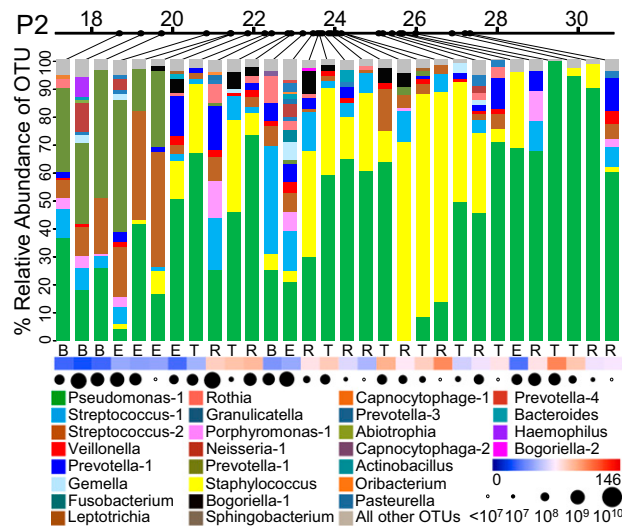
in patients S1 and S2, community diversity levels were comparable to those observed late, during periods of advanced lung disease, in the three progressing patients (Fig. 1*B*). Thus, community diversity alone is not a sufficient indicator of disease status. Of particular note is patient S3, who was homozygous for the ΔF508 CF transmembrane conductance regulator allele, a genotype more commonly associated with a more aggressive lung disease phenotype. Despite having relatively mild lung disease, this patient had airway bacterial communities that were significantly more diverse than those in the other two stable patients and comparable to those observed in early samples from the three progressing patients (Fig. 1*B* and Fig. S6). This apparent correlation between community diversity and CF genotype raises intriguing questions about the role the ΔF508 allele plays in susceptibility to airways infection.

Schluchter and colleagues (10) characterized CF lung disease phenotype by using longitudinal pulmonary function data, demonstrating the advantages of this approach over the use of a single FEV₁ measurement to describe the severity of a patient's lung disease at a particular point in time. More recently, Konstan and colleagues (11) used these same principles to clarify the distinction between a patient's overall CF lung disease severity phenotype (lung *disease aggressiveness*) and that patient's current state of lung disease (lung *disease stage*). The rate at which %FEV₁ declines over a period is an indicator of the aggressiveness of a patient's lung disease, independent of the stage of the patient's lung disease, which is defined as "early" when serial %FEV₁ measures are >70; "intermediate" when %FEV₁ measures range between 70 and 40; and "advanced" when %FEV₁ measures are <40. In this context, the three stable patients included in our study can be described as having a mild lung disease-aggressiveness phenotype, and the three progressing patients represent a more typical moderate lung disease-aggressiveness phenotype (Fig. S7).

Our study demonstrates that among patients with moderately aggressive lung disease, disease stage (early, intermediate, or advanced), as indicated by %FEV$_1$ measures, has significant bearing on estimates of the diversity of the lung bacterial community. As the three (stable) patients with mildly aggressive lung disease ultimately progress from early to intermediate stages of disease, and as antibiotic therapy intensifies, a similar decrease in diversity likely would be observed; however, continued long-term follow-up of these patients will be necessary to demonstrate this decrease. More importantly, our results show that clinical state (BETR category) at the time of sample collection has a pronounced effect on measures of diversity. It remains to be determined how finer-scale demarcation of clinical state (e.g., subcategories taking into account the exact number of days on or off antibiotic therapy relative to sample collection) might further impact measures of community diversity.

The pathophysiologic bases for the intermittent exacerbations that characterize CF are poorly understood. Speculation has focused on presumed acute changes in the composition of the airways bacterial community around the time of exacerbation. Although we detected significantly lower community diversity in samples obtained during antibiotic treatment, we observed no differences in diversity between baseline samples obtained during periods of relatively stable health and samples obtained at the time of exacerbation but before initiation of antibiotic therapy. This lack of a clear microbial signature for the onset of exacerbation suggests that exacerbations (*i*) may not be associated with specific changes in airway bacterial communities; (*ii*) may be associated with changes in only a few OTUs and/or very minor changes in several OTUs; or (*iii*) may result primarily from host-specific factors such as the nature and degree of inflammatory response to some as yet undefined stimulus. We also found no significant changes in total bacterial density by quantitative PCR in samples based on clinical state, suggesting that the reduction in some OTUs with antibiotic treatment may be offset by an increase in other, possibly more resistant, OTUs. It is important to note, however, that our quantitative PCR analysis did not differentiate between live and dead bacterial cells (12). Thus, it is likely that measures of bacterial density, particularly in samples obtained during antibiotic therapy, represent both viable and dead bacteria. Whether this distinction is important with respect to host inflammatory response and disease progression remains unclear.

It is clear from our results and those of others (5, 6) that, as most persons with CF age, lung function typically declines, antibiotic therapy both accumulates and accelerates, and the diversity of the airway community decreases. However, these confounding variables present a challenge to establishing a simple linear cause-and-effect relationship between the progression of lung disease and decreasing community diversity. Although age, lung function (%FEV$_1$), and antibiotic use were all significantly correlated with community diversity in our study, we found that the effects of age and lung function lost significance when controlled for antibiotic use, suggesting that antibiotic therapy is the stronger driver of decreasing diversity. More robust models of antibiotic use in CF that incorporate antibiotic class, spectrum of activity, route of delivery, synergistic or antagonistic interactions, and duration of therapy are needed to provide a more precise understanding of the impact of antibiotic use on microbial community diversity.

Although we found considerable overlap in the community structures from all six patients, pairwise analyses using PCoA and AMOVA showed significant community differences between patient pairs (with the exception of patients P2 and P3), suggesting that interpatient variability in community structures exceeds intrapatient variability in longitudinal sample sets. When pairwise analyses were performed on subsets of samples based on disease stage (Fig. S8 and Table S4), we found that communities from the three progressing patients become more or less distinct in the advanced stage of disease, depending on the dominant

species. More specifically, the advanced-stage communities in samples from patient P1, which were dominated by *Achromobacter*, were quite distinct from those of patients P2 and P3, which were dominated by *Pseudomonas*. This result suggests that, as community diversity decreases in the advanced stage of disease, community structures adopt a smaller number of distinct profiles, mainly described by the species that dominates the community.

