4-23-2019

Direct Lung Sampling Indicates That Established Pathogens Dominate Early Infections in Children with Cystic Fibrosis.

Peter Jorth

Zarmina Ehsan MD
Children's Mercy Hospital

Amir Rezayat

Ellen Caldwell

Christopher Pope

See next page for additional authors

Recommended Citation

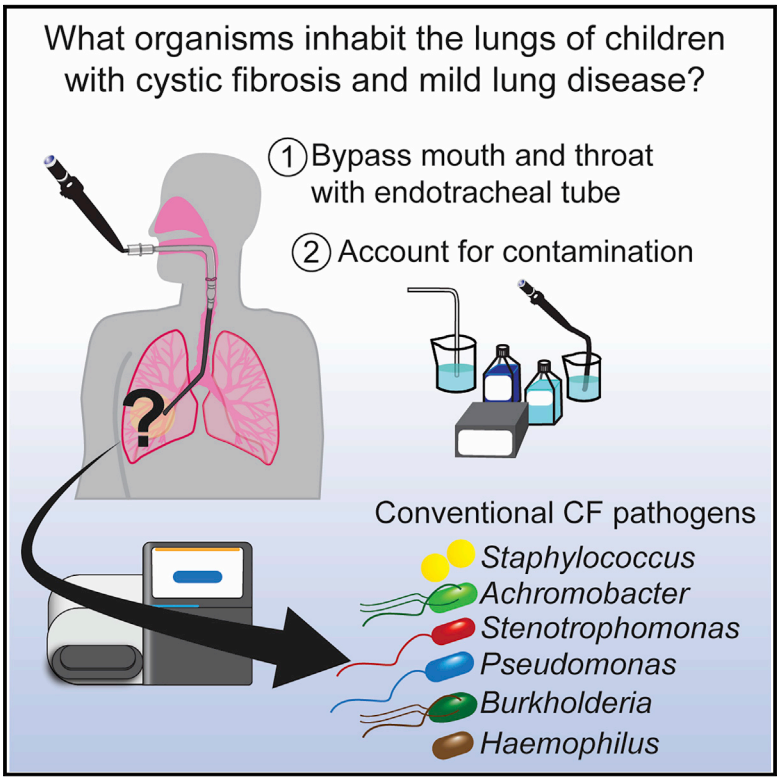
Jorth P, Ehsan Z, Rezayat A, et al. Direct Lung Sampling Indicates That Established Pathogens Dominate Early Infections in Children with Cystic Fibrosis. Cell Rep. 2019;27(4):1190-1204.e3. doi:10.1016/j.celrep.2019.03.086

Article

Cell Reports

Direct Lung Sampling Indicates That Established Pathogens Dominate Early Infections in Children with Cystic Fibrosis

Graphical Abstract



Highlights

- BALs from CF children with mild disease primarily contain established CF pathogens
- Microbiota in BALs lacking CF pathogens were indistinguishable from controls
- Rare oral organisms were found in some BALs containing abundant CF pathogens
- CF pathogens strongly correlated with inflammation and oral organisms did not

Authors

Peter Jorth, Zarmina Ehsan, Amir Rezayat, ..., Dan Benscoter, John P. Clancy, Pradeep K. Singh

Correspondence

john.clancy@cchmc.org (J.P.C.), singhpr@uw.edu (P.K.S.)

In Brief

Jorth et al. studied lung lavages from CF children using sequence-based methods and extensive controls for upper airway and reagent contamination. The results argue against the CF infection pathogenesis hypothesis postulating that diverse communities of mostly oral bacteria inhabit lungs early in disease and suggest that differences in the samples' microbial diversity were driven by changes in the abundance of conventional CF pathogens against a fairly constant background of contaminants.

Jorth et al., 2019, Cell Reports *27*, 1190–1204
April 23, 2019 © 2019
https://doi.org/10.1016/j.celrep.2019.03.086

CellPress

OPEN ACCESS
CellPress

Cell Reports
Article

Direct Lung Sampling Indicates That Established Pathogens Dominate Early Infections in Children with Cystic Fibrosis

Peter Jorth,[1,2] Zarmina Ehsan,[3,8] Amir Rezayat,[2,8] Ellen Caldwell,[4] Christopher Pope,[5] John J. Brewington,[6,7] Christopher H. Goss,[4] Dan Benscoter,[6,7] John P. Clancy,[6,7,9,*] and Pradeep K. Singh[2,9,10,*]

[1]Departments of Pathology and Laboratory Medicine, Medicine, and Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA
[2]Department of Microbiology and Medicine, University of Washington School of Medicine, Seattle, WA 98195, USA
[3]Pulmonology and Sleep Medicine, Children's Mercy-Kansas City, Kansas City, MO 64108, USA
[4]Department of Medicine, University of Washington School of Medicine, Seattle, WA 98195, USA
[5]Department of Pediatrics, University of Washington School of Medicine, Seattle, WA 98195, USA
[6]Division of Pulmonary Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA
[7]Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH 45229, USA
[8]These authors contributed equally
[9]These authors contributed equally
[10]Lead Contact
*Correspondence: john.clancy@cchmc.org (J.P.C.), singhpr@uw.edu (P.K.S.)
https://doi.org/10.1016/j.celrep.2019.03.086

SUMMARY

Culture and sequencing have produced divergent hypotheses about cystic fibrosis (CF) lung infections. Culturing suggests that CF lungs are uninfected before colonization by a limited group of CF pathogens. Sequencing suggests diverse communities of mostly oral bacteria inhabit lungs early on and diversity decreases as disease progresses. We studied the lung microbiota of CF children using bronchoscopy and sequencing, with measures to reduce contamination. We found no evidence for oral bacterial communities in lung lavages that lacked CF pathogens. Lavage microbial diversity varied widely, but decreases in diversity appeared to be driven by increased CF pathogen abundance, which reduced the signal from contaminants. *Streptococcus*, *Prevotella*, and *Veillonella* DNA was detected in some lavages containing CF pathogens, but DNA from these organisms was vastly exceeded by CF pathogen DNA and was not associated with inflammation. These findings support the hypothesis that established CF pathogens are primarily responsible for CF lung infections.

INTRODUCTION

People with cystic fibrosis (CF) suffer from chronic bacterial lung infections that culminate in respiratory failure (Burns et al., 2001; Emerson et al., 2002; Lyczak et al., 2002). Despite intensive research, key questions about CF infection pathogenesis remain obscure. One fundamental question is which bacterial species are present in the lungs of people with CF and which cause disease. Addressing this question is critical to devising effective disease prevention and treatment approaches.

Decades of study using culture-based methods suggest that the lungs of people with CF are uninfected at birth and that over time transient and then permanent infections develop (Burns et al., 2001). Importantly, a restricted group of organisms (called "CF pathogens" below) have been thought responsible for infection. The list of CF pathogens include *Haemophilus influenzae*, *Staphylococcus aureus*, *Pseudomonas aeruginosa*, *Burkholderia* spp., and a few others (Gibson et al., 2003; Lyczak et al., 2002); epidemiological and clinical observations associate these organisms with disease (Emerson et al., 2002; Gibson et al., 2003).

Recent work using DNA sequencing to identify bacteria in CF airways has produced several new ideas that have challenged this established model. First, DNA-based methods suggest that diverse communities of bacteria that were not previously thought to cause CF infections (a group we call "non-conventional organisms" below) dominate lungs before CF pathogens appear, when lung function is generally preserved (Brown et al., 2014; Coburn et al., 2015; Frayman et al., 2017; Madan et al., 2012; Muhlebach et al., 2018; Pittman et al., 2017). This finding raises the possibility that non-conventional organisms could protect lungs from pathogen infection. Alternatively, these organisms could produce initial lung injury that promotes pathogen colonization.

Second, although studies using DNA-based methods generally find that CF pathogens dominate lung microbiota in advanced disease (Coburn et al., 2015; Goddard et al., 2012; Zhao et al., 2012), sequence-based analyses indicate that diverse communities of non-conventional organisms co-exist with CF pathogens after CF pathogens become dominant (Coburn et al., 2015; Zhao et al., 2012). This finding raises the possibility that non-conventional organisms could either enhance or dampen pathogen-induced injury or treatment responses,or independently contribute to disease.

Finally, DNA-based analyses suggest that the diversity of lung microbiota may be a key disease parameter. Many studies have found that the relative abundance of non-conventional organisms and the diversity of lung microbiota in airway samples decrease as age and disease severity advances (Coburn et al., 2015; Cox et al., 2010; Fodor et al., 2012; Klepac-Ceraj et al., 2010; Stokell et al., 2015; Zhao et al., 2012). This finding raises the possibility that declining diversity could contribute to lung disease and strategies that maintain diversity could be beneficial.

Reconciling the new ideas suggested by DNA-based methods with the established pathogenesis model is challenging, as both culture- and DNA-based approaches have significant and in some ways opposing limitations. Culture analysis is based on the premise that infection is generally caused by a single or small number of pathogens that can be isolated in pure culture (Baron et al., 2013; Burns and Rolain, 2014). Thus, clinical laboratories selectively culture samples and filter results to remove organisms not considered pathogenic (Tunney et al., 2008), which are viewed as contaminants (Baron et al., 2013; Mahboubi et al., 2016). This could be problematic, as culture-based analysis could fail to identify non-conventional organisms.

In contrast, DNA-based methods cast a wider net, identifying culturable and unculturable organisms and DNA from live and dead cells without filtering for contaminants (Kennedy et al., 2014; Salter et al., 2014; Weiss et al., 2014). This could be problematic because many of the non-conventional organisms detected by sequencing are highly abundant in oropharyngeal secretions that contaminate airway samples (Bik et al., 2010; Coburn et al., 2015; Cox et al., 2010; Dewhirst et al., 2010; Fodor et al., 2012; Goddard et al., 2012; Klepac-Ceraj et al., 2010; Lim et al., 2013; Price et al., 2013; Twomey et al., 2013; Zemanick et al., 2015; Zhao et al., 2012). Contamination is particularly problematic for throat and sputum samples and when lung bacterial biomass is low (e.g., early in disease), as contaminants will have more consequential effects on measurements (Coburn et al., 2015; Salter et al., 2014). DNA-based methods could also detect bacteria that are inhaled or aspirated but killed by host defenses before they replicate. Thus, DNA-based methods will detect contaminants and dead organisms that do not actively grow in the lung.

Here, we sought to investigate the composition of lung microbiota in CF children by combining the high sensitivity of sequence-based approaches and directed bronchoscopic sampling, with measures to reduce exposure to oropharyngeal secretions and extensive controls to help identify contaminants. We focused on CF children, as the results of culture- and DNA-based methods differ most in early lung disease and understanding early infection is critical to intervention strategies to preserve lung function.

RESULTS

Subjects

We studied 22 children and young adults (mean, 13.1 years old; range, 6–21 years old) all with mild CF lung disease (mean forced expiratory volume [FEV_1], 92% ± 15% predicted) who underwent bronchoscopy for clinical indications. None of the subjects were taking cystic fibrosis transmembrane conductance regulator (CFTR) modulators. Table S1 provides additional information about study subjects.

Sampling Approach and Controls

Reducing and controlling for contamination are particularly critical when using DNA-based methods due to their sensitivity and because results are not filtered by clinical laboratories that disregard and remove organisms considered to be contaminants. Bronchoalveolar lavage (BAL) samples may become contaminated with bacterial DNA from the oropharyngeal compartment, secretions aspirated into the lung during the procedure, and non-patient sources, including the bronchoscope and reagents (Culver et al., 2003; Kennedy et al., 2014; Kirschke et al., 2003; Mehta and Minai, 1999; Salter et al., 2014; Srinivasan et al., 2003; Weiss et al., 2014).

We eliminated scope contact with oropharyngeal mucosa and likely reduced procedure-associated aspiration by performing bronchoscopy through cuffed endotracheal tubes placed during a separate laryngoscopy procedure (Figure 1A); however, some upper airway secretions likely entered the endotracheal tube during placement. We also collected mouth rinses and throat aspirates by using a sterile suction catheter immediately before bronchoscopy from each subject for comparison (Figure 1B).

Using these methods, we performed BAL of 4 lobar bronchi in each subject (with the scope in the wedged position, see STAR Methods), beginning with the right middle lobe from which 2 saline lavages were separately collected (BAL-1a and b). Lavage was subsequently performed in an adjacent segmental bronchus in the right upper lobe (BAL-2), then the left lower lobe (BAL-3), and finally in the lung segment that appeared most diseased during bronchoscopy or on prior imaging (BAL-4) (Figure 1A).

