

# SCIENTIFIC REPORTS



OPEN

## A longitudinal characterization of the Non-Cystic Fibrosis Bronchiectasis airway microbiome

T. E. Woo<sup>1,2</sup>, R. Lim<sup>2</sup>, A. A. Heirali<sup>3</sup>, N. Acosta<sup>3</sup>, H. R. Rabin<sup>2,3</sup>, C. H. Mody<sup>2,3</sup>, R. Somayaji<sup>2,3</sup>, M. G. Surette<sup>3,4</sup>, C. D. Sibley<sup>3</sup>, D. G. Storey<sup>1,3</sup> & M. D. Parkins<sup>2,3</sup>

A diverse microbiota exists within the airways of individuals with non-cystic fibrosis bronchiectasis (nCFB). How the lung microbiome evolves over time, and whether changes within the microbiome correlate with future disease progression is not yet known. We assessed the microbial community structure of 133 serial sputa and subsequent disease course of 29 nCFB patients collected over a span of 4–16 years using 16S rRNA paired-end sequencing. Interestingly, no significant shifts in the microbial community of individuals were observed during extended follow-up suggesting the microbiome remains relatively stable over prolonged periods. Samples that were *Pseudomonas aeruginosa* culture positive displayed markedly different microbial community structures compared to those that were positive for *Haemophilus influenzae*. Importantly, patients with sputum of lower microbial community diversity were more likely to experience subsequent lung function decline as defined by annual change in  $\geq -1$  FEV<sub>1</sub>% predicted. Shannon diversity values  $<1$  were more prevalent in patients with FEV<sub>1</sub> decline ( $P = 0.002$ ). However, the relative abundance of particular core microbiota constituents did not associate with risk of decline. Here we present data confirming that the microbiome of nCFB individuals is generally stable, and that microbiome-based measurements may have a prognostic role as biomarkers for nCFB.

Bronchiectasis manifests as the abnormal dilation and thickening of the airways secondary to impaired mucociliary clearance<sup>1</sup>. Studies of bronchiectasis have focused on infections caused by *Pseudomonas aeruginosa* and *Haemophilus influenzae* due to the prevalence of these pathogens in cultured respiratory samples, relationship with accelerated lung function decline, increased exacerbation frequency, and associated morbidity and mortality<sup>2,3</sup>. More recently, however, our understanding of chronic respiratory infections has evolved from a one-host-one-pathogen model to a polymicrobial model of infection<sup>4</sup>. A community of microorganisms, including bacteria, fungi, and viruses, collectively referred to as the lung microbiome, have been demonstrated in airways<sup>5–7</sup>. However, the role these organisms play in health and disease is not well understood<sup>8,9</sup>.

Our knowledge of infections in non-cystic fibrosis bronchiectasis (nCFB) lags behind that in cystic fibrosis (CF). Whereas the lung microbiome of paediatric CF patients is dynamic and changes throughout their adolescence<sup>10,11</sup>, by adulthood a core climax community evolves which remains relatively stable thereafter<sup>11–13</sup>. The observation that microbial diversity within the airways of CF patients inversely correlates with lung function in cross-sectional studies suggests that these communities may influence disease progression<sup>6,9,13</sup>. However, how the microbiota changes over time and associates with disease progression remains to be elucidated. With this study, we sought to characterize the natural history of airways communities within individuals with nCFB using a longitudinal collection of samples. Furthermore, we sought to determine if the nCFB microbiota correlated with subsequent long-term pulmonary outcomes and might thus serve as a novel biomarker to predict future clinical course.

<sup>1</sup>Department of Biological Sciences, University of Calgary, Calgary, AB, Canada. <sup>2</sup>Department of Medicine, University of Calgary, Calgary, AB, Canada. <sup>3</sup>Department of Microbiology, Immunology and Infectious Diseases, University of Calgary, Calgary, AB, Canada. <sup>4</sup>Departments of Medicine, and Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Ontario, Canada. Correspondence and requests for materials should be addressed to M.D.P. (email: [mdparkin@ucalgary.ca](mailto:mdparkin@ucalgary.ca))

Patient Demographics (n = 29)	Decliners (n = 11)	Stables (n = 18)	Total (n = 29)
Median Age at Study Enrolment (IQR)	63 (39–70)	57 (50–68)	60 (48–69)
Median FEV <sub>1</sub> at enrolment (% predicted) (IQR)	71 (45–93)	53 (39–67)	55% (40.5–74)
<b>Respiratory Comorbidities</b>			
Sinusitis	5 (45%)	10 (56%)	15 (52%)
Asthma	1 (9%)	3 (17%)	4 (14%)
Chronic Obstructive Pulmonary Disease (COPD)	0	2 (11%)	2 (7%)
<b>Treatments Recorded During Sample Collection</b>			
Inhaled corticosteroids	7 (64%)	14 (78%)	21 (72%)
Long-acting beta-agonists (LABA)	6 (55%)	14 (78%)	20 (69%)
Long-acting muscarinic antagonist (LAMA)	1 (9%)	8 (44%)	9 (31%)
Short-acting beta-agonists (SABA)	6 (55%)	12 (67%)	18 (62%)
Ipratropium bromide	3 (27%)	7 (39%)	10 (34%)
Saline	2 (18%)	4 (22%)	6 (21%)
Proton-pump inhibitors	1 (9%)	2 (11%)	3 (10%)
H <sub>2</sub> -receptor antagonists	1 (9%)	1 (6%)	2 (7%)
<b>Cultured Pathogens</b>			
<i>Pseudomonas aeruginosa</i>	6 (55%)	10 (56%)	16 (55%)
<i>Haemophilus influenzae</i>	4 (36%)	4 (22%)	8 (28%)
<i>Staphylococcus aureus</i>	3 (27%)	5 (28%)	8 (28%)
<i>Moraxella catarrhalis</i>	2 (18%)	1 (6%)	3 (10%)
<i>Streptococcus pneumoniae</i>	1 (9%)	3 (17%)	4 (14%)