Our analyses of serial samples from individual patients reinforced the impact of antibiotic therapy on community structure dynamics. BC-based PCoA plots showed marked movement of communities in samples obtained during antibiotic therapy for exacerbation relative to other samples, with the correlation between BC distance and antibiotic use exceeding that of either BC distance and age or BC distance and %FEV$_1$. Our results also suggest that CF airways communities are resilient after perturbation by antibiotics. PCoA plots of communities in serial samples from the three progressing patients generally show communities in treatment recovery samples returning to profiles resembling the communities in samples obtained before therapy, indicating that resilience, rather than stochastic reassembly of depleted communities with new species, drives community dynamics in CF. A more complete assessment of the resilience of CF airway communities will require analysis of a larger set of samples from around the time of exacerbations. An important question to address in this regard is whether the dynamic range of communities changes as diversity decreases with disease progression. It may be reasonable to expect that, as airway communities become increasingly constrained by repeated antibiotic use, they may exhibit greater resistance to further perturbation, particularly if the community becomes restricted to a small number of highly antibiotic-resistant species.

Our study reinforces the dominant role of *Pseudomonas* as a constituent of the CF airway microbiota. *Pseudomonas* was identified in all six patients and quite often represented the most abundant species in individual samples. Of particular note, the communities in patients S1 and S2 were dominated almost uniformly by *Pseudomonas* during 8–9 y of clinical stability, indicating that the presence of *Pseudomonas*, even in relatively high abundance, does not necessarily predict poor clinical outcome. Additional analyses that allow quantitative assessment of potential fluctuations in the absolute density of *P. aeruginosa* are required to understand better the ecology and clinical impact of this species in CF. Our study also reaffirms that several anaerobic species, including a number of *Veillonella* and *Prevotella* species, are common and relatively abundant members of the airway microbiota in CF (13). The facultative anaerobic species of the *S. milleri* group (*S. anginosus*, *S. constellatus*, and *S. intermedius*) deserve particular attention. Using both culture and culture-independent methods, Sibley and colleagues (4) found that these species are detected frequently in CF respiratory samples and may be associated with clinical decline in health. Our results support these observations: *S. milleri* species were both persistent (present in most samples) and relatively abundant in all patients included in our study.

Our results support the recent observations by van der Gast and colleagues (2) that the CF airway bacterial metacommunity can be partitioned into two groups. One group is comprised of taxa that are prevalent and typically abundant; the other group consists of less commonly found species occurring in low abundance. We, too, found a positive correlation between OTU occurrence and relative abundance; communities in our six patients were dominated most often by *Pseudomonas, S. milleri* species, and strictly anaerobic species. Again, however, our results highlight the importance of taking clinical state into account in characterizing the airway bacterial community in this regard. We observed that a greater number of OTUs met our definition of persistent, which considers both rate of occurrence and relative abundance, when samples were limited to those obtained in the absence of antibiotic therapy.

Given these observations, it is important to note that the method of DNA extraction may have a significant impact on measures of the relative abundance of specific taxa. For example, we have shown recently that a method of DNA extraction that does not include the use of lysozyme and lysostaphin significantly underrepresents the abundance of *Staphylococcus* in CF sputum samples (14). Thus, firm conclusions regarding the relative abundance of species in individual specimens must be made with caution. In comparing communities across multiple specimens, however, a given method of DNA extraction should impart equal bias across all samples; therefore, observed intersample differences would not be expected to be the result of the methodology used.

Finally, by using quantitative PCR we have shown, similar to the recent report by Stressmann and colleagues (15), that the total bacterial density in the airway does not increase significantly at the time of pulmonary exacerbation. Our study extends these observations to show further that, as bacterial community diversity decreases with advancing age, progression of lung disease, and increasing antibiotic use, overall bacterial density remains relatively stable over prolonged periods of time. Our data suggest that, as community diversity decreases, the density of the remaining species increases to maintain a stable airway bacterial load. This result is most evident in the terminal treatment samples from patients P1 and P3, wherein bacterial density remained high (10^8–10^9 copies/mL) in the context of markedly species-restricted airway communities and intensive antibiotic therapy.

In summary, this study describes the structure and dynamics of airway bacterial communities in persons with CF over periods of several years, relating these measures to changes in patients' clinical status and treatment. We describe diverse communities that, nevertheless, often are dominated by a few select genera. Our longitudinal analysis shows that high levels of diversity can be maintained for prolonged periods of time in patients with a mildly aggressive disease phenotype but that diversity decreases in patients with more typical progressive lung disease, being driven primarily by antibiotic therapy. We found that interpatient variability in community structures exceeds intrapatient variability in serial samples, indicating that patients harbor distinguishable communities. With decreasing diversity, communities converge to a narrower spectrum of distinct profiles, each being described by the dominant species present. Our analyses suggest that CF airway communities are resilient after antibiotic perturbation and that overall bacterial density remains relatively stable despite markedly decreasing diversity in the advanced stages of lung disease. Significant changes in community structure, diversity, or bacterial density at the time of pulmonary exacerbation were not observed. We have shown that patient clinical state and disease stage are critical in interpreting changes in the diversity and structure of the airway community, underscoring the importance of longitudinal studies in determining patterns of association between the CF airway microbiome and lung disease.

Materials and Methods

Patients, Sputum Specimens, and Medical Record Review. Sputum sample collection and medical record review were approved by the University of Michigan Institutional Review Board. Each of the 126 sputum specimens included in this study was assigned to one of four clinical states based on medical record review (see *SI Materials and Methods* for details). Antibiotic load was calculated as the sum of the number of days a patient received an antibiotic within the 30 d before sample collection (see *SI Materials and Methods* for details).

DNA Extraction and Quantitative PCR. DNA was extracted from sputum samples by treatment with Sputolysin (EMD Chemicals), mechanical disruption, and processing on an automated nucleic acid purification platform (MagNA Pure Compact System; Roche) (see *SI Materials and Methods* for details). Total bacterial load was measured by quantitative PCR using the universal primer/probe set of Nadkarni et al. (16) targeting the bacterial 16S rRNA gene (see *SI Materials and Methods* for details).

DNA Sequencing and Data Analyses. Bar-coded pyrosequencing of the 16S rRNA V3–V5 hypervariable region was performed by the Human Genome Sequencing Center at Baylor College of Medicine using protocols developed for the Human Microbiome Project (http://www.hmpdacc.org/tools_protocols/tools_protocols.php). The software package mothur v. 1.21 (17) was used to process sequences and to calculate alpha and beta diversities. Statistical analyses were performed in PASW Statistics 18 and R packages vegan and ade4 (see *SI Materials and Methods* for details).