DNA contamination contributed by reagents and the suction channels of bronchoscopes and catheters (which were used for throat aspirates) are difficult to reduce. Thus, we measured bacterial DNA in saline washes (saline from same source that was used for BALs) from each bronchoscope and catheter, immediately before bronchoscopies and throat aspirates were performed (called "scope or catheter pre-washes" or collectively called "instrument controls," below) as background measurements of contamination from non-patient sources (Figure 1B).

Measuring Background Contamination in BALs

Scope pre-washes consistently contained $\sim 10^2$ bacterial genome copies/mL (Figures S1A and S1B; Table S2) (mean, 293; range, 43–651), similar to instrument and reagent controls in other studies (Table S3). Sequencing revealed a diverse collection of microbial DNA in scope pre-washes, including CF pathogens (see Figure 1C for genera considered CF pathogens) (Figure S1C; Tables S4 for taxa and S5 for sequencing read counts and Good's coverage for each sample). The fact that DNA-based analysis of instrument control samples identified CF pathogens (see below) prior to contact with subjects highlights one challenge in interpreting sequencing results.

Two findings indicated that most bacterial DNA in scope washes originated from processing reagents, rather than the

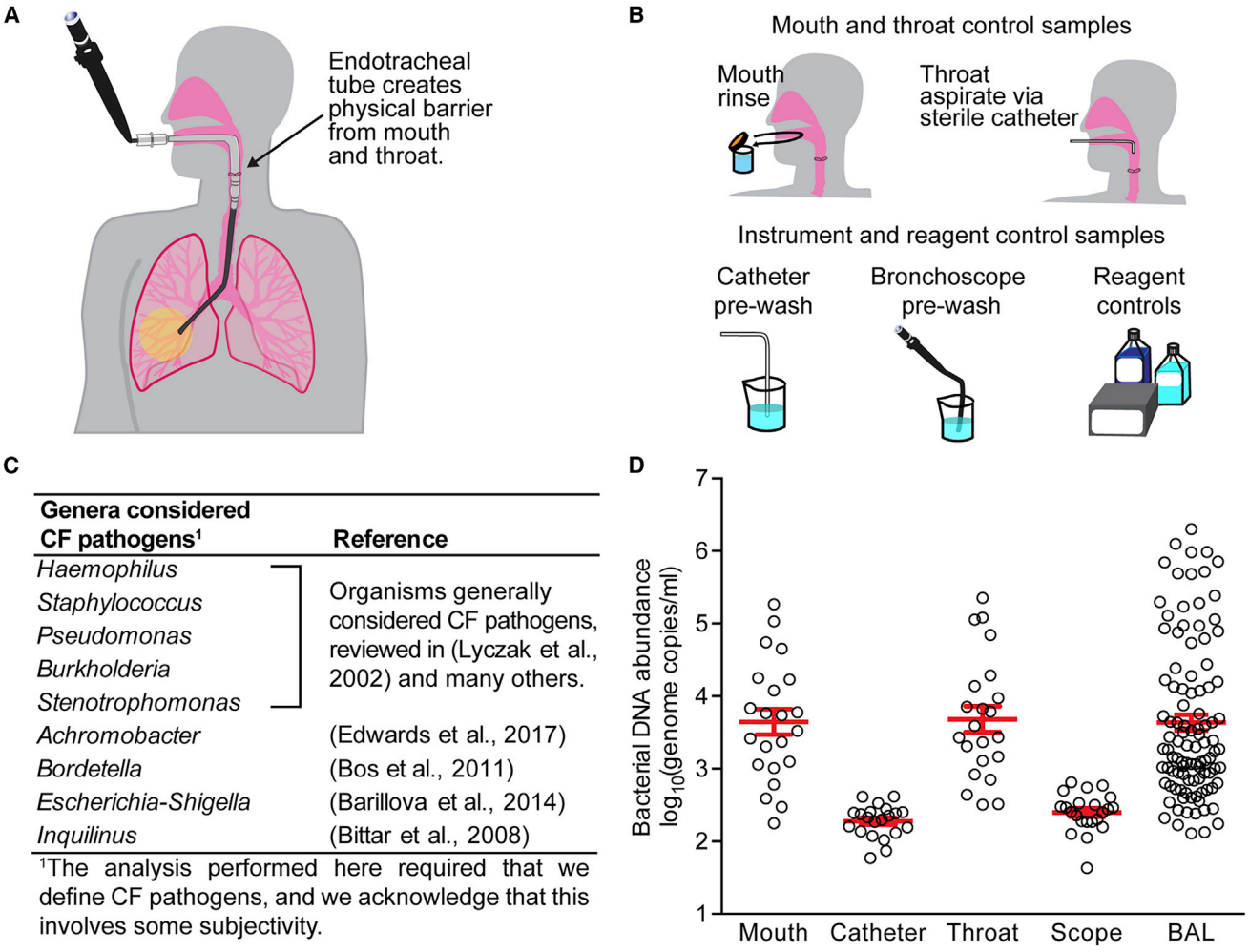


Figure 1. Study Design and Bacterial Burdens in Samples

(A) Endotracheal tubes eliminated bronchoscope contact with oropharyngeal (OP) tissues.

(B) Control samples included mouth rinses, throat aspirates, saline passed through each sterile catheter (used for throat aspirates) and bronchoscope prior to use (called scope and catheter pre-washes), and reagent controls.

(C) Organisms considered CF pathogens in this study.

(D) Total bacterial abundance detected in mouth rinse, catheter pre-wash, throat aspirate, scope pre-wash samples (n = 22 for each), and BALs (n = 102 as ~4 BAL samples were obtained for most subjects). Line indicates mean, and error bars standard deviation values. Here and below, qPCR data are the mean of 2 technical replicates.

See also Figures S1 and S2 and Tables S1, S2, and S3.

scope channel. Bacterial DNA abundances in scope and catheter pre-washes and in saline used for controls (and subsequent BALs) that was not passed through any device were similar (Figure S1A). In addition, the microbiota profiles of scope and catheter pre-washes exhibited different patterns depending on the reagent batches used for processing sequencing libraries (Figure S1C).

We tested the hypothesis that scope pre-washes would represent the contaminating DNA in subsequently obtained BALs by serially passing 10-mL saline aliquots through 3 bronchoscopes used in this study and found bacterial DNA abundance essentially unchanged in serial samples (Figure S1B). This finding indicates that the background DNA signal (likely derived from reagents, see Figure S1C) would also likely be detectable in BALs. We accounted for this contamination in the analyses of BALs described below.

BALs from CF Children Vary Widely in Bacterial DNA Abundance

We measured the total bacterial burden in all mouth, throat, BAL, and control samples. BAL bacterial DNA concentrations ranged from ~10^2 genome copies/mL (levels seen in scope pre-wash controls) to >10^6 genome copies/mL, which was higher than any other sample type (Figure 1D; Table S2). Notably, the bacterial concentration in serial BAL samples taken from the same subject sometimes differed by an order of magnitude (Figure S2; Table

S2), and BAL-4 (obtained from the lung segment that appeared most diseased) generally contained the most bacterial DNA.

BAL Samples with More Bacteria Are Less Diverse

DNA-based analyses suggest that microbial diversity may be a key disease-associated parameter, so we measured the Shannon diversity index in BALs and tried to understand differences in diversity among the samples. We focused first on the initially obtained BAL-1a samples, as they are unaffected by DNA carry-over from other lung sites. Thus, the difference between the bacterial DNA content in each paired scope pre-wash and the corresponding BAL-1a will reflect bacterial DNA recovered from the targeted lung segment. We plotted the Shannon index of each BAL-1a as a function of the DNA content difference between each BAL-1a and its paired scope pre-wash and found a strong inverse relationship; as bacterial DNA recovery from the lung increased, Shannon diversity decreased (Figure 2A; Pearson $r = -0.87$, $p < 0.0001$).

We extended this analysis to include BALs from all sites (i.e., BALs 1–4), in this case examining Shannon diversity as a function of total bacterial DNA concentration, and again found a strong negative relationship (Figure 2B; Pearson $r = -0.80$, $p < 0.001$). Notably, the negative association between Shannon diversity and bacterial DNA abundance was absent for oral and throat samples (Figure 2C; Pearson $r = -0.30$, $p = 0.18$; and Figure S3; Pearson $r = 0.05$, $p = 0.81$), which are sites known to harbor diverse resident microbiota (Dewhirst et al., 2010).

Increases in Pathogen Abundance Drive Diversity Declines

We investigated the contributions of CF pathogens (defined in Figure 1C) and non-conventional organisms to the Shannon diversity index variation of BALs by using both taxa absolute and relative abundance measurements. Absolute abundance of individual taxa was calculated by multiplying total bacterial abundance by taxa relative abundance (see experiments validating this approach in Figure S4). As shown in Figure 2D, reduced BAL Shannon diversity was associated with marked increases in the absolute abundance of CF pathogens ($r = -0.87$, $p < 0.001$, Pearson correlation). In contrast, the absolute abundance of non-conventional organisms was relatively constant over the wide range of BAL Shannon diversity observed (Figure 2E; $r = -0.21$, $p = 0.03$, Pearson correlation).

The dominant effect of CF pathogen absolute abundance in driving reduced diversity was also apparent by comparing the BALs with the highest and lowest extremes of Shannon diversity. As shown in Figure 2F, the lowest diversity BALs contained a much higher absolute abundance of CF pathogens than the highest diversity BALs, but the absolute abundance of non-conventional organisms did not significantly differ between the groups (Figure 2F).

These data suggest that the low diversity of BALs with high bacterial burdens is driven by increases in the absolute abundance of CF pathogens against a relatively static background of non-conventional organisms. Notably, plotting the data in terms of relative abundance (Figures 2G and 2H) completely obscured the differential contribution of CF pathogens and non-conventional organisms to BAL Shannon diversity (see Discussion).

The Non-conventional Organisms in BALs Are Similar to Instrument Pre-wash Controls

We focused further analysis on the non-conventional organisms in BALs because the identity and origins of these organisms are a key area of controversy in CF (O'Toole, 2018). Previous work suggests that many non-conventional organisms in CF airway samples are oral organisms (Bik et al., 2010; Coburn et al., 2015; Cox et al., 2010; Dewhirst et al., 2010; Fodor et al., 2012; Goddard et al., 2012; Klepac-Ceraj et al., 2010; Lim et al., 2013; Price et al., 2013; Twomey et al., 2013; Zemanick et al., 2015; Zhao et al., 2012). Some studies suggest that these organisms inhabit CF lungs, whereas other studies raise the possibility that they are contaminants introduced from oropharyngeal secretions. However, relatively few studies have used bronchoscopy sampling performed through endotracheal tubes placed during a separate laryngoscopy procedure as done here, and we hoped that this approach would limit oropharyngeal contamination (see above).

We began by ordering the BALs by their total bacterial DNA abundance (Figure 3A) and found that non-conventional organisms generally dominated BALs with low bacterial abundance, whereas CF pathogens dominated BALs with high abundance (Figure 3B). We then examined the similarity of each BAL to subject-paired mouth and throat samples using the Morisita-Horn dissimilarity index (Wolda, 1981).

As expected, the high biomass BALs that were dominated by CF pathogens (Figure 3B, right side of graph) were very dissimilar to their paired mouth and throat samples (Figures 3C, 3D, and 4). This finding is consistent with data indicating that CF pathogens are not abundant in the mouth and throat (Figure S5; Table S4). Surprisingly, the low bacterial biomass BALs that were dominated by non-conventional organisms (Figures 3B, left side of graph) were also very dissimilar to subject-paired mouth and throat samples (Figures 3C, 3D, and S5; Table S4).

The dissimilarity of low bacterial biomass BALs to mouth and throat samples led us to hypothesize that contaminants contributed by reagents and instruments may account for the non-conventional organism DNA in these samples. We tested this hypothesis in three ways. First, we compared BALs to their paired scope pre-wash controls and found the BALs with low bacterial DNA abundance to be more similar to subject-paired scope pre-washes than to paired mouth and throat samples (compare Figures 3E to 3C and 3D).

