Airway microbiota across age and disease spectrum in cystic fibrosis

Edith T. Zemanick1, Brandie D. Wagner1,2, Charles E. Robertson1, Richard C. Ahrens3, James F. Chmiel4, John P. Clancy5, Ronald L. Gibson6, William T. Harris7, Geoffrey Kurland8, Theresa A. Laguna9, Susanna A. McColley10, Karen McCoy11, George Retsch-Bogart12, Kurt T. Sobus13, Pamela L. Zeitlin14, Mark J. Stevens1, Frank J. Accurso1, Scott D. Sagel1 and J. Kirk Harris1

Affiliations: 1University of Colorado School of Medicine, Aurora, CO, USA. 2Colorado School of Public Health, University of Colorado Denver, Aurora, CO, USA. 3University of Iowa, Iowa City, IA, USA. 4Case Western Reserve University School of Medicine, Rainbow Babies and Children’s Hospital, Cleveland, OH, USA. 5Cincinnati Children’s Hospital Medical Center, Cincinnati, OH, USA. 6University of Washington, Seattle Children’s Hospital, Seattle, WA, USA. 7University of Alabama, Birmingham, AL, USA. 8Children’s Hospital of Pittsburgh, Pittsburgh, PA, USA. 9University of Minnesota, Minneapolis, MN, USA. 10Ann and Robert H. Lurie Children’s Hospital of Chicago and Northwestern University, Chicago, IL, USA. 11 Nationwide Children’s Hospital, Columbus, OH, USA. 12University of North Carolina, Chapel Hill, NC, USA. 13St Louis University School of Medicine, St Louis, MO, USA. 14Johns Hopkins University, Baltimore, MD, USA.

Correspondence: Edith T. Zemanick, University of Colorado School of Medicine, 13123 E. 16th Avenue, B-395, Aurora, CO 80045, USA. E-mail: edith.zemanick@childrenscolorado.org

ABSTRACT Our objectives were to characterise the microbiota in cystic fibrosis (CF) bronchoalveolar lavage fluid (BALF), and determine its relationship to inflammation and disease status.

BALF from paediatric and adult CF patients and paediatric disease controls undergoing clinically indicated bronchoscopy was analysed for total bacterial load and for microbiota by 16S rDNA sequencing. We examined 191 BALF samples (146 CF and 45 disease controls) from 13 CF centres. In CF patients aged <2 years, nontraditional taxa (e.g. Streptococcus, Prevotella and Veillonella) constituted ~50% of the microbiota, whereas in CF patients aged ≥6 years, traditional CF taxa (e.g. Pseudomonas, Staphylococcus and Stenotrophomonas) predominated. Sequencing detected a dominant taxon not traditionally associated with CF (e.g. Streptococcus or Prevotella) in 20% of CF BALF and identified bacteria in 24% of culture-negative BALF. Microbial diversity and relative abundance of Streptococcus, Prevotella and Veillonella were inversely associated with airway inflammation. Microbiota communities were distinct in CF compared with disease controls, but did not differ based on pulmonary exacerbation status in CF.

The CF microbiota detected in BALF differs with age. In CF patients aged <2 years, Streptococcus predominates, whereas classic CF pathogens predominate in most older children and adults.
Introduction
Progressive lung disease secondary to chronic airway infection and inflammation is the leading cause of morbidity and mortality in cystic fibrosis (CF) [1]. *Pseudomonas aeruginosa* and *Staphylococcus aureus* along with several other Gram-negative bacteria (*e.g.* *Burkholderia cepacia* complex, *Stenotrophomonas maltophilia* and *Achromobacter* spp.) are the pathogens most frequently cultured from CF airway samples and are associated with pulmonary decline [2]. Molecular sequencing approaches have expanded this traditional view of airway infection [3–9]. CF sputum samples are now known to contain complex bacterial communities, including facultative and obligate anaerobic organisms [10–12]. Distinguishing pathogenic bacteria that contribute to airway disease from commensal microbiota in sputum has proved challenging and remains an obstacle to clinical application of sequencing results [13, 14].