1. Harris JK, et al. (2007) Molecular identification of bacteria in bronchoalveolar lavage fluid from children with cystic fibrosis. *Proc Natl Acad Sci USA* 104:20529–20533.
2. van der Gast CJ, et al. (2011) Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. *ISME J* 5:780–791.
3. 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.
4. Sibley CD, et al. (2008) A polymicrobial perspective of pulmonary infections exposes an enigmatic pathogen in cystic fibrosis patients. *Proc Natl Acad Sci USA* 105:15070–15075.
5. Cox MJ, et al. (2010) Airway microbiota and pathogen abundance in age-stratified cystic fibrosis patients. *PLoS ONE* 5:e11044.
6. Klepac-Ceraj V, et al. (2010) Relationship between cystic fibrosis respiratory tract bacterial communities and age, genotype, antibiotics and Pseudomonas aeruginosa. *Environ Microbiol* 12:1293–1303.
7. Guss AM, et al. (2011) Phylogenetic and metabolic diversity of bacteria associated with cystic fibrosis patients. *ISME J* 5:20–29.
8. Rogers GB, Stressmann FA, Walker AW, Carroll MP, Bruce KD (2010) Lung infections in cystic fibrosis: Deriving clinical insight from microbial complexity. *Expert Rev Mol Diagn* 10:187–196.
9. Anderson MJ (2001) A new method for non-parametric multivariate analysis of variance. *Austral Ecol* 26:32–46.
10. Schluchter MD, Konstan MW, Drumm ML, Yankaskas JR, Knowles MR (2006) Classifying severity of cystic fibrosis lung disease using longitudinal pulmonary function data. *Am J Respir Crit Care Med* 174:780–786.
11. Konstan MW, Wagener JS, VanDevanter DR (2009) Characterizing aggressiveness and predicting future progression of CF lung disease. *J Cyst Fibros* 8(Suppl 1):S15–S19.
12. Rogers GB, et al. (2008) Assessing the diagnostic importance of nonviable bacterial cells in respiratory infections. *Diagn Microbiol Infect Dis* 62:133–141.
13. 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.
14. Zhao J, et al. (2012) Impact of enhanced *Staphylococcus* DNA extraction on microbial community measures in cystic fibrosis sputum. *PLoS ONE* 7(3):e33127.
15. Stressmann FA, et al. (2011) Does bacterial density in cystic fibrosis sputum increase prior to pulmonary exacerbation? *J Cyst Fibros* 10:357–365.
16. Nadkarni MA, Martin FE, Jacques NA, Hunter N (2002) Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology* 148:257–266.
17. Schloss PD, et al. (2009) Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75:7537–7541.

Supporting Information

Zhao et al. 10.1073/pnas.1120577109

SI Materials and Methods

Patients, Sputum Specimens, and Medical Record Review. Sputum specimens, collected during the course of routine medical care, were obtained from the University of Michigan Health System clinical microbiology laboratory after processing for bacterial culture and were stored at −80 °C in 0.5-mL aliquots. Specimens were assigned to one of four clinical states [baseline (B), exacerbation (E), treatment of exacerbation (T), recovering from exacerbation treatment (R)] based on medical record review, which assessed physician observations, patient-reported symptoms on the day of sputum collection, and antibiotic use within the 30 d before collection. Antibiotic load was calculated as the sum of the number of days a patient received an antibiotic for maintenance therapy or for treatment of a pulmonary exacerbation within the 30 d before collection (e.g., 1 d on one antibiotic was a load of 1; 1 d on two antibiotics was a load of 2; 2 d on one antibiotic was a load of 2; and so forth).

Definition of Exacerbation. Exacerbation of pulmonary symptoms was identified when explicit language existed in the medical record on or around the date of sputum collection indicating a cystic fibrosis (CF) pulmonary exacerbation ("doctor defined") or when a patient reported pulmonary symptoms that led to the prescription of episodic antibiotics to treat these symptoms. Relevant pulmonary symptoms included increase in cough, change in sputum production (volume and/or appearance), onset or increase of hemoptysis, increased shortness of breath, and/or decreased exercise tolerance. Any combination of two or more symptoms was considered sufficient as long as it was clear in the medical record that the antibiotic treatment was prescribed as a result of the increase in these symptoms. This definition is consistent with a recent consensus report defining exacerbation as the necessity of antibiotics indicated by a change in clinical parameters (1).

Sputum DNA Extraction. Sputum aliquots were thawed on ice and incubated with an equal volume of 10% Sputolysin (EMD Chemicals) at 37 °C for 30 min with pulse vortexing every 5 min. Samples that were not of uniform consistency after 30 min in the presence of Sputolysin were homogenized mechanically for 10 s with a tissue homogenizer (Omni International). Samples then were mixed with 0.9 volume of MagNA Pure Bacterial Lysis Buffer (Roche Applied Science), transferred to an UltraClean Fecal DNA bead tube (MO BIO Laboratories, Inc.), and agitated in a Mini-Beadbeater-8 (Biospec Products Inc.) for 1 min on maximum setting. After digesting with Proteinase K (final concentration 1.4 mg/mL) for 10 min at 65 °C, samples were again agitated in the Beadbeater and incubated at 95 °C for an additional 10 min. DNA was purified using a MagNA Pure Compact System (Roche) automated nucleic acid purification platform according to the manufacturer's DNA Bacteria v3.1 protocol.

Bar-Coded Pyrosequencing. DNA pyrosequencing was performed by the Human Genome Sequencing Center at Baylor College of Medicine. The V3, V4, and V5 hypervariable regions of the 16S rRNA gene were amplified using primer 357F (5′-CCTACGG-GAGGCAGCAG-3′) modified with the addition of the 454 FLX-titanium adaptor "B" sequence (5′-CCTATCCCCTGTG-TGCCTTGGCAGTCTCAG-3′) and primer 926R (5′-CCGTC-AATTCMTTTRAGT-3′) modified with the addition of unique six- to eight-nucleotide barcode sequences and the 454 FLX-titanium adaptor "A" sequence (5′-CCATCTCATCCCTGCGT-

GTCTCCGACTCAG-3′). Barcode and adaptor sequences are found at http://www.hmpdacc.org/doc/HMP_MDG_454_16S_Protocol_V4_2_102109.pdf. PCR amplification was performed on 2 μL of DNA template in a total volume of 25 μL containing 1× AccuPrime Buffer II (Invitrogen Corp), 320 μM of each primer, and 0.03 U/μL AccuPrime High-Fidelity Taq polymerase. Reactions were heated at 95 °C for 2 min followed by 30 cycles of 95 °C for 20 s, 50 °C for 30 s, and 72 °C for 5 min. The concentration of amplicons in each reaction was determined in triplicate using the PicoGreen fluorescent assay (Invitrogen Corp.), and amplicons were pooled in groups of 96 in equal proportion before being sequenced via 454 FLX-titanium according to the manufacturer's specifications.

DNA Sequence Processing and Analysis. Raw sequences were analyzed using mothur v. 1.21 (2) to remove sequences containing homopolymers greater than 8 bp, mismatches in the barcode or primer, one or more ambiguous bases, or an average quality score below 35 over a moving window of 50 bp. Remaining sequences that were at least 200 bp but less than 590 bp in length were curated further to remove chimeric sequences using UCHIME (3) and to reduce sequencing noise by a preclustering methodology (4) before being assigned to operational taxonomic units (OTUs) using an average neighbor algorithm with a 0.03 dissimilarity cutoff. The consensus taxonomy of each OTU was identified at the genus level using the Bayesian method (5). OTUs of interest were classified further at the species level by BLASTn. The total number of reads for each community first was normalized to 1,152, the smallest number of reads among the 126 samples, to control for differences in sequencing depth before alpha (Shannon index) (6) and beta diversity analyses were calculated. Normalization to 1,152 reads still provided very good coverage; on average, 99.4% of sequence reads were associated with OTUs that were observed at least twice.