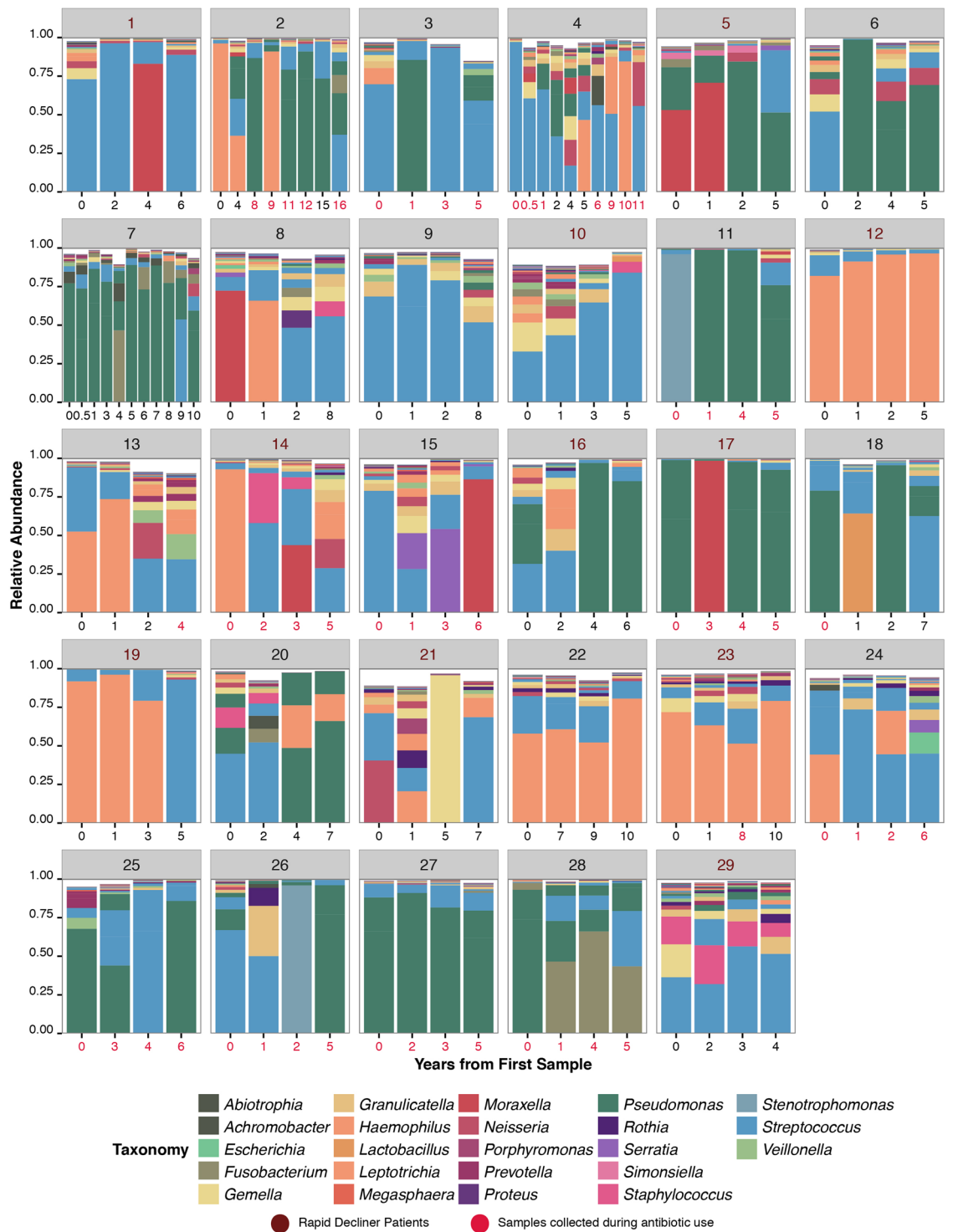
**Table 1.** Demographics and treatments used in nCFB patients at enrolment. Inter-quartile Range (IQR) and percent are indicated by the parenthesis.

## Results

**Patient Demographics.** One hundred and thirty-three sputum samples from 29 patients were included. Demographics and baseline patient characteristics are listed in Table 1. The median duration of follow-up was 5.8 years (IQR 5.0–7.9, range 4.3–16). The majority of patients were never-smokers (93%; 27/29). Etiologies of bronchiectasis included post-infectious (62%), idiopathic (34%) and ciliary dyskinesia (4%). Lung function was generally stable during follow up with a median change in FEV<sub>1</sub>% percent predicted of  $-0.29\%/year$  (IQR:  $-1.3\%+0.65\%$ ). Twelve exacerbation events were captured in the selected sputa, from 10 of the 29 (34%) patients. Two patients ultimately died and no patients received lung transplantation during the study period.

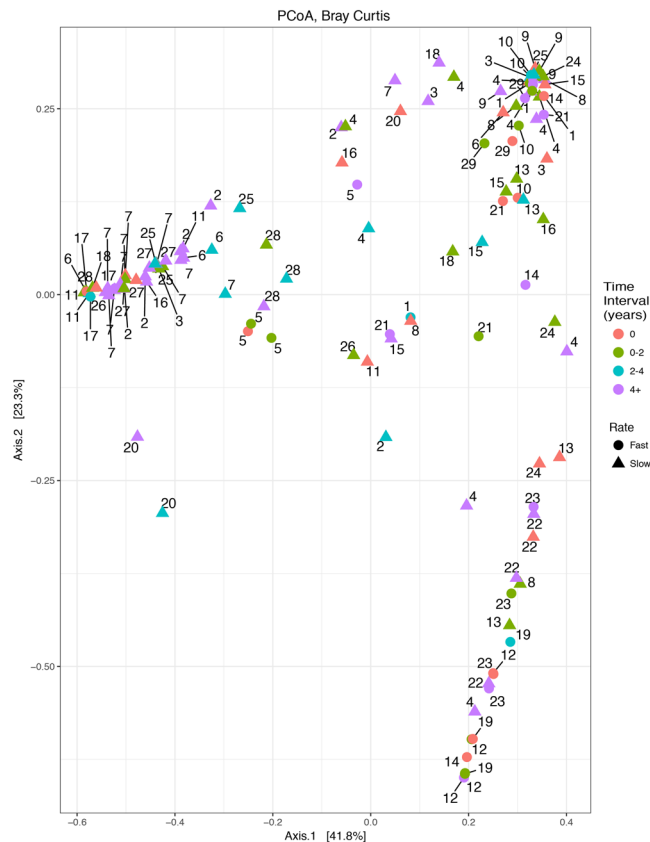
**The nCFB Microbial Community Displays Inter- and Intra-patient Variability.** A total of 7,698,972 sequences were generated (median of 49,437 sequences/sample, IQR: 33,655–73,464, range: 5136–170,812) with 3202 total observed OTUs (median of 138 OTU per sample, IQR: 90–215, range: 29–817). Across the sample population, the majority of OTUs were classified as *Streptococcus* (30%), *Pseudomonas* (28%), and *Haemophilus* (17%). The longitudinal lung microbiota of taxa found in >1% relative abundance in the dataset was visualized using taxonomic summaries (Fig. 1). Notably, we observed varying degrees of inter- and intra-patient heterogeneity. Whereas patients 22, 23 and 27 had remarkably stable lung microbiomes, patients 4, 11, 15 and 26 demonstrated considerable temporal variation. When samples collected during exacerbations were excluded, significant clustering of samples by patient was observed (PERMANOVA,  $p = 0.03$ ) (Fig. 2). In order to understand the natural history of the nCFB airways microbiota, we compared the composition of samples by first binning them into distinct time intervals (Fig. 3A–C). No significant changes in alpha-diversity was observed in any of these periods relative to the initial samples and PERMANOVA analysis did not reveal any community-wide differences ( $p = 0.54$ ). In particular, Bray-Curtis analysis did not reveal any significant inter- and intra-patient changes between the initial, intermediate or final sample intervals - although certain individuals had dynamic ranges (Fig. 3D).