Bronchoscopy with bronchoalveolar lavage fluid (BALF) collection allows identification of lower airway pathogens while limiting upper airway contamination of the sample. However, bronchoscopy is not recommended for routine surveillance in CF due to the need for sedation/anaesthesia and lack of evidence that its use improves outcomes [15, 16]. Most microbiological surveillance in CF in the USA is done by sputum collection or oropharyngeal swabs in nonexpectorating patients. Bronchoscopy is typically reserved for patients in whom infection is suspected but not identified by oropharyngeal or sputum samples, when respiratory symptoms persist despite therapy, or at the time of anaesthesia for another surgical procedure (*e.g.* sinus or gastrointestinal surgery) [17, 18].

Molecular analyses of BALF specimens in other conditions, including asthma, chronic obstructive pulmonary disease, healthy smokers, HIV infection and interstitial lung disease, have identified a lower airway microbiota distinct from upper airway bacteria (albeit with overlap) [19–22]. Small, single-centre studies of CF BALF microbiota have detected typical CF pathogens and anaerobic bacteria [23, 24]. Larger studies of lower airway BALF samples from CF patients across age and disease spectrum are lacking, and data linking microbiota, clinical characteristics and airway inflammation are limited.

To address these gaps, we collected BALF from a diverse cohort of CF patients undergoing a clinically indicated bronchoscopy at 13 CF centres in the USA. BALF samples from non-CF disease controls were collected at the coordinating site. Sequencing data were compared with standard culture, and the relationships between microbiota, BALF cytology and clinical characteristics were determined. Preliminary results from this study have been previously reported in abstract form [25].

Materials and methods

**Study design and subjects**

Patients with CF from 13 CF centres in the USA and disease controls from the coordinating site (Children’s Hospital Colorado (CHCO), Aurora, CO, USA) were recruited at the time of clinically indicated bronchoscopy. Patients with CF (aged 2 months to 50 years) who had at least 0.2 mL remnant BALF remaining after clinical testing were eligible to participate. Paediatric patients undergoing a clinical bronchoscopy for an indication other than CF were recruited as disease controls. The study was approved by the institutional review board at each site. Written informed consent and HIPPA (Health Insurance Portability and Accountability Act of 1996) authorisation were obtained from all patients aged ≥18 years or from parents or legal guardians of patients aged <18 years. Assent was obtained from patients aged 10–17 years. Clinical data at the time of bronchoscopy and Seattle Pulmonary Exacerbation Score (PES) for CF patients [26] were entered into a secure, web-based electronic database (REDCap) [27]. Participants with CF were categorised as 1) clinically stable, defined as no reported respiratory symptoms and PES <5, 2) pulmonary exacerbation, defined as respiratory symptoms leading to bronchoscopy, pulmonary exacerbation unresponsive to treatment or PES ≥5, or 3) unknown.

**Specimen collection and processing**

Bronchoscopy and BALF collection were performed following each site’s standard clinical procedure. The majority of procedures were performed using a laryngeal mask airway or endotracheal tube, minimising upper airway contamination. Standard BALF culture, including bacterial, fungal and nontuberculous mycobacterial cultures, was performed by local clinical microbiology laboratories in accordance with CF Foundation guidelines [28], with cell counts and differentials performed and recorded locally when indicated as determined by supervising physicians. Remnant BALF (1 mL) was aliquoted in cryovials and frozen within 1 h of collection at −70°C. Research samples collected at participating sites were batch-shipped overnight on dry ice to CHCO.

**Laboratory assays**

**DNA extraction and quantitative PCR**

DNA extractions were performed on 0.2 μL BALF using the Qiagen EZ1 Advanced automated extraction platform (Qiagen, Valencia, CA, USA). Total bacterial load (TBL) was measured using a quantitative
real-time PCR assay [29]. A dilution factor was applied to convert results to gene copies per millilitre. Reagent controls were analysed to determine background bacterial load and to estimate the limit of detection (LOD; defined as mean TBL\textsubscript{control}+(3.3×SD\textsubscript{control})) for TBL assay. TBL values below the LOD were recorded, but LOD is shown in the relevant figures.

**Sequencing**

Bacterial profiles were determined by broad-range amplification and sequence analysis of 16S rDNA following previously described methods and validated in prior publications [30–35]. Amplicons were generated using primers targeting ∼300 bp of the V1/V2 variable region of the 16S rRNA gene. Illumina paired-end sequencing was performed on the MiSeq platform using a 500 cycle version 2 reagent kit (Illumina, San Diego, CA, USA).