Quantitative PCR. Each 50-μL reaction contained 1× Taqman Universal PCR master mix (Applied Biosystems, Inc.), 400 nM of each primer, 100 nM of probe, and 1 μL of undiluted sputum DNA. Cycling conditions were as described (7). Each run contained nontemplate control wells and a 10-fold dilution series of *Pseudomonas aeruginosa* genomic DNA. All samples were assayed in duplicate. Standard curve samples were converted from DNA concentration to 16S rRNA copy number based on four rRNA operon copies and 6,463,464 bp per *P. aeruginosa* genome and used to extrapolate total 16S rRNA operon copy number from cycle threshold values for sputum specimens. Operon copy number and genome size were averages calculated from the sequenced *P. aeruginosa* strains PA7, PAO1, and UCBPP-PA14 provided in the Ribosomal RNA Database (8).

Data and Statistical Analyses. The Bray–Curtis distance, which incorporates both membership and abundance and is not affected by shared absence (i.e., zeros in the shared file), was used to compare beta diversity among samples. Log_2-transformed abundance of each OTU was used to dampen the strong influence of dominant OTUs on the intersample comparison while maintaining the greater weight of abundant taxa relative to rare taxa (9).

As described in the main text, the following statistical methods pertain to Fig. 1 and Fig. S1. Because this longitudinal study involved analysis of serial dependent variables, the use of ANOVA and/or a *t* test to ascertain differences between categorical vari-

ables would be inappropriate (i.e., violation of assumptions of independence and normality). Therefore, a linear mixed model that allows for dependence in measures within individual patients was applied to assess the differences in diversity between BETR categories while controlling for age (Fig. 1C). Linear mixed models were conducted using the PASW Statistics 18 software package (IBM SPSS Statistics). Simple linear models within each patient were used separately to examine relationships between Shannon index and age in the R package *Ade4*. Regression lines were drawn by applying the intercept and slopes obtained from a linear regression model with Shannon index or adjusted Shannon index as the dependent variable and age or forced expiratory volume in 1 s (%FEV$_1$) as the independent variable. (Fig. 1B and Fig. S1). Relationships between Bray–Curtis distances and patient variables, including age at time of sputum collection, %FEV$_1$ measurements obtained within 1 d of sputum collection, and antibiotic load within 30 d of sputum collection, were evaluated by the Mantel test in the R package vegan.

The remaining statistical methods pertain to results described in Table S2 and *SI Results*. Univariate and multivariate linear mixed models were performed to assess the impacts of age, %FEV$_1$, and antibiotic load on diversity, with the Shannon index as the outcome variable and age, %FEV$_1$, and antibiotic load as predictors. A P value less than 0.05 was used to indicate a significant association between a predictor and diversity. Two different units of scale for the predictors were evaluated (Table S2). In the first scale traditional units were used. In the second, standardized units, created by dividing each predictor by its corresponding population SD, were applied. The advantage of the latter scale is that the clinical impacts of these predictors across heterogeneous units of measure are comparable when predictors are considered side by side; each parameter is interpreted with respect to a one-SD change in its associated predictor.

SI Results

Community Diversity Relative to Disease Progression and Clinical State.
In Fig. 1C, the top and bottom boundaries of each box indicate 75th and 25th quartile values, respectively, and black lines inside each box represent 50th quartile (median) values. Ends of the whiskers mark the lowest and highest diversity index in each BETR category. The letters above the boxes indicate comparison results in group means; groups with the same letter do not have significant differences in means, but those indicated by different letters are significantly different. Treatment (T, $n = 30$) samples had significantly lower diversity than baseline (B, $n = 50$), exacerbation (E, $n = 17$), or recovering from exacerbation (R, $n = 21$) samples ($P < 0.05$, linear mixed model adjusted for age). R samples had significantly less diversity than E samples, as indicated by the letters above the whiskers. No significant differences in diversity were detected between B and E samples.

Impacts of Age, Antibiotic Therapy, and Lung Function on Diversity.
For a 1-y increase in age in univariate modeling, the Shannon index decreased by 0.087 [95% confidence interval (CI): 0.043, 0.132; $P < 0.001$]. However, after adjusting for antibiotic use and %FEV$_1$, the statistical significance of age vanished (0.003 decrease per year; 95% CI: −0.054, 0.060; $P = 0.91$). For every 1% increase in %FEV$_1$, the Shannon index increased by 0.019 (95% CI: 0.010, 0.027; $P < 0.001$). In the adjusted multivariate model this effect was weakened to being merely suggestive (0.007 increase; 95% CI: −0.004, 0.018; $P = 0.18$). A one-unit increase in the antibiotic use scale decreased the Shannon index by 0.011 (95% CI: 0.007, 0.014; $P < 0.001$). The univariate model including antibiotic use gave the smallest Akaike Information Criterion (AIC) value for predicting the Shannon index (228.97), indicating the best fit among the three univariate models. Antibiotic use also maintained strong statistical significance after adjusting for age and %FEV$_1$ in the multivariate model (0.009 decrease in Shannon index per unit increase; 95% CI: 0.005, 0.014; $P < 0.001$).

When translated into units of SD, the impact of antibiotic use demonstrates both statistical and clinical impact exceeding those of age and %FEV$_1$ in multivariate modeling. A one-SD higher antibiotic score is associated with a 0.306 lower Shannon index (95% CI: 0.153, 0.459 lower; $P < 0.001$), whereas for a one-SD increase in %FEV$_1$, the increase in Shannon index was 0.171 (95% CI: −0.083, 0.424; $P = 0.18$) and for a one-SD increase in age, the decrease in Shannon index was merely 0.009 (95% CI: −0.157, 0.175 lower; $P = 0.91$).

1. Bilton D, et al. (2011) Pulmonary exacerbation: towards a definition for use in clinical trials. Report from the EuroCareCF Working Group on outcome parameters in clinical trials. *J Cyst Fibros* 10(Suppl 2):S79–S81.
2. Schloss PD, et al. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75:7537–7541.
3. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27:2194–2200.
4. Huse SM, Welch DM, Morrison HG, Sogin ML (2010) Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ Microbiol* 12:1889–1898.
5. Cole JR, et al. (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* 37(Database issue):D141–D145.