**The nCFB microbiome associates with disease trajectory.** We investigated whether the baseline microbiota - as established from the first sputum sample on record - correlated with either baseline disease severity or subsequent lung function decline using our prospective collection. Notably, baseline demographics were not significantly associated with changes in microbial diversity (data not shown). In our small sample population, we did not confirm that disease severity at baseline associated with either alpha or beta-diversity (Fig. S1). We did, however, observe that baseline microbial diversity correlated with subsequent rate of lung function decline (Fig. 4). Eleven (38%) patients met our *a priori* definition for Decliner ( $> -1\%/year$  FEV<sub>1</sub>). Median rates of annual FEV<sub>1</sub> decline between Decliners were  $-1.8\%$  (IQR:  $-1.3-1$ ) vs.  $+0.4\%$  (IQR:  $-0.07-1.25$ ) in Stable patients ( $p = 0.004$ ), respectively. Baseline measurements of Observed OTU ( $p = 0.90$ ), Shannon diversity ( $p = 0.51$ ), or Simpson diversity ( $p = 0.64$ ) were not significantly different between both groups. When comparing these two groups, we identified that Decliners generally had lower alpha-diversity using the Observed OTU ( $p = 0.05$ ), but not Shannon ( $p = 0.097$ ) and Simpson ( $p = 0.06$ ) indices although similar trends were evident. When rarefied, significant differences in Shannon ( $p = 0.035$ ) and Simpson diversity ( $p = 0.033$ ), but not



**Figure 1.** Taxonomic summaries of OTU present in >1% relative abundance of all samples (n = 133) obtained from a cohort of 29 nCFB patients as a function of the years from first sample. Samples for each patient are grouped together by a unique patient identifier – as indicated by the grey heading. Patient’s numbers from the Decliner group are indicated in magenta. Analysis of the taxonomic summary reveals a unique microbiome associated with each patient and varying degrees of inter and intra-patient diversity. Samples collected under antibiotic pressure are indicated by red text indicating year of collection.

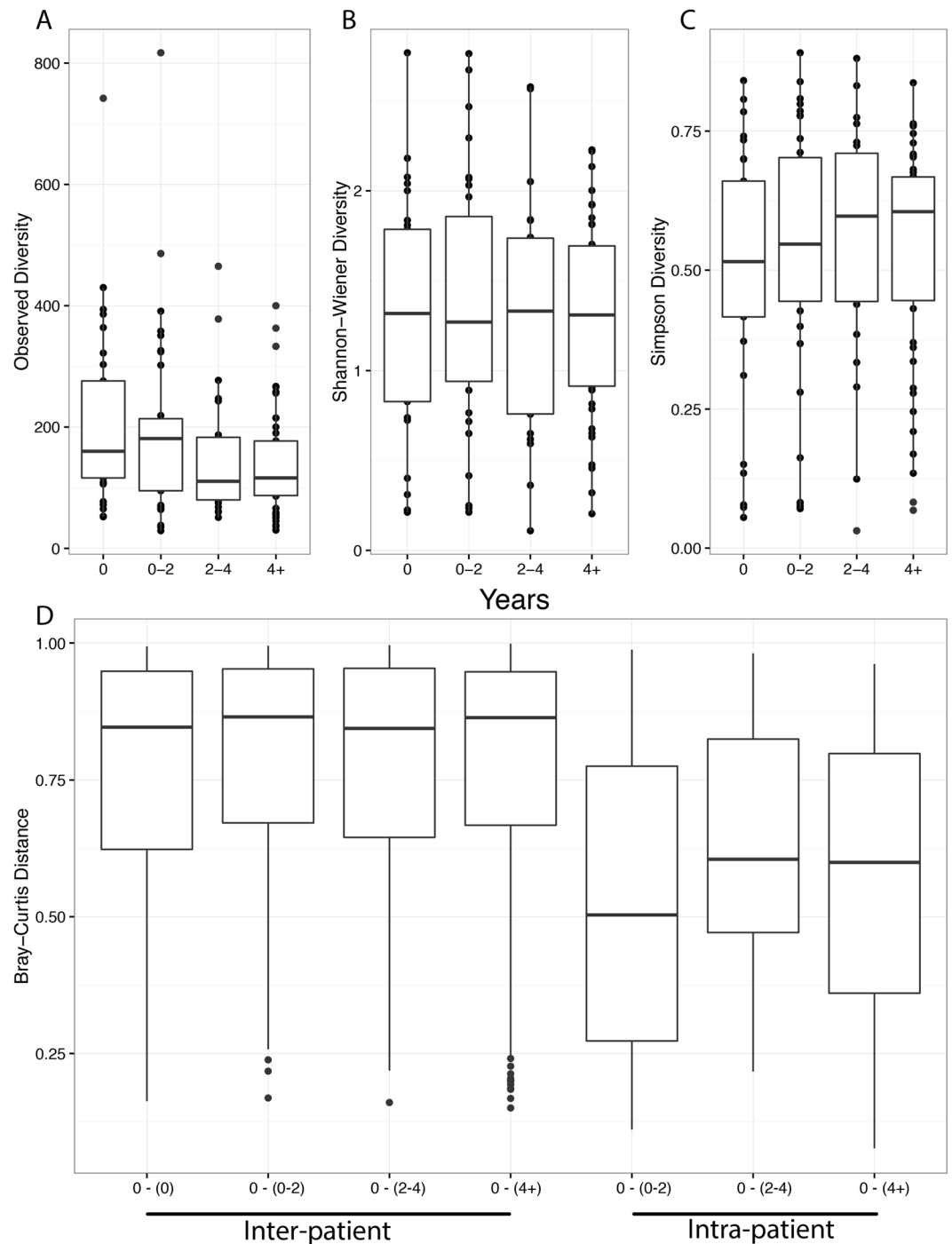
Observed OTUs (p = 0.07) were observed. However, no community-wide differences were observed between the two patient groups (p = 1.0). When using alternate definitions for Decliners of –0.5%/year and –1.5%/year the same trends were evident, but did not reach significance (data not shown).



**Figure 2.** Clustering of samples by patient during stable periods (excluding exacerbation events) was observed ( $n = 121$ ). Samples obtained from each patient are designated by the patient number adjacent to each point. Samples were categorized by color based on the time since baseline sample. Patients classified as Decliners are indicated by a circle, whereas patients with a Stable course are represented by triangles. Significant clustering was observed using PERMANOVA analysis ( $p = 0.03$ ).