**Analysis of Illumina paired-end reads**

Quality control procedures were performed on paired-end sequences as described in the supplementary material. Assembled sequences were aligned and classified with SINA version 1.2.11 [36] using the SILVA 111 database [37] as reference configured to yield the SILVA taxonomy (www.arb-silva.de). Sorted paired-end sequence data were deposited in the National Center for Biotechnology Information Sequence Read Archive (www.ncbi.nlm.nih.gov/sra) under accession number SRP044029. The software package Explicet version 2.10.5 (www.explicet.org) [38] was used for display, analysis of Good’s coverage, and calculation of richness, evenness and the Shannon diversity index, a composite measure of evenness and richness, at the rarefaction point of 497 sequences.

**Statistical analysis**

Descriptive statistics include the mean±SD or the median (range), where specified. To account for differences in sequencing depth, relative abundance of each taxon was calculated (number of sequences for specific taxa/total number of sequences×100). All comparisons between disease control and CF subjects used only the samples from paediatric subjects (defined as ≤21 years old, as disease control patients were recruited from a Children’s Hospital and included patients up to age 21 years); adult CF patients were considered separately. Wilcoxon rank-based tests and Chi-squared tests were used to compare demographic information, TBL, diversity, prevalence and relative abundance of taxa across groups. BALF samples with bacterial DNA below the limit of amplification for sequencing were assumed to be negative in calculations of prevalence and agreement with culture results. Spearman’s correlation coefficients were used to determine the correlation between TBL, ecological parameters, clinical variables (pulmonary exacerbation status, lung function, body mass index and age) and airway inflammatory markers (white blood cell count, absolute neutrophil count and percentage neutrophils). We performed a multivariate regression analysis of airway inflammatory markers and the taxa *Pseudomonas*, *Streptococcus*, *Veillonella* and *Prevotella*. We also compared microbial communities by patient age by grouping into age cohorts. TBL and diversity were compared using Kruskal–Wallis rank-based tests. Principal coordinates analysis (PCoA) using the Morisita–Horn distance was applied to visualise the relationship between microbiota communities categorised by age, CF centre, pulmonary exacerbation status and cohort (disease controls or CF). Adjustments for multiple comparisons were made using false discovery rate procedures. Analyses were performed using SAS version 9.4 (SAS Institute, Cary, NC, USA).

**Results**

**Subject characteristics and culture results**

191 BALF samples were collected from 136 paediatric CF, 10 adult CF and 45 paediatric disease control participants. Patient characteristics at the time of bronchoscopy are shown in table 1. The most common indications for bronchoscopy in CF patients (participants may have had more than one indication) were persistent cough (41%), unresponsiveness to treatment (29%), need for lower airway culture (27%), concurrent surgery (21%) and lung function decline (8%). The most common diagnoses and indications for bronchoscopy in disease control patients were asthma or wheezing (38%), pneumonia/recurrent pneumonia (36%), dysphagia/oesophageal reflux (33%) and chronic cough (24%) (supplementary tables S1 and S2). BALF bacterial cultures (obtained in 145 CF and all disease control patients) were positive in nine out of 10 (90%) adult CF, 94 out of 135 (70%) paediatric CF and 15 out of 45 (33%) disease control participants.

**Total bacterial load**

TBL was higher in paediatric CF compared with disease control BALF (median (range) 7.5 (6.8–10.7) versus 7.2 (6.4–8.7) log\textsubscript{10} rDNA copies-m\textsuperscript{-1}; p<0.01) and in those with positive bacterial cultures in CF compared with CF with negative cultures (median (range) 7.9 (6.8–10.7) versus 7.3 (6.9–8.6) log\textsubscript{10} rDNA copies-m\textsuperscript{-1}; p<0.01) (figure 1a and b). TBL did not differ between adult and paediatric CF (p=0.35). LOD
for TBL was estimated at 7.4 log10 rDNA copies·mL$^{-1}$ based on reagent controls (calculated based on mean TBL detected in reagent controls of 188 rRNA gene copies per PCR reaction).