6. Magurran AE, ed (1988) *Ecological Diversity and Its Measurement* (Princeton Univ Press, Princeton, NJ)
7. Nadkarni MA, Martin FE, Jacques NA, Hunter N (2002) Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology* 148:257–266.
8. Lee ZM, Bussema C, 3rd, Schmidt TM (2009) rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. *Nucleic Acids Res* 37(Database issue): D489–D493.
9. Dethlefsen L, Relman DA (2011) Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc Natl Acad Sci USA* 108(Suppl 1):4554–4561.

Fig. S1. Impacts of age, %FEV$_1$, and antibiotic load on community diversity in progressing patients. (*A*) Effects of age on community diversity without (*Upper*) and with (*Lower*) adjustment for antibiotic load. (*B*) Effects of %FEV$_1$ on community diversity without (*Upper*) and with (*Lower*) adjustment for antibiotic load. Circles represent individual sputum samples and are color coded to reflect BETR category (B, blue; E, red; T, green; R, yellow; U, gray). Adjusted Shannon indexes are residuals after regression with antibiotic load. Regression lines were drawn by applying the intercept and slopes obtained from a linear regression model with unadjusted or adjusted Shannon index as the dependent variable and age (*A*) or %FEV$_1$ (*B*) as the independent variable.

Fig. S2. Pairwise visualization of community structures. Principal coordinate analysis (PCoA) based on Bray–Curtis distances was used to analyze 126 samples from the six patients included in this study. Samples from two different patients are colored in each panel to compare all 15 possible patient pairs in the same principal coordinate 1 (PC1)–principal coordinate 2 (PC2) space as in Fig. 2A. Gray symbols are samples from the four patients not being compared in a panel. (A) Comparisons between pairs of stable patients (*Upper*) and between pairs of progressing patients (*Lower*). (B) Comparisons between pairs of stable and progressing patients.

Fig. S3. PCoA of community membership dynamics using Jaccard distances. The circles in each plot represent the 126 sputum samples included in the study. In each plot, samples from the indicated patient are highlighted in color and are connected by lines in chronological order. The colors reflect BETR category (B, blue; E, red; T, green; R, yellow; U, dark gray). (*Upper*) Community dynamics from the stable patients (S1, S2, and S3). (*Lower*) Community dynamics for the progressing patients (P1, P2, and P3). The scale and variations explained by PC1 (11.7%) and PC2 (6.7%) are the same in each plot. The distance betweens circles on the ordination indicate the relative similarities in community membership.

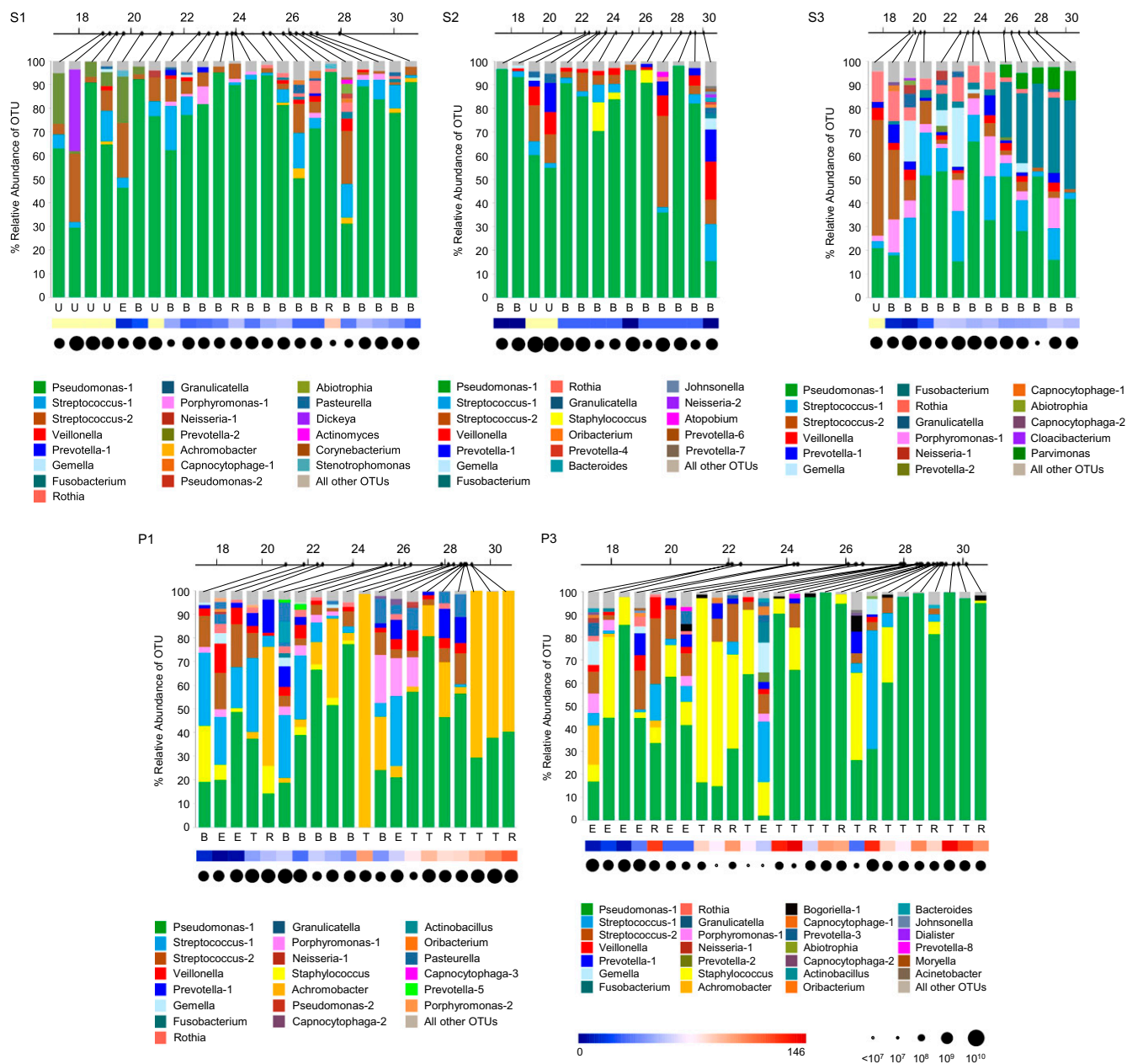**Fig. S4.** OTU relative abundance in serial sputum samples from patients S1, S2, S3, P1, and P3. Relative abundances of OTUs, each accounting for >1% of the community, are shown for each sample. Relative abundance of all remaining OTUs, each representing ≤1% of the community, are shown in gray at the top of each bar. Top line indicates patient age at time of sample collection. Letters beneath the graph indicate the BETR category (Table 2). Heat map indicates antibiotic use (*SI Materials and Methods*) at the time of sputum sample collection. Antibiotic load for U samples are in yellow to indicate no data because of incomplete medical records. Circles indicate total bacterial density (16S rRNA copies/mL sputum) based on quantitative PCR. *Pseudomonas*-1 represents *Pseudomonas aeruginosa*. *Streptococcus*-1 represents species within the *Streptococcus mitis* group. *Streptococcus*-2 represents species within the Streptococcus milleri group.
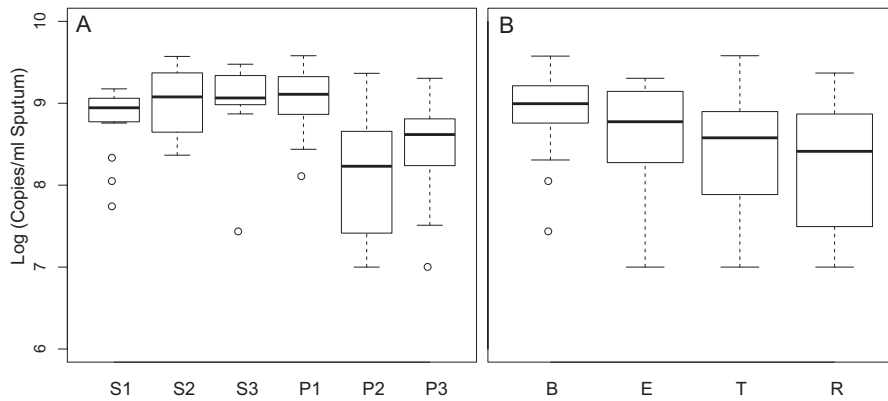
Fig. S5. Comparison of sputum bacterial densities between patients (*A*) and clinical states (*B*). Top and bottom boundaries of each box indicate 75th and 25th quartile values, respectively, and black lines inside each box represent 50th quartile (median) values. Ends of the whiskers mark the lowest and highest bacterial densities within 1.5 times the interquartile range. Outliers, defined as samples with bacterial densities <1.5 times the interquartile range, are shown as open circles. No significant differences in bacterial density were observed between patients or between clinical states by linear mixed model analyses.
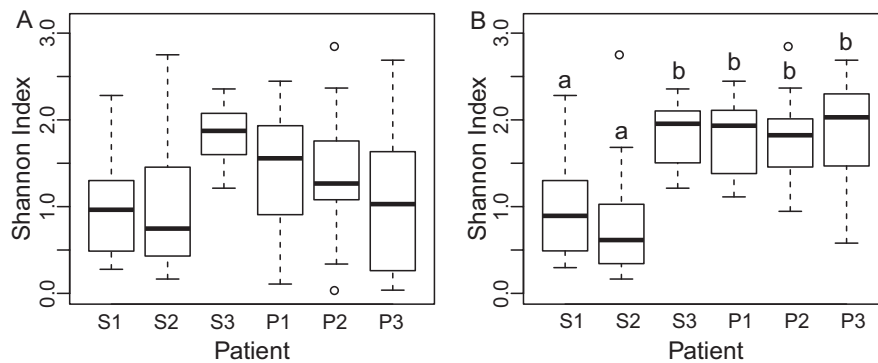
Fig. S6. Comparison of bacterial community diversity among patients included in this study. (*A*) Analysis of communities in all samples from the six patients. (*B*) Analysis of samples obtained only at baseline or exacerbation (before initiation of antibiotic therapy). The top and bottom boundaries of each box indicate 75th and 25th quartile values, respectively, and black lines inside each box represent 50th quartile (median) values. Ends of the whiskers mark the lowest and highest bacterial diversities within 1.5 times the interquartile range. Outliers, defined as samples with bacterial community diversity less than or greater than 1.5 times the interquartile range, are shown as open circles. Lowercase letters above the boxplots in *B* indicate significant pairwise differences between means ($P < 0.05$, one-way ANOVA with Bonferroni correction). Patients S1 and S2 with nonΔF508/ΔF508 *CFTR* genotypes have significantly lower community diversities than do patients S3, P1, P2, and P3, all of whom have a ΔF508/ΔF508 CF genotype.