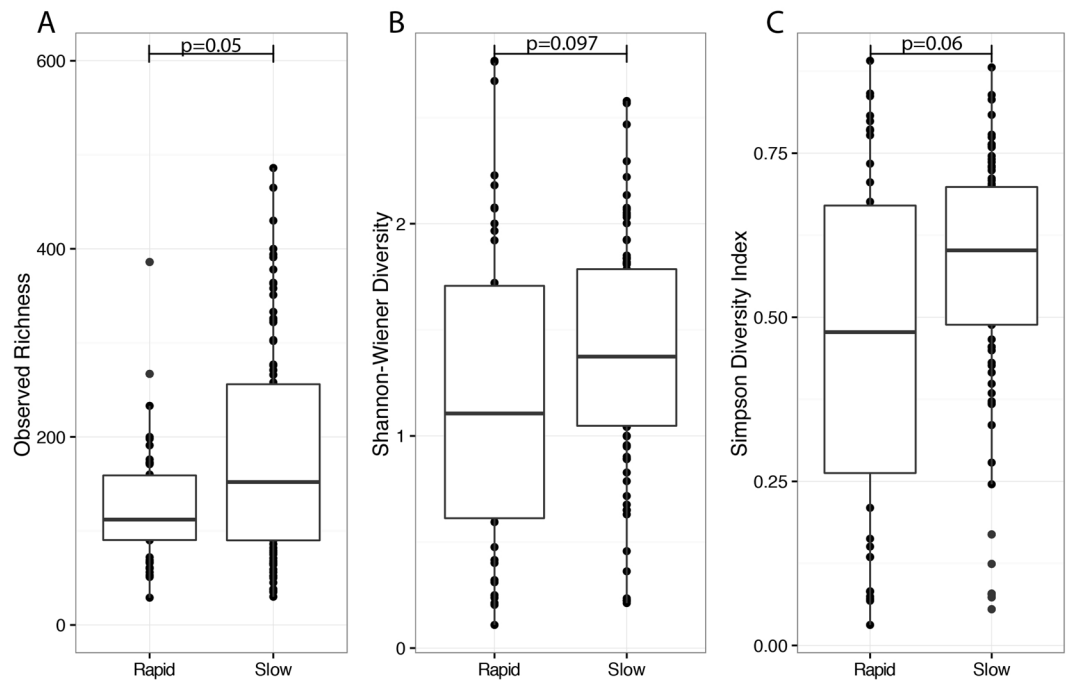
As lower microbial diversity was observed to associate with Decliner status, we investigated whether the initial Observed OTU, Simpson, and Shannon diversity measurement was correlated with change in lung function. No significant correlation between the initial diversity measures for each patient and annual rate of lung function decline was found using a generalized linear model using the Shannon ( $p = 0.61$ ), Observed ( $p = 0.23$ ) and Simpson diversity ( $p = 0.61$ ) indices (Fig. S2). Likewise, we found that no association between the abundance of specific genera (found within the overall community at  $>1\%$  of reads) and subsequent lung function decline. The initial abundance ( $t = 0$ ) of *Pseudomonas* ( $r_s = -0.21$ ), *Haemophilus* ( $r_s = 0.34$ ), *Streptococcus* ( $r_s = -0.18$ ), *Prevotella* ( $r_s = -0.29$ ), *Fusobacterium* ( $r_s = -0.18$ ), *Gemella* ( $r_s = -0.10$ ), *Granitocella* ( $r_s = -0.06$ ), *Neisseria* ( $r_s = 0.08$ ), *Rothia* ( $r_s = -0.08$ ), and *Staphylococcus* ( $r_s = 0.17$ ) within each sample did not correlate strongly with lung function for each patient. Lastly, we examined whether Shannon diversity measurements associated with the clinical prognosis of patients within our study. Using a Shannon diversity index cut-off value of 1, used for biomarker studies in CF<sup>14</sup>, we observed that the majority of samples from Stable patients had a Shannon diversity value of  $>1$  (73/93 samples) as compared to Decliners (20/40 samples) ( $p = 0.002$ ). When the analysis was constrained by patient, similar trends were observed but did not reach significance ( $p = 0.39$ ). Samples with a Shannon diversity index  $<1$  were associated with changes in the microbial community ( $p = 0.04$ ). In particular the abundance of seven genera had relative lower abundance; *Gemella*, *Lactobacillus*, *Nicoletella*, *Haemophilus*, *Bordetella*, *Staphylococcus*, and *Achromobacter* (Fig. S3) whereas *Streptococcus* and *Moraxella* were increased.

***Pseudomonas* and *Haemophilus* display an antagonistic relationship in nCFB airways.** We observed a strong correlation between airway pathogens obtained during routine real-time clinical laboratory culture and microbial community diversity determined retrospectively from these same samples. *P. aeruginosa* was isolated in 60 (45%) of the total samples (16/29 patients), whereas *H. influenzae* was only found in 17 samples (13%) (8/29 patients). Samples which were culture positive for *P. aeruginosa* trended towards coming from individuals with worse baseline lung disease relative to *H. influenzae* positive individuals (median FEV<sub>1</sub> of 53% (IQR: 36–62) versus 65% (IQR: 39–75) ( $p = 0.10$ ). We observed that the real-time isolation of *P. aeruginosa* was exclusive to *H. influenzae* (and vice-versa) such that these two organisms were never recovered together in culture from the same sample. Further differences arose when assessing the community microbiota in samples obtained in association with *P. aeruginosa*, *H. influenzae*, and in samples with neither (Fig. 5). When comparing these three groups, we found that *P. aeruginosa* culture positive samples had relatively higher Simpson ( $p = 0.01$ )



**Figure 3.** The microbial community was observed to remain relatively constant over the study period as measured by Observed OTUs (A), Shannon-Wiener (B) and Simpson diversity (C), and Bray-Curtis dissimilarity (D). Samples were categorized based on the time interval relative to first sample as: 0 (n = 29), 0–2 years (n = 36), 2–4 (n = 22), and 4 + (n = 46) years since initial sample collection. The median and interquartile ranges (IQR) are represented by the middle, top, and bottom lines of each boxplot. No statistically significant comparisons were observed.

and Shannon ( $p = 0.05$ ) diversity, but not Observed OTU diversity ( $p = 0.85$ ) as compared to *H. influenzae* culture positive samples. No significant differences in the alpha diversity were observed between either *P. aeruginosa* or *H. influenzae* culture positive samples and samples where neither of these organisms were identified. Community-wide differences were observed using PERMANOVA analysis ( $p = 0.001$ ). In particular, the relative abundance of 10 genera was found to be significantly different between samples (Fig. S4). Interestingly, the increased log-fold relative abundance of *Pseudomonas* was observed in tandem with a >5 fold decrease change in relative abundance of *Haemophilus* (Fig. S5).



**Figure 4.** Sputum samples collected from patients ( $n = 44$  samples, 11 patients) who were categorized as Decliners ( $\geq 1$  FEV<sub>1</sub>% per year) generally displayed lower levels of alpha diversity as compared to Stables ( $n = 89$ , 18 patients). No significant differences were detected using the Observed OTU (A), Shannon (B), and Simpson (C) diversity indices. The median and interquartile ranges of each group are represented by the middle, top, and bottom lines of each boxplot. Statistical analysis was performed using the Wilcoxon-signed rank test.