**Sequencing**

CF samples were more likely to have successful amplification for sequencing, with microbiota sequence data obtained from 80% of adult CF and 66% of paediatric CF compared with 27% of disease control BALF (p<0.01). TBL was lower (and at or below the LOD) in BALF samples that failed sequencing compared with those with successful amplification (see supplementary data section and supplementary figure S1). We compared patient characteristics of CF subjects between those with samples that successfully sequenced versus those that did not sequence (supplementary table S3). The primary difference between groups was in culture results, with 89% of BALF cultures positive in the sequence-positive group compared with 33% positive in the sequence-negative group (p<0.01).

**Ecology**

Shannon diversity index, evenness and richness, calculated for samples with sequencing data, were significantly higher in disease control compared with CF participants (figure 1c and supplementary table S4). We did not detect a significant difference in diversity, richness or evenness between adult and paediatric CF participants, although the number of adults was small. Shannon diversity index, richness and evenness were lowest in CF BALF with positive bacterial cultures (figure 1d and supplementary table S4).

**Taxa**

The dominant taxa (defined as taxa with the highest relative abundance) in disease control and CF BALF are shown in table 2. Taxa associated with typical CF pathogens (e.g. *Pseudomonas, Staphylococcus, Stenotrophomonas, Haemophilus, Achromobacter* and *Burkholderia*) were the dominant taxa in 47% of paediatric CF and in 60% of adult CF BALF samples; none of these taxa were dominant in disease control

---

**TABLE 1 Patient characteristics**

<table>
<thead>
<tr>
<th></th>
<th>Paediatric disease controls</th>
<th>Paediatric CF</th>
<th>Adult CF</th>
<th>p-value$^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects</td>
<td>45</td>
<td>136</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Age years</td>
<td>6 (0.8–21)</td>
<td>11 (0.2–20)</td>
<td>24 (22–42)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Female</td>
<td>18 (40)</td>
<td>65 (48)</td>
<td>6 (60)</td>
<td>0.45</td>
</tr>
<tr>
<td>Genotype</td>
<td>F508del/F508del</td>
<td>76 (56)</td>
<td>4 (40)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F508del/other</td>
<td>47 (35)</td>
<td>4 (40)</td>
<td></td>
</tr>
<tr>
<td>CF diagnosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Newborn screen¶</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meconium ileus¶</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conventional</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEV1 % pred (age ≥5 years)</td>
<td>91 [61–102] (n=5)</td>
<td>99 [38–129]  (n=97)</td>
<td>76.5 [47–94] (n=10)</td>
<td>0.12</td>
</tr>
<tr>
<td>Chronic <em>Pseudomonas</em> (&gt;50% cultures positive in past year)</td>
<td>31 (23)</td>
<td>7 (70)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antibiotics within 14 days of BALF</td>
<td>9 (20)</td>
<td>105 (77)</td>
<td>10 (100)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Pulmonary exacerbation score</td>
<td>5 (0–16) (n=129)</td>
<td>13 (0–16) (n=7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BALF culture results</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>30 (67)</td>
<td>41 (30)</td>
<td>1 (10)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td><em>Pseudomonas aeruginosa</em></td>
<td>1 (2)</td>
<td>25 (19)</td>
<td>5 (50)</td>
<td>0.01</td>
</tr>
<tr>
<td>MSSA</td>
<td>1 (2)</td>
<td>26 (19)</td>
<td>0</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>MRSA</td>
<td>0 (0)</td>
<td>21 (16)</td>
<td>3 (30)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td><em>Haemophilus influenzae</em></td>
<td>3 (7)</td>
<td>10 (7)</td>
<td>0 (0)</td>
<td>0.99</td>
</tr>
<tr>
<td><em>Stenotrophomonas maltophilia</em></td>
<td>1 (2)</td>
<td>20 (15)</td>
<td>1 (10)</td>
<td>0.02</td>
</tr>
<tr>
<td><em>Achromobacter xylosoxidans</em></td>
<td>0 (0)</td>
<td>5 (4)</td>
<td>0 (0)</td>
<td>0.33</td>
</tr>
<tr>
<td><em>Burkholderia</em></td>
<td>0 (0)</td>
<td>4 (3)</td>
<td>0 (0)</td>
<td>0.57</td>
</tr>
<tr>
<td>Nontuberculous mycobacteria</td>
<td>0 (0)</td>
<td>11 (8)</td>
<td>0 (0)</td>
<td>0.06</td>
</tr>
<tr>
<td>Bronchoscopic approach</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Endotracheal tube</td>
<td>5 (11)</td>
<td>29 (21)</td>
<td>9 (90)</td>
<td></td>
</tr>
<tr>
<td>Laryngeal mask airway</td>
<td>32 (71)</td>
<td>55 (40)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Nasal or oral route</td>
<td>7 (16)</td>
<td>11 (8)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Not recorded</td>
<td>1 (2)</td>
<td>41 (30)</td>
<td>1 (10)</td>
<td></td>
</tr>
</tbody>
</table>