Second, we plotted the relative abundance of taxa in BALs as a function of their abundance in instrument controls. Taxa that were instrument or reagent contaminants would likely show a linear abundance relationship in these two sample types, whereas taxa originating from subjects' lungs should be more abundant in BALs relative to instrument controls. As shown in Figure 3F, the relative abundance of non-conventional taxa (black points) in instrument controls and BALs exhibited a strong linear relationship. In contrast, CF pathogens (colored points) were more abundant in BALs than in instrument controls.

Third, we computationally removed CF pathogens from the results to increase the resolution of the non-conventional organisms present and compared BALs to mouth, throat, and control samples. Regardless of total bacterial DNA content, the taxonomic profiles of non-conventional organisms in

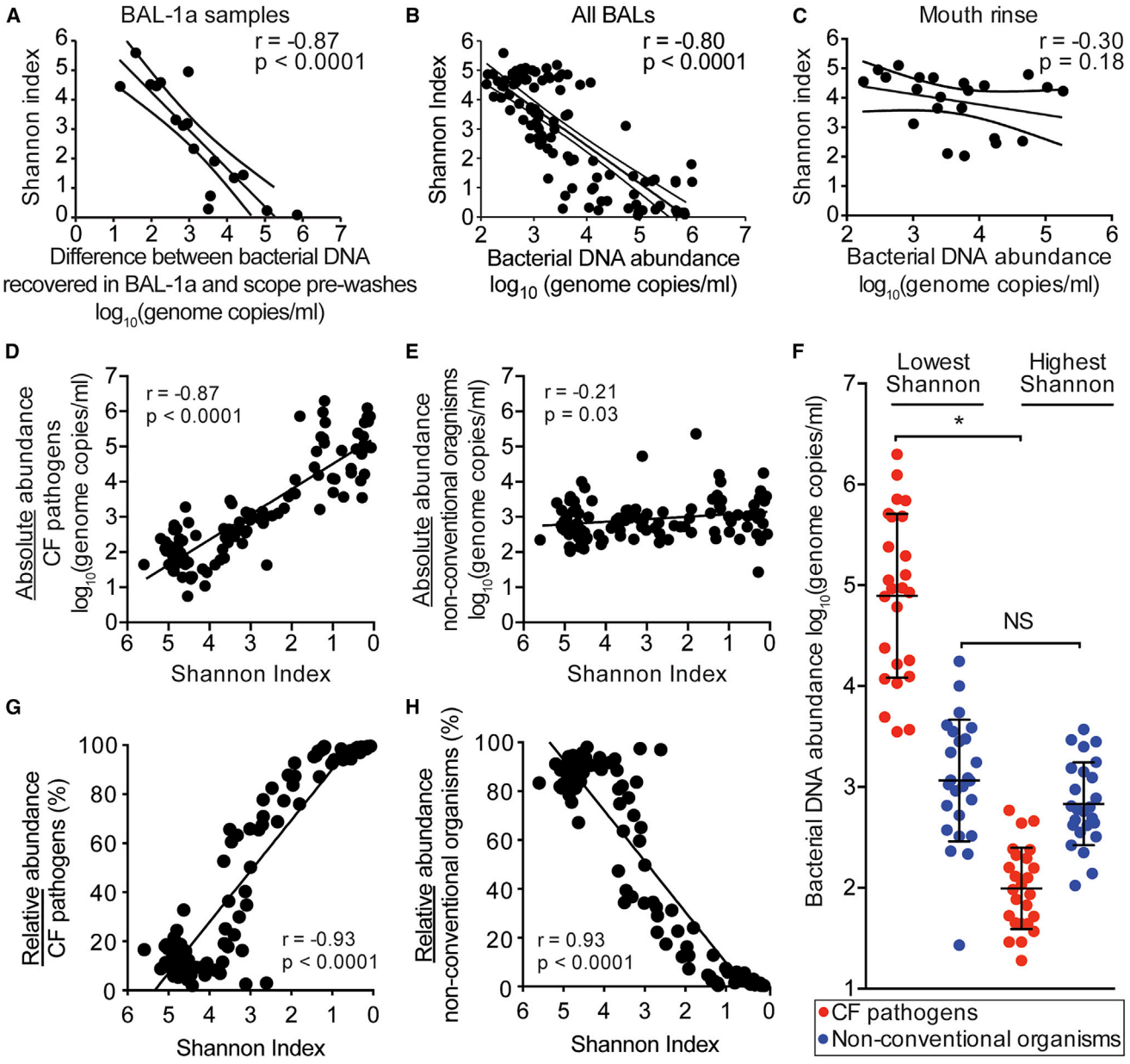


Figure 2. Changes in the Absolute Abundance of CF Pathogens Drive BAL Shannon Diversity Differences

(A) Relationship between the Shannon diversity index and bacterial DNA recovered from the first BAL obtained from each subject (BAL-1a). Bacterial DNA recovery was measured by subtracting bacterial DNA levels in patient-paired scope pre-washes from the values in subsequent BAL-1a sample; n = 17 as samples from 5 subjects (i.e., 5, 6, 8, 17, and 18) were not included because BAL samples contained lower bacterial concentrations than paired scope pre-washes.

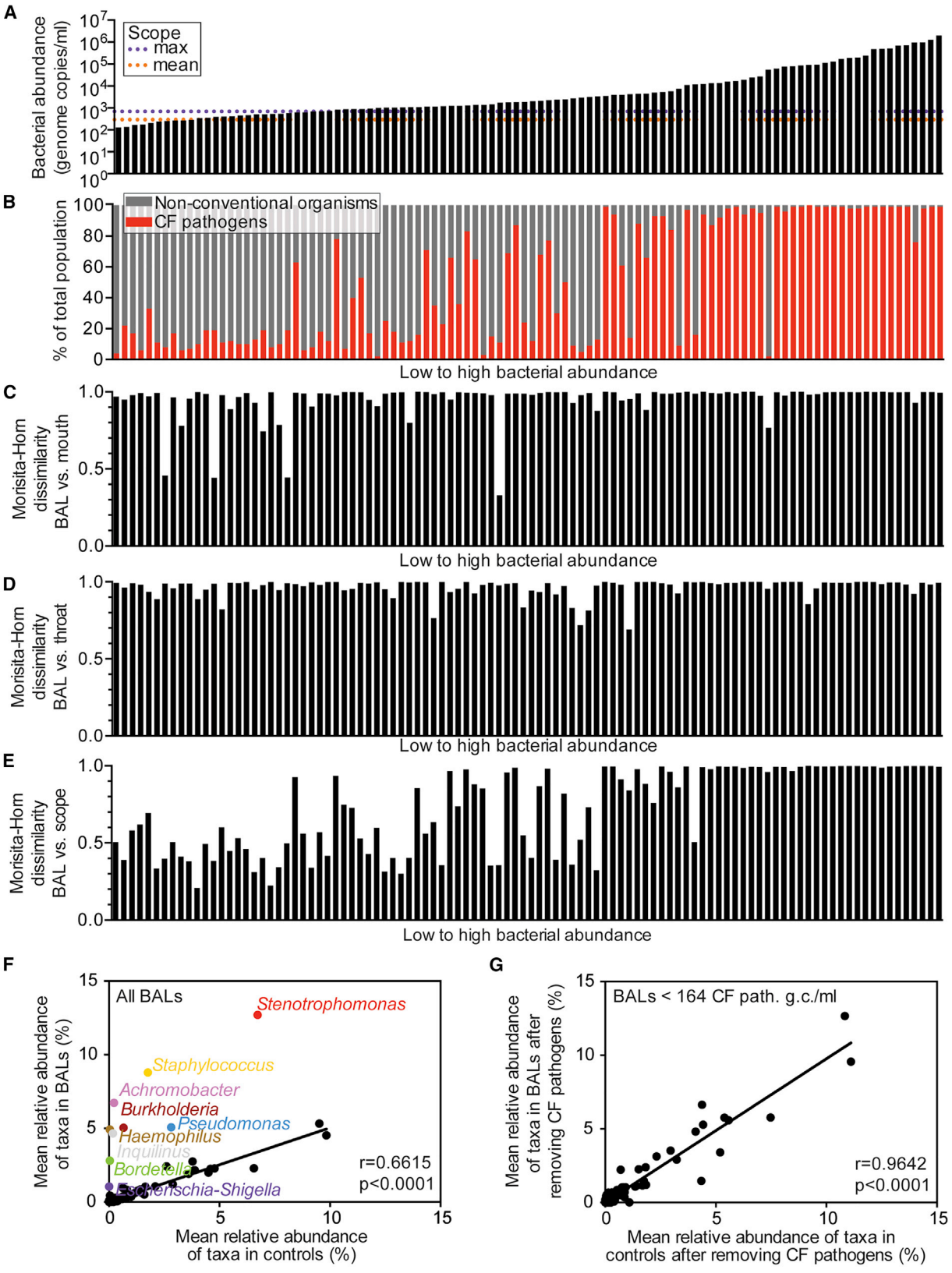
(B and C) Relationship between BAL Shannon diversity and total bacterial DNA levels in all 102 BALs (B) and 22 mouth rinses (C) collected (see Figure S3 for data from throat samples). Linear regression lines with 95% confidence intervals and Pearson correlation coefficients with p values are shown in (A)–(C).

(D and E) Relationship between the absolute abundance of CF pathogens (D) and non-conventional organisms (E) and Shannon diversity indices of BALs (n = 102). Linear regression lines and Pearson correlation coefficients with respective p values are shown.

(F) Absolute abundance of CF pathogens and non-conventional organisms in the 25 BALs with the lowest (left) and highest (right) Shannon diversity indices (mean values for each group are indicated in black bars). Error bars indicate SD values; *p < 0.005, unpaired two-tailed t test.

(G and H) Data from the 102 BALs presented in (D) and (E), but plotted in terms of relative rather than absolute abundance of CF pathogens (G) and non-conventional organisms (H). The relative abundance plots obscure the finding that CF pathogen abundance drives the Shannon diversity differences. Graphs include linear regression and Pearson correlation coefficients with respective p values. Here and below taxa abundance data obtained from sequencing are the mean of 3 technical replicates.

Also see Figures S3 and S4, Table S4 for taxa abundance data, and Table S5.



(legend on next page)

BALs were more similar to catheter and scope pre-wash controls than to mouth or throat samples (Figure 4A). Morisita-Horn dissimilarity and principal coordinates analyses corroborated these findings (Figures 4B and 4C). Furthermore, principal coordinates analyses showed that non-conventional organisms in BALs grouped with instrument control samples that were processed using the same reagent batches (Figure 4C). These findings are consistent with the strong reagent-batch effect on the microbiota identified in saline and instrument controls (Figure S1). Taken together, these findings suggest that most of non-conventional organisms in the BALs we studied were dissimilar to patient-paired oral samples and were likely contaminants contributed by processing reagents.

BALs Lacking CF Pathogens Did Not Contain Microbiota Distinguishable from Background

Although the similarity and clustering analyses above suggest that reagent contaminants contribute significantly to the non-conventional organisms in the BALs we collected, the findings do not rule out the possibility that some non-conventional organisms originate from the lungs. We used two approaches to examine this possibility further, focusing first on BALs that contained very low abundances of CF pathogens (less than 150% of the maximum pathogen DNA levels found in paired scope pre-washes, i.e., <164 CF pathogen genome copies/mL).

First, we plotted the relative abundance of non-conventional taxa found in these BALs as a function of their abundance in instrument controls. As noted above, taxa that originated from the subjects' lungs would be expected to be more abundant in BALs. As shown in Figure 3G, the abundance of non-conventional taxa in BALs and instrument controls exhibited a strong linear relationship, with values falling along the identity line.

We also used a statistical test (the negative binomial test [Love et al., 2014]) to identify non-conventional organisms over-represented in BALs, relative to paired instrument and reagent controls and found no taxa that were overrepresented in low pathogen abundance BALs as compared to scope pre-wash samples (Table 1) (see Table S6 for proof-of-principle control analyses). Thus, these analyses found no evidence for an identifiable microbiota that was distinct from instrument controls in BALs containing a low abundance of CF pathogens.

Some Pathogen-Infected BALs Contained DNA from Oral Bacteria

We used the statistical analysis described above to determine if any non-conventional organisms were overrepresented in BALs that did contain CF pathogens (i.e., BALs with >164 CF pathogen genome copies per mL, which is >150% of the maximum value in instrument and reagent controls). In contrast to the results from low pathogen abundance BALs, we found three non-conventional genera, *Streptococcus*, *Prevotella*, and *Veillonella*, overrepresented in BAL samples relative to controls (Table 1). In addition, an uncultured organism belonging to the Alcaligenaceae family was overrepresented. Alcaligenaceae includes *Achromobacter*, *Burkholderia*, and *Bordetella*, so reads assigned to Alcaligenaceae may have come from one of these CF pathogens. These findings suggest that *Streptococcus*, *Prevotella*, and *Veillonella* genera may be present at a low abundance in BALs that already contained CF pathogens.