Fig. S7. Representation of aggressiveness phenotypes of lung disease. This method for topographical depiction of disease phenotypes (1) is adapted from an original analysis by Schlucht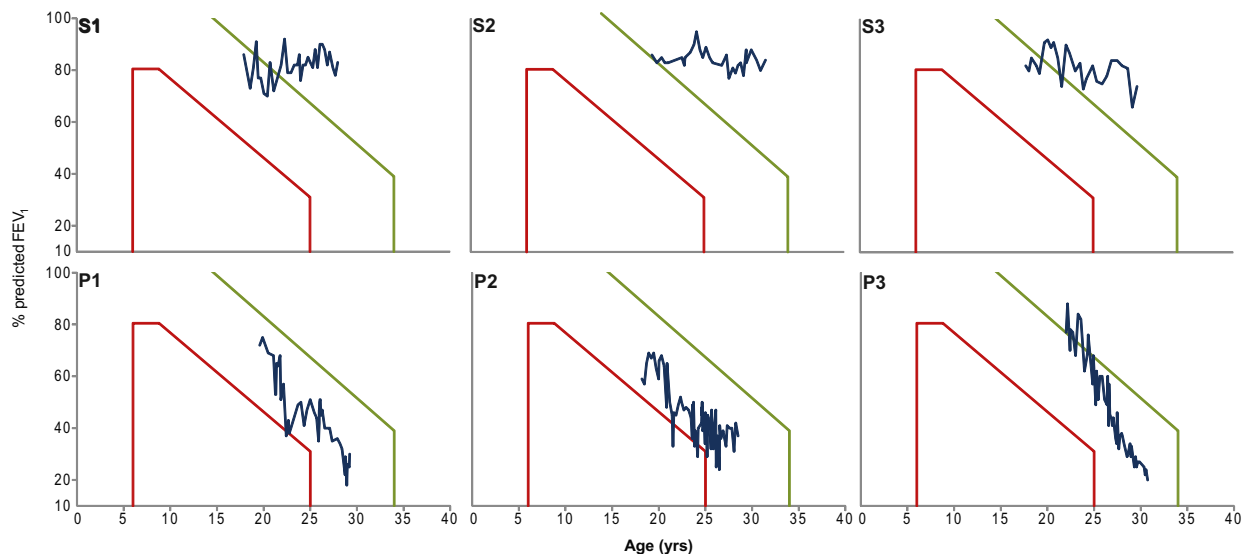er et al. (2). The change in %FEV$_1$ over time for each patient is shown in black. No patient had severely aggressive disease (within the area bounded by red line); three patients (S1, S2, and S3) had mildly aggressive disease (outside the area bounded by green line); three patients (P1, P2, and P3) had moderately aggressive disease (occupying areas between red and green lines).