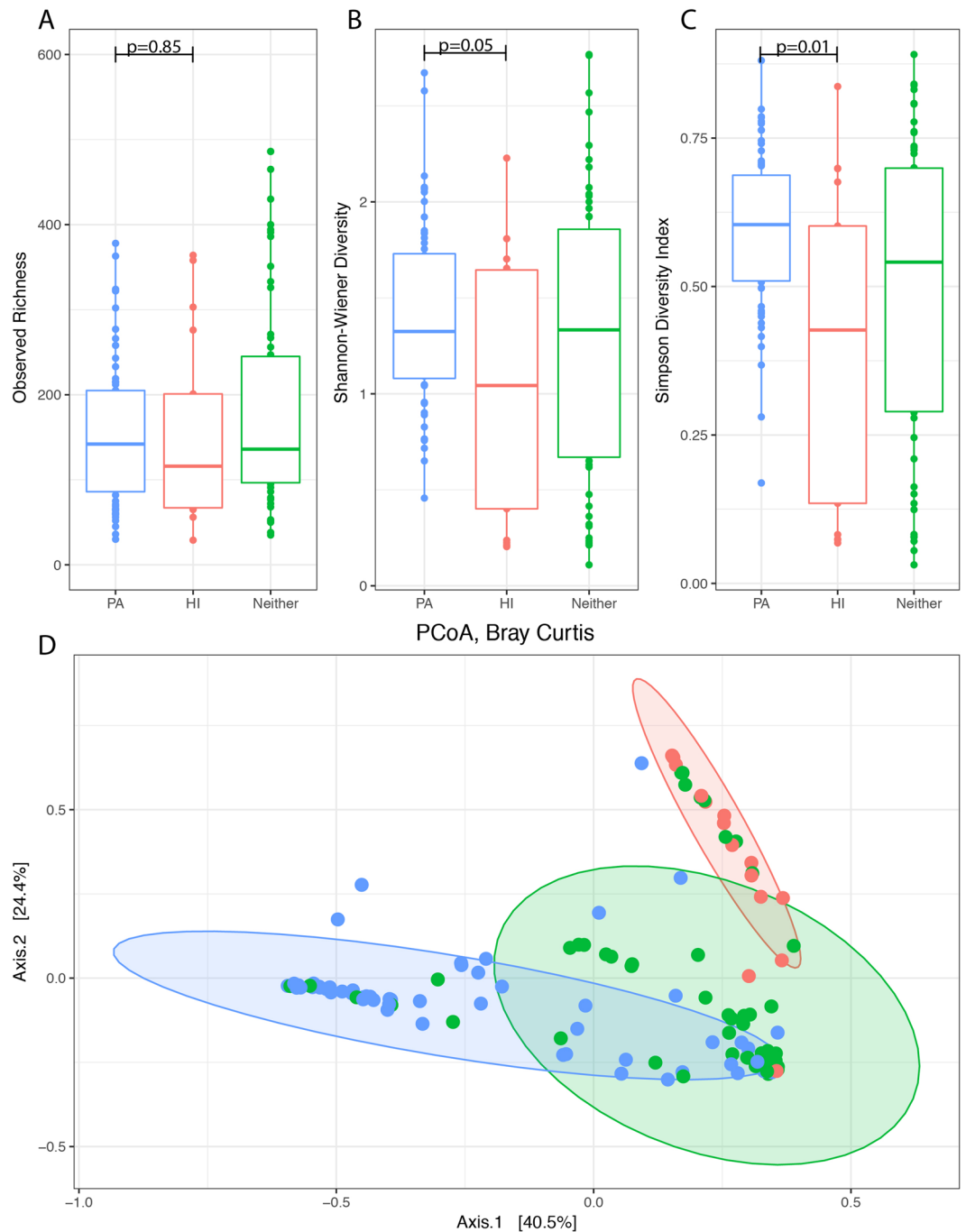
**Acute antibacterial therapies introduce transient shifts within the microbiome.** When we examined the influence of acute antimicrobial therapy with microbial changes in community structure, only fluoroquinolone use was associated with changes in diversity. Observed OTU richness ( $p = 0.004$ ) was reduced under acute antibacterial pressures, but not Shannon ( $p = 0.19$ ) or Simpson ( $p = 0.55$ ) diversity indices (Fig. 6). Interestingly, community-wide differences were observed using PERMANOVA analysis ( $p = 0.01$ ). In particular, samples obtained during fluoroquinolone use were found to have significant changes in the abundance of 12 genera. Two of the twelve genera, *Escherichia* and *Moraxella*, had increased relative abundance and the remaining had decreased relative abundance including: *Veillonella*, *Dialister*, *Azorhizophilus*, *Bulleidia*, *Capnocytophaga*, *Lactobacillus*, *Prevotella*, *Treponema*, *Haemophilus*, and *Achromobacter* (Fig. S6).

## Discussion

The lung microbiome has been suggested to play an important role in modifying disease progression for individuals with CF and nCFB<sup>6,11,12</sup>. Whereas recent studies have alluded to the relative stability of the nCFB microbiome over short periods<sup>15</sup>, we believe our ability to assess serial samples collected up to 16 years apart, provides unique insight into the longer-term adaptations of the lung microbiome. Interestingly, we observed no significant changes in the alpha and beta-diversity of the lung microbiome. This is all the more surprising given the extended duration of observation (for some patients up to 16 years). Stability was not influenced by baseline clinical characteristics or disease severity. Indeed, even over protracted periods of time, we observed, as others have over short time periods, microbiota tended to cluster by patient, alluding to unique patterns of colonization<sup>15</sup>. In comparison, studies in CF suggest that the childhood microbiota fluctuates and changes over time, while the adult lung microbiome remains relatively stable<sup>10,12,16</sup>.

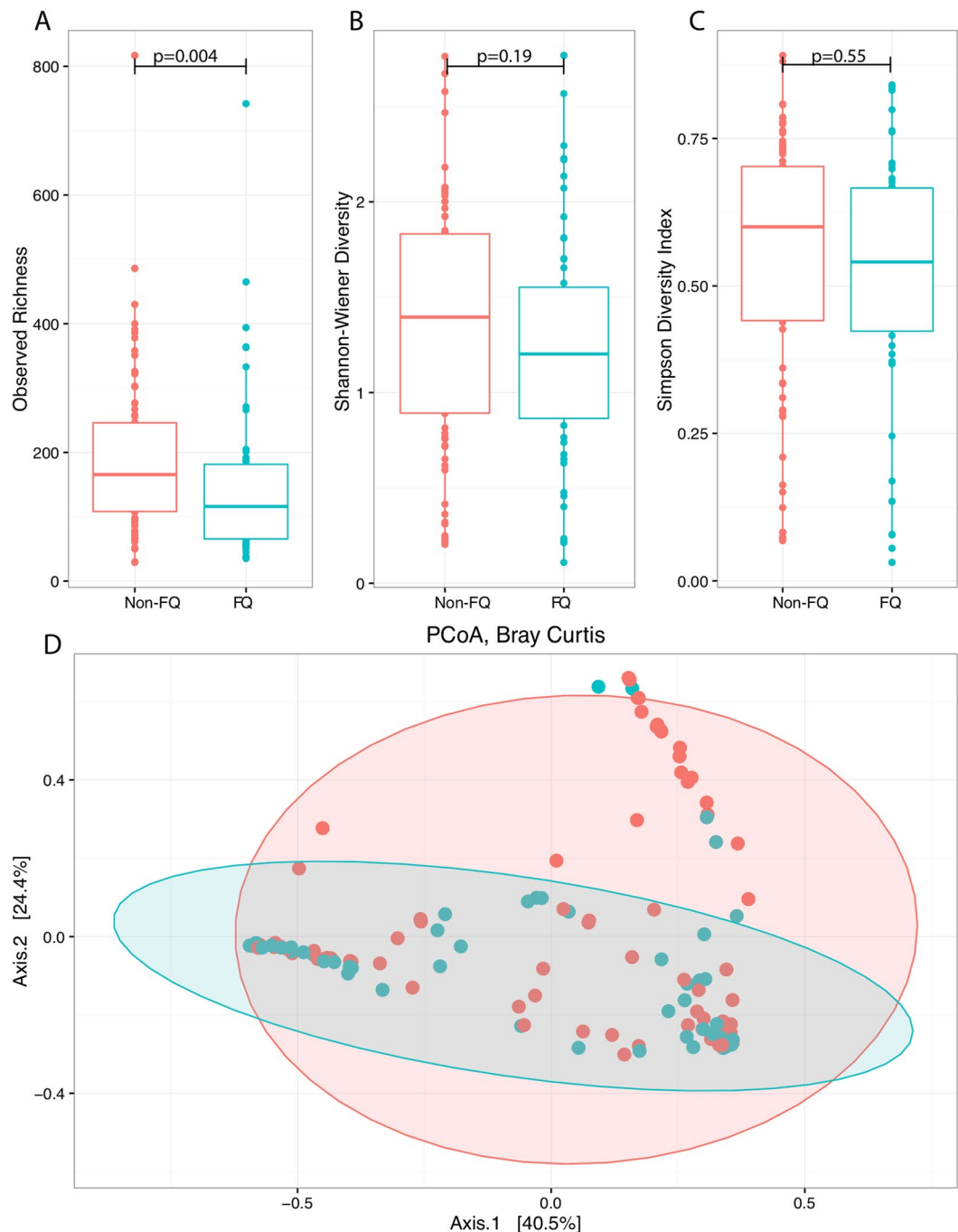
Given the limited capacity by which cultured pathogens within nCFB airways predict clinical outcomes, the potential of the microbiome to serve as a biomarker for clinical management is of great interest. For instance, early studies have alluded to the predominance of *P. aeruginosa* and *Veillonella* species as predictors for future exacerbations<sup>17</sup>. In contrast to early point prevalence studies, we observed no significant relationship between alpha diversity and static measures of lung function<sup>6,18</sup>. However, these trends are not ubiquitous amongst all studies, with other studies showing no association between diversity and lung function<sup>19</sup>. Rather, we observed a general relationship between alpha diversity and the rate at which lung function declined in these patients. It is likely that the slow rate of progressive lung disease was prohibitive in establishing community-wide differences as no significant decline was observed between baseline and end-of-study lung function measures within our cohort. Furthermore, we observed that samples with Shannon diversity values  $< 1$  tended to be associated with Decliner status and may warrant further interest in its use as a biomarker.