Data are presented as n, median [range] or n (%), unless otherwise stated. CF: cystic fibrosis; FEV1: forced expiratory volume in 1 s; BALF: bronchoalveolar lavage fluid; MSSA: methicillin-susceptible *Staphylococcus aureus*; MRSA: methicillin-resistant *Staphylococcus aureus*. #: disease controls versus paediatric CF; ¶: some patients had positive newborn screening and meconium ileus.

https://doi.org/10.1183/13993003.00832-2017
BALF apart from *Haemophilus* (4% of disease controls). Notably, 20% of paediatric and adult CF BALF contained a dominant taxon with an unknown association to disease severity and progression in CF (termed here "nontraditional taxa") (*Streptococcus, Prevotella, Bordetella, Veillonella, Moraxella, Neisseria* and *Corynebacterium*). *Bordetella* was detected in three CF BALF samples in >50% relative abundance, all from participants considered to have a pulmonary exacerbation. Significant differences in the prevalence and relative abundance of taxa were observed between paediatric CF and disease control BALF (figure 2a and b, and supplementary table S5). Specifically, no disease control BALF sample contained >1% relative abundance of *Pseudomonas, Staphylococcus, Stenotrophomonas* or *Burkholderia*. Sequencing results are displayed in a heatmap in supplementary figure S2.

We examined the differences between microbiota communities based on age, culture positivity, pulmonary exacerbation status, CF centre and bronchoscopic approach (endotracheal tube, laryngeal mask airway or nasal) using PCoA. Culture-negative samples had communities that clustered together, although there was

![Graphs showing total bacterial load (TBL) and Shannon diversity index](https://doi.org/10.1183/13993003.00832-2017)

**FIGURE 1** a, b) Total bacterial load (TBL) and c, d) Shannon diversity index detected in bronchoalveolar lavage fluid (BALF) samples from cystic fibrosis (CF) and disease control participants. a) TBL for disease control, CF paediatric and CF adult BALF. b) TBL for disease control and CF paediatric BALF based on positive (+) or negative (−) bacterial culture results. c) Shannon diversity index for disease control, CF paediatric and CF adult BALF. d) Shannon diversity index for disease control and CF paediatric BALF based on (+) or negative (−) bacterial culture results. TBL was highest and Shannon diversity index lowest in CF BALF with positive cultures. TBL was measured on all BALF samples, whereas the Shannon diversity index was calculated only for samples with successful amplification and sequencing. Boxes show 25–75th interquartile range (IQR) with whiskers showing 1.5 times the IQR. Median indicated by a solid line in the box. Outliers are shown as individual data points. TBL limit of detection based on reagent control samples is indicated by the dashed line. *: p<0.05; **: p<0.01 for statistically significant differences.
overlap with culture-positive samples (supplementary figure S3). Communities from young patients tended to cluster together, although there was overlap between age groups. Communities did not cluster by pulmonary exacerbation status. We further examined the impact of collection approach on bacterial communities. Only 11 (8%) of the CF samples were collected via the nasal route and these were from the same CF centre; seven of these samples had adequate bacterial load for sequencing and had microbiota communities that clustered together by PCoA (by collection approach and by site). Clustering was not observed when comparing the laryngeal mask airway versus endotracheal tube approach or by other sites.