Streptococcus, *Prevotella*, and *Veillonella* Are Vastly Outnumbered by Pathogens

Examination of the profiles of non-conventional organisms (with pathogens removed) (Figure 4A) suggested that *Streptococcus*, *Prevotella*, and *Veillonella* were more abundant in BALs that contained more bacterial DNA. To examine this further, we plotted the absolute abundance of *Streptococcus*, *Prevotella*, and *Veillonella* as a function of the total BAL bacterial DNA concentration (Figure 5A) and CF pathogen abundance (Figure 5B) and found a strong positive relationship. With the exception of 6 (out of 102) BALs that contained trace levels of *Streptococcus* DNA (~10 genome copies/mL), *Streptococcus*, *Prevotella*, and *Veillonella* were absent in BALs containing <1,000 bacterial genome copies/mL (Figure 5A) or <100 CF pathogen genome copies/mL (Figure 5B).

We examined the ratio of *Streptococcus*, *Prevotella*, and *Veillonella* to CF pathogens in BALs and found CF pathogens in far excess abundance. On average, there was 725-fold more pathogen DNA than *Streptococcus*, 2,334-fold more pathogen DNA than *Prevotella*, and 6,476-fold more pathogen DNA than *Veillonella* (Figure 5C). Thus, these three non-conventional organisms were almost always found in BALs that already contained abundant pathogens (Figure 5B) and, when present, were far outnumbered by CF pathogens (Figure 5C). These findings are consistent with supplemental data presented by Hogan et al. (2016) (see their

Figure 3. Microbiota of BALs with Low CF Pathogen Abundance Are More Similar to Paired Scope Washes Than to Paired Mouth and Throat Samples

(A) Total bacterial abundance in BAL samples, ordered from lowest to highest. Reference lines indicate mean and maximum scope pre-wash bacterial DNA loads. Ordering of BAL samples (n = 102) is maintained in A–E.

(B) Relative abundance of CF pathogens (red bars) and non-conventional organisms (gray bars) in BAL samples.

(C–E) Morisita-Horn dissimilarity indices of BALs as compared to subject-paired mouth (C), throat (D), and instrument control (E) samples; 0 is maximum similarity, 1 is maximum dissimilarity. Dissimilarities were calculated using all identified bacterial operational taxonomic units (OTUs).

(F) Mean relative abundance of taxa in all 102 BALs plotted as a function of their mean relative abundance in instrument controls. CF pathogen taxa (colored points) are labeled. The black line indicates the linear regression fit to non-conventional organism taxa (black points).

(G) Mean relative abundance of all non-conventional taxa in 34 BALs with low CF pathogen abundance (<164 CF pathogen genome copies/mL) plotted as a function of their mean relative abundance in instrument controls. CF pathogens were computationally removed for this analysis. The black line indicates the linear regression fit to the non-conventional organism OTUs.

In (F) and (G), Pearson correlations are indicated for each fit (r and p values).

See also Figure S5.

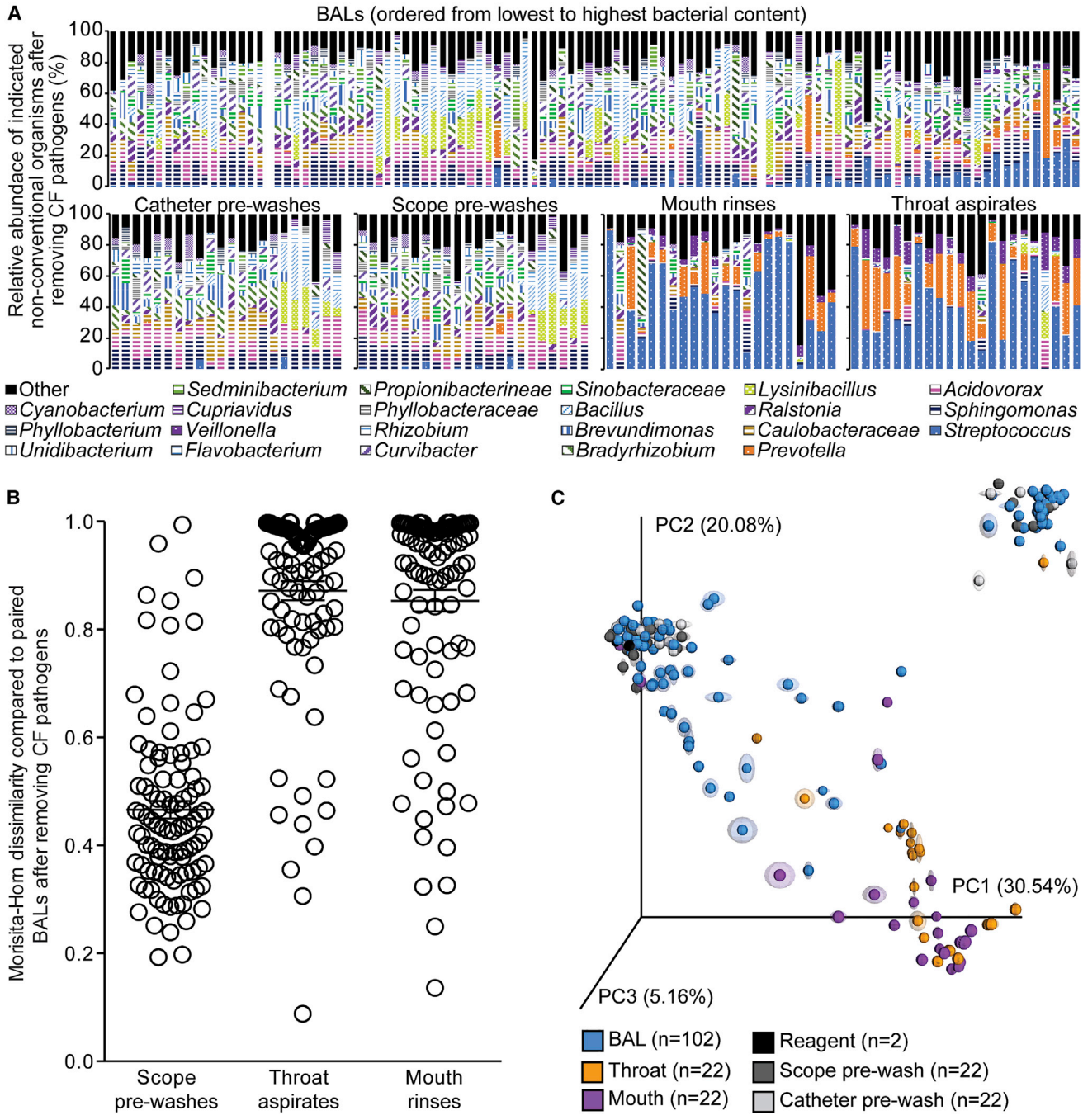


Figure 4. Non-conventional Organisms in BALs Are Similar to Scope and Catheter Pre-washes but Different from Paired Mouth and Throat Samples

(A) Relative abundance of non-conventional taxa detected in all 102 BALs (top); and 22 catheter and scope pre-washes, mouth and throat samples (bottom) after CF pathogens were computationally removed.

(B) Morisita-Horn dissimilarity of non-conventional organisms in all 102 BALs (after pathogens were removed) as compared to non-conventional organisms in subject-paired scope pre-wash, mouth rinse, and throat aspirate samples. Mean and SEM are indicated.

(C) Principal coordinates analysis of Morisita-Horn dissimilarities of non-conventional organisms (after pathogens removed) in BAL, mouth rinse, throat aspirate, and instrument control samples described in (A). The scope and catheter pre-wash control samples separate into two groups along principle coordinate 1 according to the batch in which they were sequenced.

Table 1. Non-conventional Taxa Significantly Overrepresented in BALs Relative to Controls, See Also Table S6

BALs without CF Pathogens[a]				
No non-conventional taxa were significantly overrepresented in BALs without CF pathogens as compared to controls[b]				
BALs with CF Pathogens[c]				
Taxa Overrepresented versus Controls[b]	Relative Abundance in Controls[d] (%)	Relative Abundance in BALs[e] (%)	FoldDifference[f]	p Value[g]
Streptococcus	0.88	4.90	3	1.9×10^{-4}
Prevotella	0.40	2.23	2	1.6×10^{-5}
Veillonella	0.07	0.48	3	5.9×10^{-7}
Alcaligenaceae	0.00	1.72	19	9.9×10^{-19}

[a]BALs considered "without CF pathogens" contained less than 150% maximum value in instrument and reagent controls (i.e., <164 CF pathogen genome copies/mL).
[b]Based on results of negative binomial test (see text).
[c]BALs considered "with CF pathogens" contained more than 150% of the maximum value in instrument and reagent controls (i.e., >164 CF pathogen genome copies/mL).
[d]Mean relative abundance of OTU in all catheter pre-washes, scope pre-washes, and reagent controls studied.
[e]Mean relative abundance of OTU in all BALs studied.
[f]Fold difference calculated by dividing reads of indicated taxa in BALs by level in controls.
[g]p values are for taxa abundance in BAL relative to controls.

Table S2), which indicated that CF pathogens outnumbered *Streptococcus*, *Prevotella*, and *Veillonella* by ~1,000- to ~4,000-fold in BALs of CF adults with mild to moderate lung disease (Figure S6). Notably, these oral bacteria were far more abundant in subject-paired sputum samples than in BALs (Figure S6).

Multivariate Analyses Implicate CF Pathogens in Lung Inflammation

Although CF pathogens are generally acknowledged to cause CF lung inflammation, the effect of oral organisms that may inhabit lungs is less clear. Indeed, both beneficial (Filkins et al., 2012; Zemanick et al., 2013) and detrimental (Conrad et al., 2013) effects on inflammation have been postulated.

We first used pairwise linear regression models and found that the BAL abundance of CF pathogens, *Streptococcus*, *Veillonella*, and *Prevotella* each correlated with neutrophil elastase, a robust marker of disease-causing inflammation (Sagel et al., 2007) (Figure 5D). However, *Streptococcus*, *Veillonella*, and *Prevotella* and CF pathogen abundances were strongly correlated in these BALs (Figure 5B), so pairwise associations could be spurious.

We, thus, used multivariate regression modeling, and again found CF pathogen abundance significantly associated with neutrophil elastase levels in all models tested (Figure 5E). Indeed, for each $\log_{10}$ increase in CF pathogen abundance, neutrophil elastase increased by nearly 0.4 $\log_{10}$ μg/mL. For reference, the primary anti-inflammatory medicine used in CF (azithromycin) was found to reduce neutrophil elastase by 0.2 $\log_{10}$ μg/mL (Saiman et al., 2003). In contrast, models incorporating *Prevotella*, *Streptococcus*, or *Veillonella* abundance identified no positive or negative effect of these organisms (Figure 5E). *Prevotella*, *Streptococcus*, and *Veillonella* exhibited significant collinearity and could not be analyzed in the same model.