Amongst the pathogens isolated from the airways, *P. aeruginosa* and *H. influenzae* play an important role. The presence of these organisms within the airways has been associated with decreased microbial diversity, worsened clinical outcomes, and increased morbidity and mortality<sup>4,11,20,21</sup>. Importantly, the interactions



**Figure 5.** Observed OTU (A), Shannon (B), and Simpson (C) diversity, and Bray-Curtis dissimilarity (D) illustrate clustering of *P. aeruginosa* and *H. influenzae* positive samples. Differences in the sputum microbial diversity of samples collected in relation to the growth of *Pseudomonas aeruginosa* (PA, n = 60) and *Haemophilus influenzae* (HI, n = 17) and in samples with neither (Neither, n = 56). The ellipses for each group represent the 90% confidence interval. The median and interquartile ranges (IQR) are represented by the middle, top, and bottom lines of each boxplot. Statistically significant comparisons were shown using the p-value.

between these airway pathogens is of great interest, with an antagonistic relationship observed in both cultured and culture-independent microbiome<sup>4,8</sup>. Similarly, we believe that these trends may reflect both direct and indirect competition between these two species. It is possible that it is the accessory microbiome associated with *Pseudomonas*-dominant patients that may exclude the colonization of *H. influenzae* and vice-versa<sup>8</sup>. Further, this competition may extend to other organisms and may explain the change in genera observed within our study. Though our V3/V4 sequencing of the 16S rRNA allowed only for genus level classification, our complimentary



**Figure 6.** Acute fluoroquinolones use was associated with reduced community richness as represented by the Observed OTU. Samples collected during acute fluoroquinolone use (FQ,  $n = 55$ ) were compared to samples collected in their absence (No-FQ,  $n = 78$ ). Significant differences were seen in Observed OTUs (A), but not for Shannon (B) or Simpson diversity (C). Significant community-wide differences were found using PERMANOVA when permutations were constrained by patient and visualized using PCoA ( $p = 0.01$ ).

analysis of real-time culture data revealed that the only *Pseudomonas* and *Haemophilus* species identified from these samples were *P. aeruginosa* and *H. influenzae*.

The impact of acute exacerbations on microbial communities is not well understood – with no significant changes in the microbiome associated with the onset of these events<sup>19</sup>. Antibiotic use is the mainstay of management of exacerbations in patients with CF and nCFB. Indeed, antibiotic therapy may influence the respiratory microbiota and play an important role in clinical prognosis<sup>4,11,20,21</sup>. Within our study and in congruence with prior findings<sup>15</sup>, we suggest that the broad-spectrum effects of fluoroquinolone decrease the overall richness without affecting the evenness of the microbial community.



Importantly, this retrospective study of prospectively collected samples has a number of limitations. While our sample size of 29 patients is limited, it remains comparable in size to similar microbiome studies in nCFB<sup>6,19</sup>, and even longitudinal studies in CF<sup>16,21</sup>. Furthermore, we believe that the repeated longitudinal measurements analysis – assessing a time period vastly greater than prior works – provides unique insight into the natural history of the disease well beyond cross-sectional studies. Notably, as the primary focus of our study was to assess longitudinal changes in the microbial community, our analysis may have been insensitive to the influence of antimicrobial therapies or the incidence of pulmonary exacerbations on microbial diversity – as seen in previous studies<sup>15,16,22,23</sup>. Given our use of “opportunistically collected samples” during clinical encounters, it may be that our analysis is insensitive to detect changes in the microbial diversity prior to the onset of an exacerbation. As such, studies characterizing samples in their longitudinal relation to such acute events are important in highlighting other factors which may lead to onset of exacerbation<sup>21</sup>. Furthermore, as the vast majority of samples were collected during stable periods it is possible that little antibacterial selective pressures persisted to drive changes in the microbial community. Previous studies in CF have observed that microbial diversity did not change significantly during periods of stability<sup>16</sup>, which we suspect also extends to nCFB. Lastly, our collection of samples is derived from a regional referral clinic and patients are not exclusively seen by nCFB clinic providers and therefore may not be all encompassing.

The Calgary nCFB Biobank represents a unique opportunity in which to characterize the longitudinal nCFB lung microbiome. To our knowledge, this study represents the longest examination of the nCFB microbiome to-date suggesting that the lung microbiome is relatively stable over time, and is highly individualized. We suggest that the lung microbial diversity may be an important contributor to clinical course – although this is likely but one of many host, organism/community and environmental factors. Furthermore, we confirmed the reciprocal relationship between *P. aeruginosa* and *H. influenzae* in nCFB airways<sup>4,8</sup>. This study has provided further insight into the longitudinal microbiome of individuals with nCFB, and suggests that further studies exploring the microbiome’s association with subsequent clinical outcomes are warranted in our attempt to develop biomarkers for patient prognostication.

## Methods

**Patients and samples.** The Calgary Bronchiectasis Biobank (1998–current) is a prospectively inventoried collection of sputum obtained from nCFB adults (>18 years) as part as their routine clinical care and stored at –80 °C. Patients followed at the Calgary Bronchiectasis Clinic have symptoms consistent with and radiographic evidence of bronchiectasis<sup>24</sup>. For inclusion in this study patients had to have ≥4 sputum samples spanning ≥4 years and a diagnosis of nCFB. Individuals with bronchiectasis due to CF were excluded.