1. Konstan MW, Wagener JS, VanDevanter DR (2009) Characterizing aggressiveness and predicting future progression of CF lung disease. *J Cyst Fibros* 8(Suppl 1):S15–S19.
2. Schluchter MD, Konstan MW, Drumm ML, Yankaskas JR, Knowles MR (2006) Classifying severity of cystic fibrosis lung disease using longitudinal pulmonary function data. *Am J Respir Crit Care Med* 174:780–786.
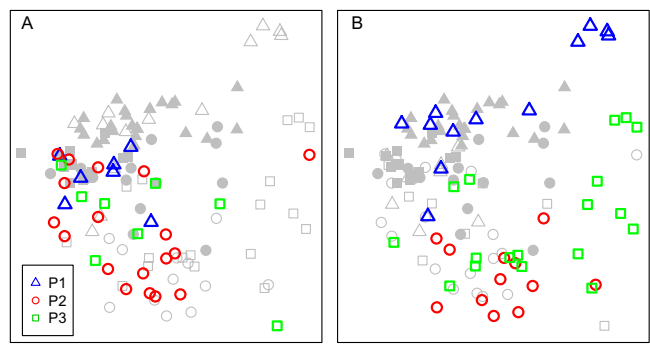
Fig. S8. Community structures in samples from the three progressing patients. PCoA based on Bray–Curtis distances was used to compare communities in samples obtained when %FEV$_1$ was >40 (*A*) or ≤40 (*B*). In the early stages of disease communities could not be differentiated by patient; however, with disease progression, communities from patient P1 became distinct from those from patients P2 and P3 (see also Table S4). The scale and variations explained PC1 (12.7%) and PC2 (6.4%) are the same in each plot. The distances between circles on the ordination indicate the relative similarities in community membership.

Table S1. Prevalence of OTUs

Phylum	OTU	S1 (23)	S2 (14)	S3 (13)	P1 (20)	P2 (30)	P3 (26)	Stable (50)	Progress (76)	All (126)
									Patient groups (no. of specimen)	
Proteobacteria	Pseudomonas-1	100	100	92.3	95.0	96.7	100.0	98.0	97.4	97.6
	Pseudomonas-2	4.3	0.0	0.0	5.0	0.0	0.0	2.0	1.3	1.6
	Achromobacter	26.1	0.0	0.0	85.0	0.0	11.5	12.0	26.3	20.6
	Neisseria-1	4.3	0.0	15.4	10.0	13.3	3.8	6.0	9.2	7.9
	Neisseria-2	0.0	7.1	0.0	0.0	0.0	0.0	2.0	0.0	0.8
	Pasteurella	4.3	0.0	0.0	15.0	10.0	0.0	2.0	7.9	5.6
	Actinobacillus	0.0	0.0	0.0	5.0	10.0	7.7	0.0	7.9	4.8
	Stenotrophomonas	8.7	0.0	0.0	0.0	0.0	0.0	4.0	0.0	1.6
	Dickeya	4.3	0.0	0.0	0.0	0.0	0.0	2.0	0.0	0.8
	Haemophilus	0.0	0.0	0.0	0.0	3.3	0.0	0.0	1.3	0.8
	Acinetobacter	0.0	0.0	0.0	0.0	0.0	3.8	0.0	1.3	0.8
Firmicutes	Streptococcus-1	78.3	71.4	100.0	65.0	70.0	46.2	82.0	60.5	69.0
	Streptococcus-2	91.3	71.4	100.0	75.0	73.3	57.7	88.0	68.4	76.2
	Veillonella	34.8	78.6	69.2	75.0	56.7	34.6	56.0	53.9	54.8
	Staphylococcus	0.0	21.4	0.0	40.0	80.0	73.1	6.0	67.1	42.9
	Granulicatella	21.7	21.4	30.8	35.0	23.3	11.5	24.0	22.4	23.0
	Gemella	8.7	21.4	38.5	20.0	20.0	15.4	20.0	18.4	19.0
	Oribacterium	0.0	7.1	0.0	5.0	6.7	3.8	2.0	5.3	4.0
	Parvimonas	0.0	0.0	38.5	0.0	0.0	0.0	10.0	0.0	4.0
	Abiotrophia	4.3	0.0	7.7	0.0	3.3	3.8	4.0	2.6	3.2
	Johnsonella	0.0	7.1	0.0	0.0	0.0	3.8	2.0	1.3	1.6
	Dialister	0.0	0.0	0.0	0.0	0.0	3.8	0.0	1.3	0.8
	Moryella	0.0	0.0	0.0	0.0	0.0	3.8	0.0	1.3	0.8
Bacteroidetes	Prevotella-1	13.0	42.9	76.9	55.0	56.7	34.6	38.0	48.7	44.4
	Prevotella-2	26.1	0.0	15.4	0.0	33.3	3.8	16.0	14.5	15.1
	Prevotella-3	0.0	0.0	0.0	0.0	6.7	7.7	0.0	5.3	3.2
	Prevotella-4	0.0	7.1	0.0	0.0	3.3	0.0	2.0	1.3	1.6
	Prevotella-5	0.0	0.0	0.0	10.0	0.0	0.0	0.0	2.6	1.6
	Prevotella-6	0.0	7.1	0.0	0.0	0.0	0.0	2.0	0.0	0.8
	Prevotella-7	0.0	7.1	0.0	0.0	0.0	0.0	2.0	0.0	0.8
	Prevotella-8	0.0	0.0	0.0	0.0	0.0	3.8	0.0	1.3	0.8
	Porphyromonas-1	34.8	0.0	84.6	45.0	30.0	15.4	38.0	28.9	32.5
	Porphyromonas-2	0.0	0.0	0.0	10.0	0.0	0.0	0.0	2.6	1.6
	Capnocytophaga-1	13.0	0.0	7.7	0.0	3.3	11.5	8.0	5.3	6.3
	Capnocytophaga-2	0.0	0.0	7.7	10.0	3.3	7.7	2.0	6.6	4.8
	Capnocytophaga-3	0.0	0.0	0.0	5.0	0.0	0.0	0.0	1.3	0.8
	Bacteroides	0.0	7.1	0.0	0.0	3.3	3.8	2.0	2.6	2.4
	Cloacibacterium	0.0	0.0	7.7	0.0	0.0	0.0	2.0	0.0	0.8
	Sphingobacterium	0.0	0.0	0.0	0.0	3.3	0.0	0.0	1.3	0.8
Actinobacteria	Rothia	30.4	7.1	84.6	45.0	33.3	15.4	38.0	30.3	33.3
	Bogoriella-1	0.0	0.0	0.0	0.0	26.7	26.9	0.0	19.7	11.9
	Actinomyces	4.3	0.0	0.0	0.0	0.0	0.0	2.0	0.0	0.8
	Corynebacterium	4.3	0.0	0.0	0.0	0.0	0.0	2.0	0.0	0.8
	Atopobium	0.0	7.1	0.0	0.0	0.0	0.0	2.0	0.0	0.8
	Bogoriella-2	0.0	0.0	0.0	0.0	3.3	0.0	0.0	1.3	0.8
Fusobacteria	Fusobacterium	17.4	7.1	84.6	5.0	3.3	7.7	32.0	5.3	15.9
	Leptotrichia	0.0	0.0	0.0	0.0	3.3	0.0	0.0	1.3	0.8