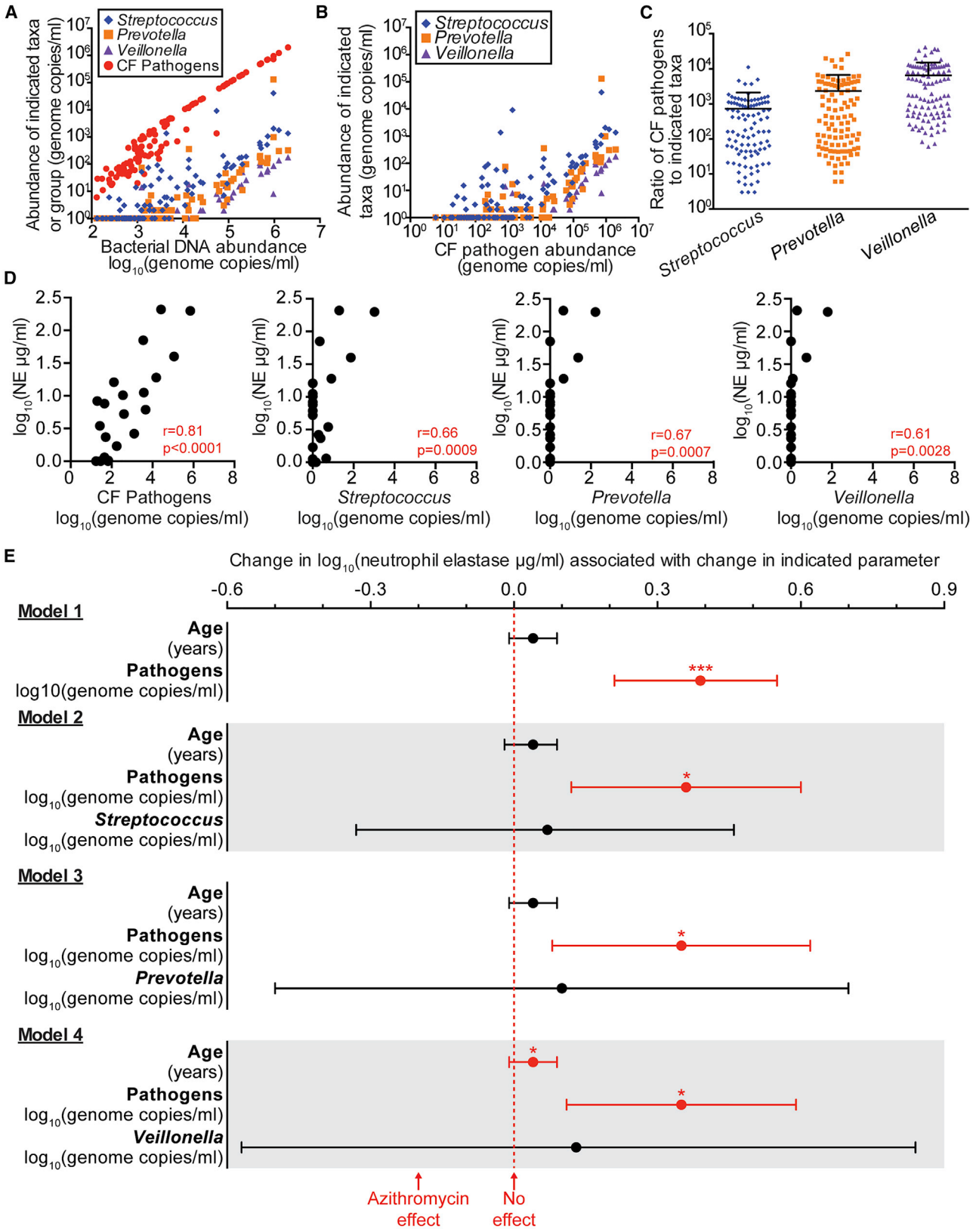
Effect of Treatment on Results

We investigated the potential effects of antibiotic and anti-inflammatory medications on the data. We found no differences in BAL bacterial DNA abundance or neutrophil elastase in subjects that were (n = 16) or were not (n = 6) taking antibiotics and were (n = 15) or were not (n = 7) taking anti-inflammatories (Figures S7B–S7E). Principle coordinate analysis of Morisita-Horn dissimilarities showed that the BALs did not cluster based on antibiotic use (Figure S7F). Indeed, the batch of reagents used for processing appeared to have a more pronounced effect than antibiotics (Figure S7G). Furthermore, the abundance of *Streptococcus*, *Veillonella*, and *Prevotella* was actually higher in BALs from antibiotic-treated subjects (Figure S7H). These observations may be related to our finding that *Streptococcus*, *Veillonella*, and *Prevotella* were more abundant in BALs with more CF pathogens and argue against antibiotic use as an explanation for the low abundance of oral organisms detected in these BALs.

Sequencing Can Produce False Positive Results

The strong association between CF pathogens and inflammation indicates that the accurate detection of pathogens is critical. We compared CF pathogen identification by clinical laboratory cultures and sequencing, limiting analysis to BALs studied by both methods. As shown in Figure 6A, in 11/14 cases, the CF pathogen genera identified by the clinical laboratory were also identified as highly abundant by 16S sequencing. In 2/18 BALs (from subjects 7 and 8), the clinical lab identified multiple CF pathogens, and in these cases, only 1 pathogen was detected by sequencing. However, culture detection of CF pathogens was generally corroborated by sequencing results.

In contrast, our data indicate that it could be challenging to use sequencing data in isolation to determine if CF pathogens are present or absent. On average, 13% of the sequences in scope and catheter pre-washes mapped to CF pathogens (Figure S1C) even though these control samples were obtained prior to any contact with subjects. Furthermore, BALs from the 3 subjects that were culture negative for pathogens (subjects 6, 9, and 15) and whose BAL-1a had low levels of neutrophil elastase contained CF pathogen sequences at a relative abundance of 5%, 14%, and 16% (Figure 6B).



(legend on next page)

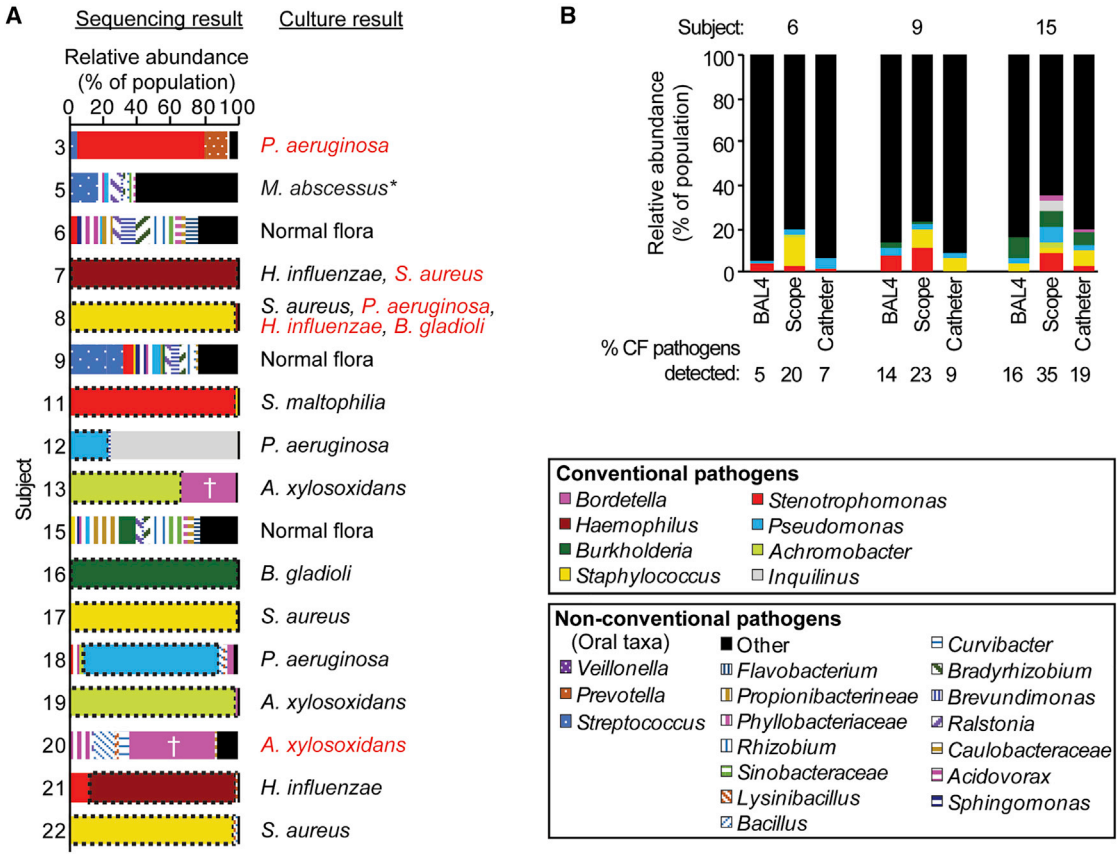


Figure 6. Sequence-Based Bacterial Identification Generally Agrees with Culture but Can Produce False Positives

(A) Sequencing and culture results of 18 BALs (i.e., BAL-4 samples) that were studied by both approaches. Taxa relative abundance (from sequencing) are shown at left, and dashed boxes indicate taxa detected by culture. Culture results are listed at right, and pathogens indicated in red were not detected by sequencing at appreciable levels (<1% relative abundance). We investigated the *Bordetella* results (indicated by a cross [†]) by subjecting the *Bordetella* reference sequence in the Silva database to BLAST analysis in the NCBI 16S rRNA database, and found the *Bordetella* reference highly similar to *Achromobacter* sp. Thus, reads mapping to *Bordetella* may have originated from *Achromobacter*, as indicated by culture results. *M. abscessus* (indicated by *) may not have been detected by sequencing because procedures used may not have lysed mycobacterial cells.

(B) CF pathogens were detected in culture-negative BALs. Stacked bar graphs show relative abundance of CF pathogens (colored bars) and non-conventional organisms (black bars) in 3 culture-negative BALs and their subject-paired instrument control samples. The percentage of bacterial DNA mapping to CF pathogens is listed below each bar.

These findings are consistent with supplemental data from Hogan et al., (2016) (Table S2 in Hogan et al., 2016) indicating that washes from bronchoscopes inserted to the level of the vocal cords, but not entering subjects' lungs, contained an average of 24% CF pathogens. Notably, the CF pathogen genera present in the scope wash controls and culture-negative BALs from our study and scope washes from Hogan et al., (2016) included *Pseudomonas*, *Burkholderia*, *Achromobacter*, and others that typically prompt clinical treatment. These data suggest that accurate sequence-based detection of CF pathogens could be problematic without extensive measures to minimize and account for background contamination.

Figure 5. CF Pathogens Vastly Outnumber Non-conventional CF Organisms in BALs and Are Associated with Lung Inflammation

(A) The absolute abundance of CF pathogens, *Streptococcus*, *Prevotella*, and *Veillonella* in all 102 BALs plotted as a function of total BAL bacterial abundance.

(B) The absolute abundance of *Streptococcus*, *Prevotella*, and *Veillonella* in all 102 BALs plotted as a function of CF pathogen abundance.

(C). Ratios of the abundance of CF pathogens to *Streptococcus*, *Prevotella*, and *Veillonella* in all 102 BALs. Mean and SD are indicated.

(D) Pairwise correlation analyses of neutrophil elastase levels with CF pathogens, *Prevotella*, *Streptococcus*, and *Veillonella* in 22 BALs. Pearson correlation coefficients and p values are shown, and red designates significant correlation (see Figure S7A for correlations between neutrophil elastase and age, BMI, and FEV_1 and bacterial DNA abundance Shannon diversity). In (A)–(D), each point represents measurements from individual BALs.

(E) Four multivariate regression models show that the abundance of CF pathogens positively correlated with neutrophil elastase, whereas *Prevotella*, *Streptococcus*, and *Veillonella* exhibited no positive or negative correlation. Error bars indicate 95% confidence limits of parameter estimates. Red indicates factors that significantly correlated with neutrophil elastase in each multivariate model (*$p < 0.05$, ***$p < 0.0005$). Red dashed line indicates no effect on neutrophil elastase, and the red arrow shows the negative effect of azithromycin on neutrophil elastase levels identified in Saiman et al., (2003) for reference.

See also Figures S6 and S7.

DISCUSSION

We studied BALs from children with CF to address the following questions about CF infection pathogenesis.

Do Diverse Communities of Mostly Oral Organisms Dominate CF Lungs before Pathogens?

The analyses comparing BALs to subject-paired mouth and throat samples indicated that the bacterial DNA in BALs with low (essentially background) levels of CF pathogens were generally very different from subjects' mouth and throat samples (see Figures 3, 4, and S5) but similar to subject-paired instrument controls. A statistical test corroborated these findings, finding no non-conventional taxa that were statistically more abundant in low-pathogen BALs than in instrument controls. Thus, these findings do not support the idea that diverse communities of oral organisms colonize lungs before CF pathogens.

Do Diverse Communities of Non-conventional Organisms Co-exist with Pathogens?

Neither the negative binomial test nor the analysis that computationally removed CF pathogen reads found evidence for high-diversity communities of non-conventional organisms in BALs in which pathogens were abundant (Figure 4). However, this analysis did identify three non-conventional taxa (*Streptococcus*, *Prevotella*, and *Veillonella*) that were over-represented in these BALs relative to controls. Notably, these organisms were not over-represented (relative to instrument controls) in BALs that lacked pathogens (Table 1, Figures 3G, 4A, and 5B), and when present, they were vastly outnumbered by pathogens (Figure 5C). Furthermore, multiple regression analysis found that whereas CF pathogen abundance exhibited a strong positive association with airway inflammation, neither positive nor negative effects of the low-abundance *Streptococcus*, *Prevotella*, or *Veillonella* were apparent (Figure 5E).