**Clinical Information and Definitions.** Patient demographic and clinical information was collected through detailed chart review. Pathogens detected through traditional culture-based approaches from these samples were reported in real-time from the clinical microbiology laboratory and these data were extracted during chart review. Baseline patient characteristics including bronchiectasis aetiology, gender, respiratory therapies, antibacterial therapies and co-morbidities were analyzed for their relationship to microbial diversity. Bronchiectasis aetiology was classified as post-infectious, idiopathic, or other. Lung disease stages were classified by spirometry as determined using the Knudson calculation as advanced; Forced expiratory volume in one second (FEV<sub>1</sub>) < 40% predicted, moderate; 40–70%, mild; ≥70%<sup>18</sup>. Rates of lung function decline were calculated by dividing the FEV<sub>1</sub>% by the total number of years followed for each patient. Patients with an annual decline in FEV<sub>1</sub>% predicted of ≥1% were *a priori* considered Decliners (D)<sup>25,26</sup> whereas those whose decline was less than this were considered Stables (S). Samples were separately coded and considered to be collected during acute antimicrobial therapy if obtained within two weeks of a new antibiotic administration (as acute parenteral antibiotic therapies have been associated with transient changes in the microbiota of suppurative lung disease)<sup>27–29</sup>. In order to assess differences in community composition over time, samples were categorized based on years from collection relative to the initial sample (0), >0.25–2 years, >2–4 years, and >4 years for each patient over the study period. The study was performed in accordance with relevant guidelines and regulations of the University of Calgary and consistent with those required by Scientific Reports. The Conjoint Health Research Board of the University of Calgary has approved the ongoing collection and maintenance of the biobank and the samples maintained in it (REB16-0035). Furthermore, the CHREB has granted the investigators permission to evaluate patient outcomes associated with biobank derived samples REB16-0854.

## DNA Isolation, 16s rRNA Variable 3–4 (V3/V4) Amplification, and Sequence Processing.

Genomic DNA was isolated from sputum samples as previously described<sup>30,31</sup>. Barcoded universal primers adapted from Bartam *et al.* were used to amplify the V3 and V4 region of the 16s rRNA prior to MiSeq Illumina sequencing<sup>32</sup>. Reagent blanks were run for each set of DNA extractions and DNA amplification. If contamination was seen in the technical controls, samples were excluded and the DNA extraction and amplification was subsequently repeated. Technical replicates were not sequenced. Sequence analysis was carried out using previously published Perl scripts<sup>33</sup>. Low quality reads and primer sequences were removed using Cutadapt<sup>34</sup> and paired-end reads were merged using PANDAseq<sup>35</sup>. Operational taxonomic units (OTUs) were then generated using AbundantOTU+ based on ≥97% similarity of sequences and classified taxonomically using the Ribosomal Database Project Classifier and the Greengenes database<sup>36–38</sup>. No samples were excluded based on our *a priori* requirement cut-off of >3500 reads. Bacterial OTU tables were generated by the removal of singletons (sequences with single reads) and non-bacterial OTU’s (Supplementary Table 1). Singletons were removed in an effort to more accurately evaluate community diversity by reducing the impact of base substitutions, low-quality reads, variable read lengths, non-target amplification, and undetected chimeric sequences<sup>39,40</sup>. The removal of singletons, while debatable, has been shown to reduce alpha-diversity but does significantly alter beta-diversity and the removal of singletons likely reflects a more accurate community<sup>39</sup>.

**Microbial Community and Statistical Analysis.** Analysis of the resulting OTU tables was performed in R (V 3.99) using the phyloseq package<sup>41</sup>. For alpha-diversity and beta-diversity calculations, we included OTUs that were present in  $\geq 1\%$  of total relative abundance. Alpha-diversity was calculated using Observed OTUs as measure of richness, and Shannon and Simpson's diversity indices which include richness and evenness were all calculated<sup>6,42,43</sup>. Unless otherwise stated, all analysis was performed without the use of rarefaction as it is suggested to avoid introducing false positives<sup>44</sup>. However, a secondary analysis of alpha-diversity was performed after rarefying the samples to a read depth of 15,000. This analysis was not included within the manuscript unless a discrepancy with the non-rarefied data was observed. If discrepancies arose, both the non-rarefied and rarefied data are presented. Alpha diversity indices were compared using Wilcoxon Signed Rank for paired non-parametric factors, and Mann-Whitney U test for unpaired non-normally distributed variables. Beta-diversity differences in the microbiome were analyzed using Bray-Curtis distances and visualized using principal coordinate analysis. Community-wide OTU level differences were assessed using the permutational multivariate analysis of variance (PERMANOVA) using distance matrices using 999 permutations following proportional normalization of the input data<sup>13,45</sup>. If community-wide differences were identified, OTUs belonging to the same genus were analysed using the DeSeq2 package using the test = "wald" and fitType = "parametric" settings<sup>46</sup>. For multiple testing, we utilized the Benjamin Hochberg multiple test correction. Log abundance plots were visualized using ggplot2<sup>47</sup>. A generalized linear model was constructed in R to examine the relationship between microbial diversity and repeated lung function measurements. Each analyzed variable was tested independently. The correlation between the abundance of individual genera and FEV<sub>1</sub> was examined using the Spearman rank correlation coefficient ( $r_s$ ). Analysis of clinical characteristic and dichotomous variables was done using the Chi-Squared test and Fischer's exact test using Prism 5.0 (GraphPad). Asymmetrically distributed variables were represented using median and interquartile ranges (IQR).

## Data Availability

The data from which this study derives is available for analysis as per Scientific Reports requirements (PRJNA514329).