For each patient or set of patients, the proportion of samples positive for OTUs with a relative abundance of >1% in at least one sample was calculated. Numbers are in percent (%) and represent the proportion of the total number of samples that contain the specified OTU. Values are color coded as follows to indicate the percent of samples within the group positive for each OTU: white, no samples positive; light pink, >0% to <25% samples positive; dark pink, ≥25% to <50% samples positive; purple, ≥50% to <75% samples positive; violet, ≥75% samples positive. *Pseudomonas*-1 represents *Pseudomonas aeruginosa*. *Streptococcus*-1 represents species within the *Streptococcus mitis* group. *Streptococcus*-2 represents species within the *Streptococcus milleri* group.

Table S2. Linear mixed model results assessing impact of age, %FEV₁, and antibiotic load on bacterial community diversity

Predictors	Coefficient (95% CI)	P value	AIC
Univariate linear mixed model (traditional units)			
Age (per year)	−0.087 (−0.132, −0.043)	<0.001	239.48
FEV₁ (per 1%)	0.019 (0.010, 0.027)	<0.001	242.98
Antibiotics (per unit)	−0.011 (−0.014, −0.007)	<0.001	228.97
Multivariate linear mixed model (traditional units)			
Age (per year)	−0.003 (−0.060, 0.054)	0.91	
FEV₁ (per 1%)	0.007 (−0.004, 0.018)	0.18	240.93
Antibiotics (per unit)	−0.009 (−0.014, −0.005)	<0.001	
Univariate linear mixed model (standardized units)			
Age (per SD)	−0.254 (−0.383, −0.124)	<0.001	237.35
FEV₁ (per SD)	0.437 (0.231, 0.643)	<0.001	236.67
Antibiotics (per SD)	−0.362 (−0.483, −0.241)	<0.001	221.96
Multivariate linear mixed model (standardized units)			
Age (per SD)	−0.009 (−0.175, 0.157)	0.91	
FEV₁ (per SD)	0.171 (−0.083, 0.424)	0.18	225.46
Antibiotics (per SD)	−0.306 (−0.459, −0.153)	<0.001	

Univariate models study each predictor separately in predicting diversity. Multivariate models account for all predictors simultaneously in predicting diversity. AIC, Akaike Information Criterion.

Table S3. Average relative abundance (%) of persistent OTUs

S1	S2	S3	P1	P2	P3
All samples					
Pseudomonas-1* (75.0)	*Pseudomonas*-1 (75.2)	*Pseudomonas*-1 (34.3)	*Pseudomonas*-1 (39.2)	*Pseudomonas*-1 (46.6)	*Pseudomonas*-1 (61.2)
Streptococcus-1* (5.3)	*Streptococcus*-1 (3.4)	*Streptococcus*-1 (11.8)	*Streptococcus*-1 (10.3)	*Streptococcus*-1 (6.1)	*Streptococcus*-2 (5.7)
Streptococcus-2* (7.0)	*Streptococcus*-2 (7.2)	*Streptococcus*-2 (9.3)	*Streptococcus*-2 (5.7)	*Streptococcus*-2 (7.0)	*Staphylococcus* (16.0)
	Veillonella (4.0)	*Veillonella* (2.1)	*Veillonella* (3.3)	*Veillonella* (1.2)	
		Prevotella-1 (2.5)	*Prevotella*-1 (3.5)	*Prevotella*-1 (3.3)	
		Fusobacterium (13.7)	*Achromobacter* (23.2)	*Staphylococcus* (18.7)	
		Porphyromonas-1 (6.6)			
		Rothia (5.4)			
Total† 87.3	89.8	85.7	85.2	82.9	82.9
Only samples obtained at baseline and at exacerbation (before antibiotic treatment)					
Pseudomonas-1 (76.0)	*Pseudomonas*-1 (78.2)	*Pseudomonas*-1 (35.4)	*Pseudomonas*-1 (38.5)	*Pseudomonas*-1 (30.9)	*Pseudomonas*-1 (42.5)
Streptococcus-1 (6.3)	*Streptococcus*-1 (3.3)	*Streptococcus*-1 (12.5)	*Streptococcus*-1 (17.0)	*Streptococcus*-1 (8.4)	*Streptococcus*-2 (8.7)
Streptococcus-2 (5.9)	*Streptococcus*-2 (6.1)	*Streptococcus*-2 (6.0)	*Streptococcus*-2 (8.0)	*Streptococcus*-2 (15.1)	*Staphylococcus* (13.8)
	Veillonella (3.3)	*Veillonella* (1.9)	*Veillonella* (4.2)	*Veillonella* (1.3)	**Veillonella (2.5)**
		Prevotella-1 (2.5)	*Prevotella*-1 (2.6)	*Prevotella*-1 (3.4)	**Prevotella-1 (3.0)**
		Fusobacterium (14.8)	*Achromobacter* (7.9)	*Staphylococcus* (6.1)	**Streptococcus-1 (6.3)**
		Porphyromonas-1 (7.0)	**Porphyromonas-1 (5.5)**	**Porphyromonas-1 (2.0)**	**Porphyromonas-1 (3.0)**
		Rothia (4.7)	**Rothia (1.7)**	**Prevotella-2 (19.5)**	
			Staphylococcus (4.2)		
Total† 88.2	90.9	84.8	89.6	86.7	79.8

OTUs in bold are considered persistent only when samples are limited to baseline and exacerbation samples.
Pseudomonas-1 represents *Pseudomonas aeruginosa*. *Streptococcus*-1 represents species within the *Streptococcus mitis* group. *Streptococcus*-2 represents species within the "*Streptococcus milleri*" group.
†Sum of the average relative abundances of each persistent OTU.

Table S4. Pairwise comparison of community structures using AMOVA

	P1	P1L	P2	P2L	P3	P3L
P1						
P1L	NS					
P2	NS	*				
P2L	*	*	NS			
P3	NS	NS	NS	NS		
P3L	*	*	*	NS	NS	

Pairwise comparison of samples collected when $\%FEV_1$ was >40 (P1, P2, P3) or when $\%FEV_1$ was ≤40 (P1L, P2L, P3L). NS, not significant.

*Indicates statistically significant difference in community structures (*P* values less than the Bonferroni adjusted significance level of 0.0024). Communities from early stages of disease are not significantly different between patients; however, in the late stage of disease, communities from P1 become significantly different from those in P2 and P3.