Relationship between CF BALF microbiota, airway inflammation and clinical characteristics
Airway inflammation was measured by BALF cell counts and results are given in supplementary table S6 and supplementary figure S4. In CF patients, lower diversity was associated with older age and increased airway inflammation, but not with forced expiratory volume in 1 s, potentially due to lack of spirometry data in children aged <6 years (figure 3). Streptococcus, Veillonella and Prevotella were associated with younger age and decreased markers of airway inflammation. In a multivariate regression analysis, the associations of Streptococcus, Veillonella and Prevotella with airway inflammatory markers remained when controlling for Pseudomonas relative abundance. Lower diversity was also associated with current antibiotic use (p=0.01). Review of antibiotic use at the time of BALF collection indicated that most CF participants (77% paediatric CF and 100% adult CF) were on at least one inhaled, oral or intravenous antibiotic (median (interquartile range) number of antibiotics 1 (1–4)), with 33 different antibiotics reported, limiting our ability to compare antibiotic effects.

Pulmonary exacerbations and microbiota
CF patients were categorised as clinically stable (n=44 (30%)), pulmonary exacerbation (n=90 (62%)) or unknown (n=12 (8%)). The most common indications recorded for clinically stable CF participants undergoing bronchoscopy (participants may have had more than one indication) included concurrent surgery (55%), physician-determined need for lower airway culture (20%), part of routine clinical CF care (18%) and no indication provided (7%). P. aeruginosa relative abundance and percentage neutrophils were higher in those with pulmonary exacerbation compared with clinically stable participants (supplementary table S7). We did not detect a significant difference in TBL or Shannon diversity index based on pulmonary exacerbation status (figure 4a and b). Total cell count and absolute neutrophils between groups also did not differ (data not shown). Microbiota communities also did not cluster by pulmonary exacerbation status using PCoA (supplementary figure S3).

### TABLE 2 Dominant taxa identified by sequencing in bronchoalveolar lavage fluid (BALF) samples from disease control and cystic fibrosis (CF) paediatric and adult participants

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Disease control</th>
<th>Paediatric CF</th>
<th>Adult CF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative (no sequence data)</td>
<td>45</td>
<td>136</td>
<td>10</td>
</tr>
<tr>
<td>Achromobacter+&lt;sup&gt;#&lt;/sup&gt;</td>
<td>0</td>
<td>6 (4)</td>
<td>0</td>
</tr>
<tr>
<td>Bordetella</td>
<td>0</td>
<td>2 (1.5)</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Burkholderia</td>
<td>0</td>
<td>2 (1.5)</td>
<td>0</td>
</tr>
<tr>
<td>Corynebacterium</td>
<td>0</td>
<td>1 (1)</td>
<td>0</td>
</tr>
<tr>
<td>Haemophilus+&lt;sup&gt;#&lt;/sup&gt;</td>
<td>2 (4)</td>
<td>11 (8)</td>
<td>0</td>
</tr>
<tr>
<td>Moraxella</td>
<td>1 (2)</td>
<td>1 (1)</td>
<td>0</td>
</tr>
<tr>
<td>Mycoplasma</td>
<td>1 (2)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Neisseria</td>
<td>0</td>
<td>1 (1)</td>
<td>0</td>
</tr>
<tr>
<td>Prevotella</td>
<td>2 (4)</td>
<td>7 (5)</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Pseudomonas+&lt;sup&gt;+&lt;/sup&gt;</td>
<td>0</td>
<td>14 (10)</td>
<td>4 (40)</td>
</tr>
<tr>
<td>Staphylococcus+&lt;sup&gt;+&lt;/sup&gt;</td>
<td>0</td>
<td>15 (11)</td>
<td>2 (20)</td>
</tr>
<tr>
<td>Stenotrophomonas+&lt;sup&gt;+&lt;/sup&gt;</td>
<td>0</td>
<td>16 (12)</td>
<td>0</td>
</tr>
<tr>
<td>Streptococcus</td>
<td>6 (13)</td>
<td>12 (9)</td>
<td>0</td>
</tr>
<tr>
<td>Veillonella</td>
<td>0</td>
<td>2 (1.5)</td>
<td>0</td>
</tr>
</tbody>
</table>

Data are presented as n or n (%). <sup>+</sup>: taxa associated with typical CF pathogens. <sup>+</sup>: taxa associated with typical CF pathogens. "Pseudomonas+" contains sequences assigned to the taxa Pseudomonadales and Pseudomonas. "Achromobacter+" contains sequences assigned to the taxa Alcaligenaceae and Achromobacter.