## References

- Weycker, D., Edelsberg, J., Oster, G. & Tino, G. Prevalence and economic Burden of Bronchiectasis. *Clin. Pulm. Med.* **12**, 205–209 (2005).
- Davies, G., Wells, A. U., Doffman, S., Watanabe, S. & Wilson, R. The effect of *Pseudomonas aeruginosa* on pulmonary function in patients with bronchiectasis. *Eur. Respir. J.* **28**, 974–9 (2006).
- Ho, P. L. *et al.* The effect of *Pseudomonas aeruginosa* infection on clinical parameters in steady-state bronchiectasis. *Chest* **114**, 1594–1598 (1998).
- Purcell, P. *et al.* Polymicrobial airway bacterial communities in adult bronchiectasis patients. *BMC Microbiol.* **14**, 130 (2014).
- Sibley, C. D. *et al.* Culture enriched molecular profiling of the cystic fibrosis airway microbiome. *PLoS One* **6**, 25–27 (2011).
- Rogers, G. B. *et al.* Clinical measures of disease in adult non-CF bronchiectasis correlate with airway microbiota composition. *Thorax* **68**, 731–7 (2013).
- Green, H. & Jones, A. M. The Microbiome and Emerging Pathogens in Cystic Fibrosis and Non-Cystic Fibrosis Bronchiectasis. *Semin. Respir. Crit. Care Med.* **36**, 225–235 (2015).
- Rogers, G. B., van der Gast, C. J. & Serisier, D. J. Predominant pathogen competition and core microbiota divergence in chronic airway infection. *ISME J.* **9**, 217–225 (2015).
- van der Gast, C. J. *et al.* Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. *ISME J.* **5**, 780–791 (2011).
- Cox, M. J. *et al.* Airway Microbiota and Pathogen Abundance in Age-Stratified Cystic Fibrosis Patients. *PLoS One* **5**, e11044 (2010).
- Coburn, B. *et al.* Lung microbiota across age and disease stage in cystic fibrosis. *Sci. Rep.* **5**, 10241 (2015).
- Carmody, L. A. *et al.* The daily dynamics of cystic fibrosis airway microbiota during clinical stability and at exacerbation. *Microbiome* **3**, 12 (2015).
- Acosta, N. *et al.* The Evolving Cystic Fibrosis Microbiome: A Comparative Cohort Study Spanning Sixteen Years. *Ann. Am. Thorac. Soc.* **14**, 1288–1297 (2017).
- Acosta, N. *et al.* Sputum microbiota is predictive of long-term clinical outcomes in young adults with cystic fibrosis. *Thorax* **73**, 1016–1025 (2018).
- Cox, M. J. *et al.* Longitudinal assessment of sputum microbiome by sequencing of the 16S rRNA gene in non-CF bronchiectasis patients. *PLoS One* **12**, e0170622 (2017).
- Zhao, J. *et al.* Decade-long bacterial community dynamics in cystic fibrosis airways. *Proc. Natl. Acad. Sci. USA* **109**, 5809–14 (2012).
- Rogers, G. B. *et al.* A Novel Microbiota Stratification System Predicts Future Exacerbations in Bronchiectasis. *Ann. Am. Thorac. Soc.* **11**, 496–503 (2014).
- Paganin, P. *et al.* Changes in cystic fibrosis airway microbial community associated with a severe decline in lung function. *PLoS One* **10**, 1–19 (2015).
- Tunney, M. M. *et al.* Lung microbiota and bacterial abundance in patients with bronchiectasis when clinically stable and during exacerbation. *Am. J. Respir. Crit. Care Med.* **187**, 1118–1126 (2013).
- Rogers, G. B., Bruce, K. D., Martin, M. L., Burr, L. D. & Serisier, D. J. The effect of long-term macrolide treatment on respiratory microbiota composition in non-cystic fibrosis bronchiectasis: an analysis from the randomised, double-blind, placebo-controlled BLESS trial. *Lancet. Respir. Med.* **2**, 988–996 (2014).
- Whelan, F. J. *et al.* Longitudinal sampling of the lung microbiota in individuals with cystic fibrosis. *PLoS One* **12**, e0172811 (2017).
- Carmody, L. A. *et al.* Changes in cystic fibrosis airway microbiota at pulmonary exacerbation. *Ann. Am. Thorac. Soc.* **10**, 179–187 (2013).
- Collie, D. *et al.* Lung Microbiota Changes Associated with Chronic *Pseudomonas aeruginosa* Lung Infection and the Impact of Intravenous Colistimethate Sodium. *PLoS One* **10**, e0142097 (2015).
- Chalmers, J. D. *et al.* The EMBARC European Bronchiectasis Registry: protocol for an international observational study. *ERJ Open Res.* **2**, 00081–2015 (2016).
- Fardon, T. *et al.* Identifying Modifiable Risk Factors for Rapid Lung Function Decline in Bronchiectasis. In *American journal of respiratory and critical care medicine* A2878 (2016).
- Martínez-García, M. A. Factors Associated With Lung Function Decline in Adult Patients With Stable Non-Cystic Fibrosis Bronchiectasis. *Chest* **132**, 1565 (2007).
- Smith, D. J. *et al.* Pyrosequencing reveals transient cystic fibrosis lung microbiome changes with intravenous antibiotics. *Eur. Respir. J.* **44**, 922–930 (2014).
- Stressmann, F. A. *et al.* Long-term cultivation-independent microbial diversity analysis demonstrates that bacterial communities infecting the adult cystic fibrosis lung show stability and resilience. *Thorax* **67**, 867–873 (2012).

29. Nielsen, S. *et al.* Disrupted progression of the intestinal microbiota with age in children with cystic fibrosis. *Sci. Rep.* **6**, 1–11 (2016).
30. Thevaranjan, N. *et al.* Streptococcus pneumoniae colonization disrupts the microbial community within the upper respiratory tract of aging mice. *Infect. Immun.* **84**, IAI.01275-15 (2016).
31. Whelan, F. J. *et al.* The loss of topography in the microbial communities of the upper respiratory tract in the elderly. *Ann. Am. Thorac. Soc.* **11**, 513–521 (2014).
32. Bartram, A. K., Lynch, M. D. J., Stearns, J. C., Moreno-Hagelsieb, G. & Neufeld, J. D. Generation of Multimillion-Sequence 16S rRNA Gene Libraries from Complex Microbial Communities by Assembling Paired-End Illumina Reads. *Appl. Environ. Microbiol.* **77**, 3846–3852 (2011).
33. Whelan, F. J. & Surette, M. G. A comprehensive evaluation of the sl1p pipeline for 16S rRNA gene sequencing analysis. *Microbiome* **5**, 100 (2017).
34. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
35. Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G. & Neufeld, J. D. PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* **13**, 31 (2012).
36. DeSantis, T. Z. *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72**, 5069–5072 (2006).
37. Ye, Y. Identification and Quantification of Abundant Species from Pyrosequences of 16S rRNA by Consensus Alignment. *Proceedings. (IEEE. Int. Conf. Bioinformatics Biomed)*. **2010**, 153–157 (2011).
38. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–7 (2007).
39. Auer, L., Mariadassou, M., O'Donohue, M., Klopp, C. & Hernandez-Raquet, G. Analysis of large 16S rRNA Illumina data sets: Impact of singleton read filtering on microbial community description. *Mol. Ecol. Resour.* **17**, e122–e132 (2017).
40. Schirmer, M. *et al.* Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res.* **43**, e37 (2015).
41. McMurdie, P. J. & Holmes, S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* **8**, e61217 (2013).
42. Rogers, G. B. *et al.* Reducing bias in bacterial community analysis of lower respiratory infections. *ISME J.* **7**, 697–706 (2013).
43. Gihring, T. M., Green, S. J. & Schadt, C. W. Massively parallel rRNA gene sequencing exacerbates the potential for biased community diversity comparisons due to variable library sizes. *Environ. Microbiol.* **14**, 285–290 (2012).
44. McMurdie, P. J. & Holmes, S. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Comput. Biol.* **10**, e1003531 (2014).
45. Anderson, M. A new method for non parametric multivariate analysis of variance. *Austral Ecol.* **26**, 32–46 (2001).
46. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
47. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag, New York, 2009).

## Acknowledgements

We would like to thank our clinical partners at the Calgary Bronchiectasis Clinic and Calgary Laboratory Services for their assistance in procuring and maintaining samples in the biobank. This study was funded by a grant from the Lung Association of Alberta and Northwest Territories and a grant from the Natural Sciences and Engineering Research Council of Canada (RGPIN 203595-2013). T.E.W. was supported by a scholarship from Natural Sciences and Engineering Research Council of Canada and Alberta Innovates Technology Futures.

## Author Contributions

T.E.W. was responsible for the collection strains. T.E.W. and A.H. performed the genomic extractions and DNA amplification. T.E.W., R.L., H.R.R. and C.H.M. were responsible for patient data collection. T.E.W., N.A., M.G.S., R.S., M.D.P. and C.S. contributed to the statistical analysis. All authors contributed to the preparation of the manuscript. The Calgary Non-CF Bronchiectasis Biobank is maintained by H.R.R. and M.D.P. D.G.S. and M.D.P. envisioned the project together and serve as guarantors of the work.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-42862-y>.

**Competing Interests:** T.E.W., R.L., A.A.H., N.A., C.H.M., R.S. and C.D. declare no competing interests. M.D.P., H.R.R. and M.G.S. have received research support from Gilead Sciences.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019