How Do We Interpret Differences in Microbial Diversity Indices?

Studies using DNA-based analyses suggest that microbial diversity measurements (which are calculated from relative abundance data) are a key disease parameter as strong inverse relationships between diversity and lung function and age have been observed (Coburn et al., 2015; Cox et al., 2010; Fodor et al., 2012; Klepac-Ceraj et al., 2010; Stokell et al., 2015; Zhao et al., 2012). Indeed, our study produced analogous findings, as we found strong inverse relationships between Shannon diversity and BAL bacterial abundance (Figure 2B) and neutrophil elastase (Figure S7A).

However, interpreting BAL diversity differences in the context of absolute bacterial abundance data and accounting for contaminants present in subject-paired scope pre-washes suggested that decreases in BAL diversity were due to an increased absolute abundance of CF pathogens added to a relatively constant background of non-conventional organisms that were highly similar to the organisms present in scope pre-wash controls (see Figures 2D–2F). In other words, decreases in BAL diversity appeared to be due primarily to increases in CF pathogens reducing the signal from contaminants. Importantly, an analogous mechanism could operate in sputum samples as increases in CF pathogen abundance could reduce the signal from diverse bacterial communities in contaminating oropharyngeal secretions.

Implications for CF Infection Pathogenesis

Although our findings were consistent with two major predictions of the established pathogenesis model (i.e., identifiable microbiota were not detected before pathogen infection and pathogens likely drive inflammation), our data raise questions about the established model's postulate that only CF pathogens are present. The low abundance of *Streptococcus*, *Prevotella*, and *Veillonella* DNA found in BALs that contained CF pathogens raises the possibility that these organisms can cause superinfection of pathogen-infected lungs. If this were the case, they could contribute to disease independently or by interactions with established pathogens. However, previous work also suggests that patients with more cough and more advanced disease may experience more aspiration (Blondeau et al., 2008; Blondeau et al., 2010); thus, it is possible that the low abundance of *Streptococcus*, *Prevotella*, and *Veillonella* DNA in BALs could be a consequence of increased aspiration of oral secretions. These findings, together with the regression modeling data finding no association between *Streptococcus*, *Prevotella*, and *Veillonella* abundance with pathogen-associated inflammation indicate that additional work is needed to determine if and how these organisms might affect disease.

Comparing Results with Other Studies

Comparing results from studies that use DNA-based analyses is challenging because of differences in patient characteristics, specimen collection and processing and sequencing approaches, and methods used to control for contaminants. Although several factors may explain differences between our results and previous work analyzing CF airway microbiomes (Carmody et al., 2013; Coburn et al., 2015; Cox et al., 2010; Filkins et al., 2012; Frayman et al., 2017; Hogan et al., 2016; Pittman et al., 2017; Renwick et al., 2014; Whiteson et al., 2014; Zemanick et al., 2013; Zemanick et al., 2015; Zhao et al., 2012), we speculate that one key factor was the use of cuffed endotracheal tubes for bronchoscopy that were inserted by laryngoscopy under general anesthesia. This approach eliminated scope contact with the oropharynx and likely reduced contamination with oropharyngeal secretions. The saliva of CF children has been reported to contain an average of 10^{10} bacteria/mL (Zemanick et al., 2015). Thus, mixing as little as 1/100 of a milliliter of saliva into 10 mL of BAL fluid or sputum could produce a signal reporting $\sim 10^6$ oral bacteria per mL. The sampling approach used here likely reduced upper airway contamination and, thus, decreased detection of oral organisms. Other studies have used different approaches to control for contaminants, including using multiple bronchoscopes and protected brushes to sample the airways (Hogan et al., 2016), sequencing reagent control samples (Muhlebach et al., 2018), and sequencing oropharyngeal control samples (Zemanick et al., 2015), and the different methods could produce different results.

Previous work using BAL is somewhat more analogous to this study, but comparisons are still hampered by methodological

differences. For example, studies that reached different conclusions about the abundance of oral organisms in CF lungs either studied infants or very young children (Frayman et al., 2017; Muhlebach et al., 2018; Pittman et al., 2017), did not analyze data in reference to paired instrument or oropharyngeal controls (Hogan et al., 2016), and in some cases performed BAL through laryngeal masks (Frayman et al., 2017; Pittman et al., 2017; Renwick et al., 2014) or trans-orally (Hogan et al., 2016) rather than by endotracheal tubes. Finally, as indicated above, interpreting differences in diversity indices between samples using absolute and relative taxa abundance values can lead to different conclusions. We speculate that these factors account for some of the different conclusions reached between this and other studies.

Study Limitations

Our study had several limitations. First, we collected cross-sectional samples and more complete information about early infections requires longitudinal sampling. For example, it is possible that oral organisms could be present periodically and transiently affect disease manifestations. Second, subjects underwent bronchoscopy for varying clinical indications and were on physician-directed (rather than protocol-directed) treatments, and this likely introduced variability to the findings. The use of bronchoscopies for clinical indications and the lack of longitudinal samples are difficult to avoid given the risks of general anesthesia, endotracheal intubation, and bronchoscopy in children.

Third, the DNA-based methods we used do not distinguish organisms that are living and replicating in the lung from those that are deposited in the airways and then rapidly killed. This is particularly important for interpreting the low abundance of *Streptococcus*, *Prevotella*, and *Veillonella* in BALs dominated by pathogens, as work in non-CF subjects suggests that oral bacteria may be aspirated and rapidly cleared from lungs (Dickson et al., 2017). It is also important to note that although the statistical analysis found no evidence for these or other oral organisms in BALs without pathogens, it is possible that they are present at very low abundances and could not be resolved from background contaminants. In addition, this analysis compared the values and variance of taxa present in all BAL and controls as groups, and some BALs could be outliers in terms of taxa composition.

Fourth, it was impractical to directly measure the absolute abundance of each taxon identified. Thus, we calculated absolute abundances from measurements of total bacterial genome copies and taxa relative abundance. Our validation experiments suggest this approach is reasonable. Finally, we sampled a limited number of lung sites (usually 4), only used BAL, did not study non-CF children, CF children less than 5 years old, or those with advanced disease, and did not have a cohort undergoing defined pulmonary exacerbations. Sampling more sites, protected specimen brushes, and studying different subjects could produce different results.

Conclusions

Identifying the bacteria infecting the lungs of CF children is extremely challenging because the lungs lie distal to the heavily colonized oropharynx, lung bacterial biomass can be low, and reagents contain significant contaminating DNA. Additional work is needed that uses stringent sampling, extensive controls, absolute abundance measurements, and perhaps novel approaches (such as differential immunoglobulin coating to identify bacterial origin). This work will be critical because future progress depends upon accurately determining which bacteria are present in the lungs of people with CF and contribute to disease.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Human subjects
- METHOD DETAILS
  - Study design and sample collection
  - DNA isolation
  - Bacterial quantification
  - 16S rRNA gene sequencing
  - 16S rRNA gene sequencing analyses
  - Quantification of OTU absolute abundances
  - Neutrophil elastase activity assays
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2019.03.086.

ACKNOWLEDGMENTS

We thank the participants who made this study possible. P.J. was supported by fellowships from the Cystic Fibrosis Foundation (JORTH14F0 and JORTH17F5). This study was supported by the Cystic Fibrosis Foundation, University of Washington and University of Cincinnati's CF Research Development programs, NIH (K24HL102246, P30DK089507, and P30DK1117467), and the Burroughs Wellcome Fund (BWF1006700). We thank C. Penland for advice and encouragement, and L. Hoffman and C. Manoil for helpful discussions and critical review of the manuscript.

AUTHOR CONTRIBUTIONS

Conceptualization: P.J., Z.E., A.R., C.H.G., D.B., J.P.C., and P.K.S.; Methodology: P.J., Z.E., A.R., E.C., C.P., C.H.G., D.B., J.P.C., and P.K.S.; Formal Analysis: P.J., E.C., C.H.G.; Investigation: P.J., Z.E., A.R., C.P., and D.B.; Resources: P.J., Z.E., D.B., J.P.C., and P.K.S.; Data Curation, P.J. and E.C.; Writing – Original Draft, P.J., J.P.C., and P.K.S.; Writing – Review and Editing: P.J., Z.E., J.P.C., and P.K.S.; Visualization: P.J.; Supervision: C.H.G., J.P.C., and P.K.S.; Project Administration: P.J., J.P.C., and P.K.S.; Funding Acquisition: P.J., J.P.C., and P.K.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 13, 2018
Revised: December 17, 2018
Accepted: March 22, 2019
Published: April 23, 2019

REFERENCES

Baron, E.J., Miller, J.M., Weinstein, M.P., Richter, S.S., Gilligan, P.H., Thomson, R.B., Jr., Bourbeau, P., Carroll, K.C., Kehl, S.C., Dunne, W.M., et al. (2013). A guide to utilization of the microbiology laboratory for diagnosis of infectious diseases: 2013 recommendations by the Infectious Diseases Society of America (IDSA) and the American Society for Microbiology (ASM)(a). Clin. Infect. Dis. *57*, e22–e121.

Bik, E.M., Long, C.D., Armitage, G.C., Loomer, P., Emerson, J., Mongodin, E.F., Nelson, K.E., Gill, S.R., Fraser-Liggett, C.M., and Relman, D.A. (2010). Bacterial diversity in the oral cavity of 10 healthy individuals. ISME J. *4*, 962–974.

Blondeau, K., Dupont, L.J., Mertens, V., Verleden, G., Malfroot, A., Vandenplas, Y., Hauser, B., and Sifrim, D. (2008). Gastro-oesophageal reflux and aspiration of gastric contents in adult patients with cystic fibrosis. Gut *57*, 1049–1055.

Blondeau, K., Pauwels, A., Dupont, Lj., Mertens, V., Proesmans, M., Orel, R., Brecelj, J., López-Alonso, M., Moya, M., Malfroot, A., et al. (2010). Characteristics of gastroesophageal reflux and potential risk of gastric content aspiration in children with cystic fibrosis. J. Pediatr. Gastroenterol. Nutr. *50*, 161–166.

Brown, P.S., Pope, C.E., Marsh, R.L., Qin, X., McNamara, S., Gibson, R., Burns, J.L., Deutsch, G., and Hoffman, L.R. (2014). Directly sampling the lung of a young child with cystic fibrosis reveals diverse microbiota. Ann. Am. Thorac. Soc. *11*, 1049–1055.

Burns, J.L., and Rolain, J.M. (2014). Culture-based diagnostic microbiology in cystic fibrosis: can we simplify the complexity? J. Cyst. Fibros. *13*, 1–9.

Burns, J.L., Gibson, R.L., McNamara, S., Yim, D., Emerson, J., Rosenfeld, M., Hiatt, P., McCoy, K., Castile, R., Smith, A.L., and Ramsey, B.W. (2001). Longitudinal assessment of Pseudomonas aeruginosa in young children with cystic fibrosis. J. Infect. Dis. *183*, 444–452.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. Nat. Methods *7*, 335–336.

Carmody, L.A., Zhao, J., Schloss, P.D., Petrosino, J.F., Murray, S., Young, V.B., Li, J.Z., and LiPuma, J.J. (2013). Changes in cystic fibrosis airway microbiota at pulmonary exacerbation. Ann. Am. Thorac. Soc. *10*, 179–187.

Coburn, B., Wang, P.W., Diaz Caballero, J., Clark, S.T., Brahma, V., Donaldson, S., Zhang, Y., Surendra, A., Gong, Y., Elizabeth Tullis, D., et al. (2015). Lung microbiota across age and disease stage in cystic fibrosis. Sci. Rep. *5*, 10241.

Conrad, D., Haynes, M., Salamon, P., Rainey, P.B., Youle, M., and Rohwer, F. (2013). Cystic fibrosis therapy: a community ecology perspective. Am. J. Respir. Cell Mol. Biol. *48*, 150–156.

Cox, M.J., Allgaier, M., Taylor, B., Baek, M.S., Huang, Y.J., Daly, R.A., Karaoz, U., Andersen, G.L., Brown, R., Fujimura, K.E., et al. (2010). Airway microbiota and pathogen abundance in age-stratified cystic fibrosis patients. PLoS One *5*, e11044.