https://doi.org/10.1183/13993003.00832-2017
**FIGURE 2** a) Prevalence (% of samples with specific taxa detected) and b) relative abundance of the most frequently detected taxa and taxa associated with traditional cystic fibrosis (CF) pathogens from disease control (DC) and paediatric CF bronchoalveolar lavage fluid (BALF). a) Prevalence of bacterial taxa detected from CF and disease control BALF. Most taxa were detected at higher prevalence in CF compared with disease control BALF. b) Median relative abundance of bacterial taxa associated with nontraditional and traditional CF bacteria detected from CF and disease control BALF. *Streptococcus, Prevotella, Veillonella, Neisseria* and *Porphyromonas* were detected at higher relative abundance in disease control compared with CF BALF. Examining traditional CF taxa, *Staphylococcus* was detected at higher relative abundance in CF compared with disease control BALF. Although not statistically significant, taxa associated with *Pseudomonas*, *Stenotrophomonas, Burkholderia* and *Achromobacter* had a wider range of relative abundance in CF compared with disease controls. *"Pseudomonas"* contains sequences assigned to the taxa *Pseudomonadales* and *Pseudomonas*. "Achromobacter"* contains sequences assigned to the taxa *Alcaligenaceae* and *Achromobacter*. Median relative abundance was calculated based on BALF samples positive for a given taxa. Relative abundance is displayed as median (line), boxes indicate 25–75th interquartile range, and whiskers extend from minimum and maximum values. *: p<0.05; **: p<0.01 for statistically significant differences.

**FIGURE 3** Heatmap of associations with clinical characteristics, microbiota and airway inflammation in cystic fibrosis patients. Spearman’s correlation coefficients for total bacterial load, diversity and relative abundance of bacterial taxa related to age, lung function (forced expiratory volume in 1 s [FEV1] % pred) and bronchoalveolar lavage fluid cell counts (total white blood cells, absolute neutrophil count and percentage neutrophils). *"Pseudomonas"* contains sequences assigned to the taxa *Pseudomonadales* and *Pseudomonas*. "Achromobacter"* contains sequences assigned to the taxa *Alcaligenaceae* and *Achromobacter*. Red/brown: positive correlations; blue/black: negative correlations. For example, lower diversity was associated with older age and higher markers of inflammation. Higher relative abundance of *Prevotella, Streptococcus* and *Veillonella* was associated with younger age and less inflammation. Positive (up) and negative (down) correlations are also indicated by the direction of the arrowhead symbol in the heatmap. Correlation values >0.5 have p-values <0.05.
Microbiota and age in CF

Given the differences in airway infection known to occur by age in CF, we sought to examine the microbiota by age group [39]. Patient characteristics within age groups (<2, 2–5, 6–10, 11–17, 18–24 and ≥25 years) are shown in supplementary table S8. TBL did not differ across age groups, whereas diversity was higher in younger age groups (figure 4c and d). Taxa associated with CF, most prominently Pseudomonas, were present in higher median relative abundance in older age groups (figure 5a). Conversely, Streptococcus was highest in young children (aged <6 years) in whom it was present at >20% relative abundance, compared with those aged >10 years in whom median relative abundance decreased to <3%. By age 6 years, a median of 70% of the bacterial communities consisted of traditional CF taxa (Pseudomonas, Staphylococcus, Haemophilus, Stenotrophomonas and Burkholderia) (figure 5b). TBL and diversity for disease control patients by age are shown in the supplementary data section; although the numbers were small, the difference in diversity by age was not seen in disease controls as it was in CF (supplementary figure S5).

We specifically examined taxa reported in prior CF microbiota and expanded culture analyses [40–42]. Rothia was detected with median relative abundance of 0.15%, ranging from 0.97% in those aged <2 years.
to 0% in those aged ≥18 years. *Gemella* was detected with median relative abundance of 0.15%, ranging from 0.55% in those aged <2 years to 0% in those aged 18–24 years, and 0.19% in those aged ≥25 years.

*Enterobacteriaceae* was detected with median relative abundance 1.5% in those aged <2 years and 0.3% in those aged 2–5 years, but was not detected from any participant aged >11 years.