Culver, D.A., Gordon, S.M., and Mehta, A.C. (2003). Infection control in the bronchoscopy suite: a review of outbreaks and guidelines for prevention. Am. J. Respir. Crit. Care Med. *167*, 1050–1056.

Dewhirst, F.E., Chen, T., Izard, J., Paster, B.J., Tanner, A.C., Yu, W.H., Lakshmanan, A., and Wade, W.G. (2010). The human oral microbiome. J. Bacteriol. *192*, 5002–5017.

Dickson, R.P., Erb-Downward, J.R., Freeman, C.M., McCloskey, L., Falkowski, N.R., Huffnagle, G.B., and Curtis, J.L. (2017). Bacterial Topography of the Healthy Human Lower Respiratory Tract. MBio *8*, e02287-16.

Emerson, J., Rosenfeld, M., McNamara, S., Ramsey, B., and Gibson, R.L. (2002). Pseudomonas aeruginosa and other predictors of mortality and morbidity in young children with cystic fibrosis. Pediatr. Pulmonol. *34*, 91–100.

Faro, A., Wood, R.E., Schechter, M.S., Leong, A.B., Wittkugel, E., Abode, K., Chmiel, J.F., Daines, C., Davis, S., Eber, E., et al.; American Thoracic Society Ad Hoc Committee on Flexible Airway Endoscopy in Children (2015). Official American Thoracic Society technical standards: flexible airway endoscopy in children. Am. J. Respir. Crit. Care Med. *191*, 1066–1080.

Filkins, L.M., Hampton, T.H., Gifford, A.H., Gross, M.J., Hogan, D.A., Sogin, M.L., Morrison, H.G., Paster, B.J., and O'Toole, G.A. (2012). Prevalence of streptococci and increased polymicrobial diversity associated with cystic fibrosis patient stability. J. Bacteriol. *194*, 4709–4717.

Fodor, A.A., Klem, E.R., Gilpin, D.F., Elborn, J.S., Boucher, R.C., Tunney, M.M., and Wolfgang, M.C. (2012). The adult cystic fibrosis airway microbiota is stable over time and infection type, and highly resilient to antibiotic treatment of exacerbations. PLoS One *7*, e45001.

Frayman, K.B., Armstrong, D.S., Carzino, R., Ferkol, T.W., Grimwood, K., Storch, G.A., Teo, S.M., Wylie, K.M., and Ranganathan, S.C. (2017). The lower airway microbiota in early cystic fibrosis lung disease: a longitudinal analysis. Thorax *72*, 1104–1112.

Gibson, R.L., Burns, J.L., and Ramsey, B.W. (2003). Pathophysiology and management of pulmonary infections in cystic fibrosis. Am. J. Respir. Crit. Care Med. *168*, 918–951.

Goddard, A.F., Staudinger, B.J., Dowd, S.E., Joshi-Datar, A., Wolcott, R.D., Aitken, M.L., Fligner, C.L., and Singh, P.K. (2012). Direct sampling of cystic fibrosis lungs indicates that DNA-based analyses of upper-airway specimens can misrepresent lung microbiota. Proc. Natl. Acad. Sci. USA *109*, 13769–13774.

Hisert, K.B., Heltshe, S.L., Pope, C., Jorth, P., Wu, X., Edwards, R.M., Radey, M., Accurso, F.J., Wolter, D.J., Cooke, G., et al. (2017). Restoring Cystic Fibrosis Transmembrane Conductance Regulator Function Reduces Airway Bacteria and Inflammation in People with Cystic Fibrosis and Chronic Lung Infections. Am. J. Respir. Crit. Care Med. *195*, 1617–1628.

Hogan, D.A., Willger, S.D., Dolben, E.L., Hampton, T.H., Stanton, B.A., Morrison, H.G., Sogin, M.L., Czum, J., and Ashare, A. (2016). Analysis of Lung Microbiota in Bronchoalveolar Lavage, Protected Brush and Sputum Samples from Subjects with Mild-To-Moderate Cystic Fibrosis Lung Disease. PLoS One *11*, e0149998.

Horn, H.S. (1966). Measurement of "overlap" in comparative ecological studies. Am. Nat. *100*, 419–424.

Jorth, P., Turner, K.H., Gumus, P., Nizam, N., Buduneli, N., and Whiteley, M. (2014). Metatranscriptomics of the human oral microbiome during health and disease. MBio *5*, e01012–e01014.

Kennedy, K., Hall, M.W., Lynch, M.D., Moreno-Hagelsieb, G., and Neufeld, J.D. (2014). Evaluating bias of illumina-based bacterial 16S rRNA gene profiles. Appl. Environ. Microbiol. *80*, 5717–5722.

Kirschke, D.L., Jones, T.F., Craig, A.S., Chu, P.S., Mayernick, G.G., Patel, J.A., and Schaffner, W. (2003). Pseudomonas aeruginosa and Serratia marcescens contamination associated with a manufacturing defect in bronchoscopes. N. Engl. J. Med. *348*, 214–220.

Klepac-Ceraj, V., Lemon, K.P., Martin, T.R., Allgaier, M., Kembel, S.W., Knapp, A.A., Lory, S., Brodie, E.L., Lynch, S.V., Bohannan, B.J., et al. (2010). Relationship between cystic fibrosis respiratory tract bacterial communities and age, genotype, antibiotics and Pseudomonas aeruginosa. Environ. Microbiol. *12*, 1293–1303.

Lim, Y.W., Schmieder, R., Haynes, M., Willner, D., Furlan, M., Youle, M., Abbott, K., Edwards, R., Evangelista, J., Conrad, D., and Rohwer, F. (2013). Metagenomics and metatranscriptomics: windows on CF-associated viral and microbial communities. J. Cyst. Fibros. *12*, 154–164.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Lyczak, J.B., Cannon, C.L., and Pier, G.B. (2002). Lung infections associated with cystic fibrosis. Clin. Microbiol. Rev. *15*, 194–222.

Madan, J.C., Koestler, D.C., Stanton, B.A., Davidson, L., Moulton, L.A., Housman, M.L., Moore, J.H., Guill, M.F., Morrison, H.G., Sogin, M.L., et al. (2012). Serial analysis of the gut and respiratory microbiome in cystic fibrosis in infancy: interaction between intestinal and respiratory tracts and impact of nutritional exposures. MBio *3*, e00251-12.

Mahboubi, M.A., Carmody, L.A., Foster, B.K., Kalikin, L.M., VanDevanter, D.R., and LiPuma, J.J. (2016). Culture-Based and Culture-Independent Bacteriologic Analysis of Cystic Fibrosis Respiratory Specimens. J. Clin. Microbiol. *54*, 613–619.

Mehta, A.C., and Minai, O.A. (1999). Infection control in the bronchoscopy suite. A review. Clin. Chest Med. *20*, 19–32, ix.

Muhlebach, M.S., Zorn, B.T., Esther, C.R., Hatch, J.E., Murray, C.P., Turkovic, L., Ranganathan, S.C., Boucher, R.C., Stick, S.M., and Wolfgang, M.C. (2018). Initial acquisition and succession of the cystic fibrosis lung microbiome is associated with disease progression in infants and preschool children. PLoS Pathog. *14*, e1006798.

O'Toole, G.A. (2018). Cystic Fibrosis Airway Microbiome: Overturning the Old, Opening the Way for the New. J. Bacteriol. *200*, e00561-17.

Pittman, J.E., Wylie, K.M., Akers, K., Storch, G.A., Hatch, J., Quante, J., Frayman, K.B., Clarke, N., Davis, M., Stick, S.M., et al.; Australian Respiratory Early Surveillance Team for Cystic Fibrosis (2017). Association of Antibiotics, Airway Microbiome, and Inflammation in Infants with Cystic Fibrosis. Ann. Am. Thorac. Soc. *14*, 1548–1555.

Price, K.E., Hampton, T.H., Gifford, A.H., Dolben, E.L., Hogan, D.A., Morrison, H.G., Sogin, M.L., and O'Toole, G.A. (2013). Unique microbial communities persist in individual cystic fibrosis patients throughout a clinical exacerbation. Microbiome *1*, 27.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. *41*, D590–D596.

Renwick, J., McNally, P., John, B., DeSantis, T., Linnane, B., and Murphy, P.; SHIELD CF (2014). The microbial community of the cystic fibrosis airway is disrupted in early life. PLoS One *9*, e109798.

Sagel, S.D., Chmiel, J.F., and Konstan, M.W. (2007). Sputum biomarkers of inflammation in cystic fibrosis lung disease. Proc. Am. Thorac. Soc. *4*, 406–417.

Saiman, L., Marshall, B.C., Mayer-Hamblett, N., Burns, J.L., Quittner, A.L., Cibene, D.A., Coquillette, S., Fieberg, A.Y., Accurso, F.J., and Campbell, P.W., 3rd.; Macrolide Study Group (2003). Azithromycin in patients with cystic fibrosis chronically infected with Pseudomonas aeruginosa: a randomized controlled trial. JAMA *290*, 1749–1756.

Salter, S.J., Cox, M.J., Turek, E.M., Calus, S.T., Cookson, W.O., Moffatt, M.F., Turner, P., Parkhill, J., Loman, N.J., and Walker, A.W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol. *12*, 87.

SAS (2014). SAS 9.4 (SAS Institute Inc., Cary, NC).

Srinivasan, A., Wolfenden, L.L., Song, X., Mackie, K., Hartsell, T.L., Jones, H.D., Diette, G.B., Orens, J.B., Yung, R.C., Ross, T.L., et al. (2003). An outbreak of Pseudomonas aeruginosa infections associated with flexible bronchoscopes. N. Engl. J. Med. *348*, 221–227.

Stokell, J.R., Gharaibeh, R.Z., Hamp, T.J., Zapata, M.J., Fodor, A.A., and Steck, T.R. (2015). Analysis of changes in diversity and abundance of the microbial community in a cystic fibrosis patient over a multiyear period. J. Clin. Microbiol. *53*, 237–247.

Tunney, M.M., Field, T.R., Moriarty, T.F., Patrick, S., Doering, G., Muhlebach, M.S., Wolfgang, M.C., Boucher, R., Gilpin, D.F., McDowell, A., and Elborn, J.S. (2008). Detection of anaerobic bacteria in high numbers in sputum from patients with cystic fibrosis. Am. J. Respir. Crit. Care Med. *177*, 995–1001.

Twomey, K.B., Alston, M., An, S.Q., O'Connell, O.J., McCarthy, Y., Swarbreck, D., Febrer, M., Dow, J.M., Plant, B.J., and Ryan, R.P. (2013). Microbiota and metabolite profiling reveal specific alterations in bacterial community structure and environment in the cystic fibrosis airway during exacerbation. PLoS One *8*, e82432.

Weiss, S., Amir, A., Hyde, E.R., Metcalf, J.L., Song, S.J., and Knight, R. (2014). Tracking down the sources of experimental contamination in microbiome studies. Genome Biol. *15*, 564.

Whiteson, K.L., Meinardi, S., Lim, Y.W., Schmieder, R., Maughan, H., Quinn, R., Blake, D.R., Conrad, D., and Rohwer, F. (2014). Breath gas metabolites and bacterial metagenomes from cystic fibrosis airways indicate active pH neutral 2,3-butanedione fermentation. ISME J. *8*, 1247–1258.

Wolda, H. (1981). Similarity indices, sample size and diversity. Oecologia *50*, 296–302.

Zemanick, E.T., Harris, J.K., Wagner, B.D., Robertson, C.E., Sagel, S.D., Stevens, M.J., Accurso, F.J., and Laguna, T.A. (2013). Inflammation and airway microbiota during cystic fibrosis pulmonary exacerbations. PLoS One *8*, e62917.

Zemanick, E.T., Wagner, B.D., Robertson, C.E., Stevens, M.J., Szefler, S.J., Accurso, F.J., Sagel, S.D., and Harris, J.K. (2015). Assessment of airway microbiota and inflammation in cystic fibrosis using multiple sampling methods. Ann. Am. Thorac. Soc. *12*, 221–229.

Zhao, J., Schloss, P.D., Kalikin, L.M., Carmody, L.A., Foster, B.K., Petrosino, J.F., Cavalcoli, J.D., VanDevanter, D.R., Murray, S., Li, J.Z., et al. (2012). Decade-long bacterial community dynamics in cystic fibrosis airways. Proc. Natl. Acad. Sci. USA *109*, 5809–5814.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial Strains		
Pseudomonas aeruginosa PAO1	Manoil Lab, University of Washington	PAO1
Biological Samples		
Human cystic fibrosis bronchoalveolar lavage samples, oral washes, and post-pharyngeal aspirates	Cincinnati Children's Hospital	N/A
Chemicals, Peptides, and Recombinant Proteins		
RNA-Bee	Tel-test	Cat# CS501B
RNAlater	Invitrogen	Cat# AM7021
Lysing matrix B	MP Biomedical	Cat# 116911500
Chloroform	Fisher	Cat# C607SK-1
Isopropanol	Fisher	Cat# A416-500
Linear acrylamide	Thermo Fisher Scientific	Cat# AM9520
TE Buffer pH 8.0	Thermo Fisher Scientific	Cat# AM9849
Ethanol	Fisher	Cat# A4094
3 M sodium acetate pH 5.5	Invitrogen	Cat# AM9740
SYBR FAST qPCR Master Mix	Kapa Biosystems	Cat# KK4602
HiFi HotStart ReadyMix	Kapa Biosystems	Cat# KK2602
600 cycle v3 MiSeq Reagent Kit	Illumina	Cat# MS-102-3003
MeO-suc-Ala-Ala-Pro-Ala-*p*-nitroanilide	Sigma Aldrich	Cat# M4765
Critical Commercial Assays		
Library Quantification Kit for Illumina	Kapa Biosystems	Cat# KK4828
High Sensitivity D1000 ScreenTape	Agilent	Cat# 5067-5584
SequalPrep Normalization Plates	Applied Biosystems	Cat# A1051001
Deposited Data		
Bacterial 16S rRNA sequencing data	NCBI SRA	Genbank: PRJNA495633
Oligonucleotides		
Primers for bacterial 16S rRNA sequencing, see Table S7	This study	N/A
Primer 16S-CCF-qPCR-F: CCAGACTCCTACGGGCAGC	This study	N/A
Primer 16S-926R-qPCR-F: CCGTCAATTYYTTTRAGTTT	This study	N/A
Software and Algorithms		
Qiime	Caporaso et al., 2010	http://qiime.org
SAS	SAS, 2014	https://www.sas.com/en_us/home.html
DESeq2	Love et al., 2014	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
Prism GraphPad	Prism	https://www.graphpad.com/scientific-software/prism/
Microsoft Excel	Microsoft	https://products.office.com/en-US/excel

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Pradeep K. Singh (singhpr@uw.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human subjects

Patients were enrolled at the Cincinnati Children's Hospital Medical Center with IRB approval. Twenty-two subjects between the ages of 2-21 years old (14 females and 8 males) were recruited for participation in this prospective study if they had a documented diagnosis of CF and were undergoing clinically indicated bronchoscopies. Participation in the study was voluntary, and participants and/or parents provided informed consent prior to enrollment. We did not examine associations between subjects' gender and results as the study objectives were to investigate using stringent sampling methods, multiple types of control samples, taxa absolute abundance data, and within-subject comparisons to interpret BAL microbiome and inflammatory marker data.

METHOD DETAILS

Study design and sample collection

Samples were collected in the following order to limit and control for possible contamination sources. (1) An oral saline rinse was collected to sample the mouth. (2) Sterile saline was drawn through a sterile catheter and saved as the first instrument control sample (called "catheter pre-wash"). (3) Using the catheter, a post-pharyngeal aspirate was collected. (4) Before beginning bronchoscopy, the second instrument control sample was collected by drawing sterile saline through the bronchoscope (called "scope pre-wash"). (5) Patients were intubated to reduce oropharyngeal contamination during the bronchoscopy.

BAL for clinical care is generally performed in only 1-2 lobes as per the ATS guidelines (Faro et al., 2015), and use either fixed volumes of 10 to 50 mL per aliquot or weight-based volumes, typically 1 ml/kg per aliquot. The right middle lobe and lingula are the lung segments most frequently sampled because they provide maximal return. Here we sampled four sites to increase the lung regions sampled. The bronchoscope was inserted into the airways without suctioning if possible, and 2 BAL samples were collected consecutively from two adjacent segmental bronchi of the right middle lobe (BAL-1a and BAL-1b). The second lavage (BAL-2) was then performed in a segmental bronchus of the right upper lobe. The bronchoscope was then repositioned to the left lower lobe (BAL-3) and in the lung segment that appeared most diseased during bronchoscopy, or on prior imaging (BAL-4).

All samples were immediately preserved in RNAlater solution to enable isolation of DNA and RNA from each sample. For a subset of samples, a portion of the untreated sample was immediately frozen for later comparison to the RNAlater treated samples. Taxonomic analyses showed that preservation in RNAlater did not affect the microbes detected in the samples by sequencing. Following collection, the BAL-4 was sent to the Cincinnati Children's Hospital Medical Center Clinical Laboratory for microbial culturing.

DNA isolation

BAL, mouth rinse, throat aspirate, and bronchoscope and catheter pre-wash samples were thawed on ice, and DNA was isolated using RNAbee, as previously described (Jorth et al., 2014). For each sample, 100 μL sample was combined with 1 mL RNA-Bee in a lysing matrix B bead beating tube (MP Biomedical) and bead beat 3 times for 60 s each, incubating on ice for 1 m in between each bead beating. Lysed supernatants were mixed with 200 μl chloroform (Fisher) and shaken vigorously for 45 s. Samples were centrifuged at 13,100 *g* for 30 m at 4°C to separate organic and aqueous phases. Organic phases from each sample were mixed with 500 μL TE buffer pH 8.0 (Thermo Fisher Scientific), shaken vigorously for 30 s, and incubated for 10 m on a tube roller at 25°C. Samples were centrifuged at 13,100 *g* for 30 m at 4°C to separate organic and aqueous phases. The DNA-containing aqueous phase was combined with 750 μL 100% ethanol, 25 μL 3 M sodium acetate pH 5.5, and 1 μL 5 mg/ml linear acrylamide (Thermo Fisher Scientific) to precipitate DNA. Samples were mixed by gentle inversion and incubated for at least 1 h at −80°C. Samples were centrifuged at 16,100 *g* for 15 m at 4°C to pellet DNA. DNA pellets were resuspended in 750 μL 75% ethanol and centrifuged at 16,100 *g* for 5 m at 4°C to wash the DNA. A second 75% ethanol was performed, and ethanol was removed. DNA pellets were air-dried for 5 m at 25°C and DNA was dissolved in 22 μL TE buffer pH 8.0.

Bacterial quantification

Bacteria were quantified in triplicate with 16S qPCR using 16S-CCF-qPCR-F and 16S-926R-qPCR-R primers (Table S7) with SYBR FAST qPCR Master Mix (Kapa Biosystems) using the manufacturer's protocol with a CFX96 Touch Real-Time PCR Detection System (Biorad). Genomic DNA from *P. aeruginosa* PAO1 was used to generate a standard curve for quantification of genome copies per ml of sample. A negative control sample (no template) was included with every standard curve. For every experiment, a standard curve was run alongside samples.

16S rRNA gene sequencing

Sequencing libraries were produced by PCR amplification of 16S rRNA gene V3-V5 variable regions in triplicate reactions using 5 μL DNA per reaction with HiFi HotStart ReadyMix (Kapa Biosystems) using the manufacturer's protocol on a CFX96 Touch Real-Time PCR Detection System (Biorad). PCR was performed in triplicate for each library. Indexed primers 926BC01-926BC96 (Table S7) were used to uniquely identify each library, and primers contained Illumina-compatible DNA sequences for MiSeq sequencing. For each library the triplicate PCR reactions were pooled. Pooled library reactions were purified and normalized using SequalPrep Normalization Plates (Applied Biosystems), according to the manufacturer's instructions. Normalized indexed sequencing libraries

were combined and quantified using qPCR with the Library Quantification Kit for Illumina sequencing platforms (Kapa Biosystems). Library size was verified and by Tape Station analysis with a High Sensitivity D1000 ScreenTape (Agilent) and taken into account for quantification. The 16S rRNA gene sequencing libraries were sequenced on an Illumina MiSeq using paired 300 bp sequencing with a 600 cycle v3 MiSeq Reagent Kit (Illumina) using custom primers MiSeq-926R, MiSeq-926R-Index, and MiSeq-CCF, which were spiked into MiSeq reagent cartridge wells 12, 13, and 14, respectively (Table S7).

16S rRNA gene sequencing analyses

Qiime was used to assign sequencing reads to OTUs, calculate Shannon alpha diversity for each sample, and calculate beta diversity using the Morisita-Horn method (Caporaso et al., 2010; Horn, 1966). For each sequencing library, the 300 bp read 1 sequencing reads were processed for Qiime using custom scripts. Reads were assigned OTUs by Uclust searching against the Silva rRNA database (Quast et al., 2013). For each sample, reads were rarefied ten times to equal depth, 10,000 reads per sample. Alpha diversity was quantified by calculating Shannon diversity for rarefied samples. Morisita-Horn jackknifed beta diversity was calculated with a depth of 10,000 reads per sample when all taxa were included.

Pathogen-free taxonomic profiles were generated by filtering out Staphylococcus, Pseudomonas, Haemophilus, Burkholderia, Stenotrophomonas, Achromobacter, Burkholderia, Bordetella, Inquilinus, and Escherichia-Shigella from the OTU table. Morisita-Horn jackknifed beta diversity was calculated with a depth of 900 reads per sample when only non-conventional CF organisms were analyzed. To identify potential BAL-specific non-conventional organisms we compared all BAL samples with greater than 465 genome copies/ml (since these BAL samples have more bacterial content than the instrument controls and would likely have true non-conventional organisms present) to the control group, which consisted of all catheter pre-wash, bronchoscope pre-wash, and reagent control samples. As a proof-of-principle we also compared the oropharyngeal samples consisting of all oral and throat samples to the same control samples. DESeq2 was used to compare non-conventional CF taxa in BAL versus control samples, as well as oropharyngeal versus control samples (Love et al., 2014).

Quantification of OTU absolute abundances

OTUs were quantified in this study by multiplying the total bacterial abundance determined by 16S rRNA gene qPCR by the relative abundances of each OTU determined by bacterial 16S rRNA gene sequencing. As proof of principle, the same analysis was performed using our previous 16S rRNA gene qPCR and sequencing data from CF sputum collected in Ireland (Hisert et al., 2017). To assess the validity of this quantification method, Pearson correlation analysis was used to compare the extrapolated OTU quantities for *Streptococcus, Prevotella, Pseudomonas, Staphylococcus,* and *Haemophilus* to the quantities of these organisms determined by genera-specific qPCR analysis.

Neutrophil elastase activity assays

Neutrophil elastase activity assays were performed on the BAL1a samples from all 22 patients by spectrophotometric detection of the hydrolysis of MeO-suc-Ala-Ala-Pro-Ala-*p*-nitroanilide (Sigma Chemical Co, St. Louis, MO). As controls, untreated BAL-1a samples were compared to RNAlater treated BAL-1a samples from subjects 18, 19, and 20, and negligible differences were observed.

QUANTIFICATION AND STATISTICAL ANALYSIS

Pearson and Spearman correlation analyses were performed using Prism GraphPad. Multivariate linear regression was used to examine independent effects on BAL neutrophil elastase of CF pathogens, *Streptococcus, Prevotella,* and *Veillonella* adjusted for subject age. Neutrophil elastase was non-normally distributed and required log transformation. We employed model diagnostics as follows: assessment of Studentized residuals; assessment of leverage by plotting leverage versus residuals-square, Cook's Distance (Cook's D), DFIT (a measure of overall measure of influence) and DFBETA. Models were rerun without influential points and outliers to assess changes in coefficients. Quantile-quantile (q-q) plots (qqplots) of residuals were made to assess distribution of residuals. No influential data points were identified that altered the coefficients. Additional plots were of residuals against predicted values were done to assess for evidence of heteroscedasticity along with the White test. Coefficients are presented as the point estimate with 95% confidence intervals and p values. A p value of ≤ 0.05 was considered significant by convention. Graphs and statistical analyses were prepared with Prism GraphPad, Qiime, Microsoft Excel, and SAS® software (Caporaso et al., 2010; SAS, 2014).

DATA AND SOFTWARE AVAILABILITY

Sequencing read data are available in the NCBI Sequence Read Archive under Bioproject accession number PRJNA495633.