*Mycobacterium* was detected with median relative abundance of 0.01%.

Comparisons with culture

Results of comparisons between sequencing and culture are shown in table 3. There were 42 CF BALF samples with negative cultures, of which 32 did not amplify. For the 10 out of 42 (24%) culture-negative specimens with sequencing data, two contained >98% *Haemophilus* and the remaining had mixed anaerobic taxa. *Prevotella* was present in six out of 10 samples with >10% relative abundance and five had *Pseudomonas*, albeit in low abundance (0.006–1.13% relative abundance). Relative abundance and TBL data from all 10 culture-negative, sequence-detectable BALF samples are shown in supplementary figure S6.

![Cross-sectional plots of major taxa detected by sequencing by age group in cystic fibrosis (CF).](https://doi.org/10.1183/13993003.00832-2017)

**FIGURE 5** Cross-sectional plots of major taxa detected by sequencing by age group in cystic fibrosis (CF). a) Median relative abundance, based on samples with detectable taxa, for *Pseudomonas+*, *Stenotrophomonas*, *Streptococcus*, *Prevotella*, *Staphylococcus*, *Haemophilus*, *Burkholderia* and *Veillonella*. b) Combined median relative abundance for traditional CF taxa (*Pseudomonas+*, *Staphylococcus*, *Stenotrophomonas*, *Haemophilus* and *Burkholderia*) and taxa not traditionally identified from culture (*Streptococcus*, *Neisseria*, *Porphyromonas*, *Prevotella* and *Veillonella*). "*Pseudomonas*+" contains sequences assigned to the taxa *Pseudomonadales* and *Pseudomonas*.

<table>
<thead>
<tr>
<th></th>
<th>BALF culture-positive</th>
<th>BALF sequencing-positive</th>
<th>BALF culture- and sequencing-positive (sensitivity)</th>
<th>BALF sequencing-positive but culture-negative</th>
<th>Relative abundance of taxa detected by sequencing but not culture</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Pseudomonas</em>+</td>
<td>30 [21]</td>
<td>73 [60]</td>
<td>28 [93]</td>
<td>45 [31]</td>
<td>0.10 [0.01–2.61]</td>
</tr>
<tr>
<td><em>Staphylococcus</em></td>
<td>49 [34]</td>
<td>79 [54]</td>
<td>39 [80]</td>
<td>40 [30]</td>
<td>0.05 [−0.01–1.52]</td>
</tr>
<tr>
<td><em>Achromobacter</em>+</td>
<td>5 [3]</td>
<td>38 [26]</td>
<td>4 [80]</td>
<td>35 [24]</td>
<td>0.06 [0.01–99.49]</td>
</tr>
<tr>
<td><em>Burkholderia</em></td>
<td>4 [3]</td>
<td>17 [12]</td>
<td>3 [75]</td>
<td>14 [10]</td>
<td>0.02 [−0.01–0.45]</td>
</tr>
<tr>
<td><em>Stenotrophomonas</em></td>
<td>21 [14]</td>
<td>65 [45]</td>
<td>19 [99]</td>
<td>47 [32]</td>
<td>0.07 [−0.01–68.97]</td>
</tr>
<tr>
<td><em>Haemophilus</em></td>
<td>10 [7]</td>
<td>66 [46]</td>
<td>10 [100]</td>
<td>57 [39]</td>
<td>0.25 [0.01–99.43]</td>
</tr>
<tr>
<td><em>Mycobacterium</em></td>
<td>11 [8]</td>
<td>5 [3]</td>
<td>2 [18]</td>
<td>3 [2]</td>
<td>0.06 [0.01–0.22]</td>
</tr>
</tbody>
</table>

Data are presented as n (%) or median [range]. BALF: bronchoalveolar lavage fluid. Sequencing has a higher sensitivity than standard culture to detect low quantities of most taxa other than *Mycobacterium*. One BALF sample did not have culture results reported; thus, data from 145 BALF specimens are presented here. BALF samples with bacterial DNA that did not amplify adequately for sequencing were assumed to be negative. "*Pseudomonas*+" contains sequences assigned to the taxa *Pseudomonadales* and *Pseudomonas*. "*Achromobacter*+" contains sequences assigned to the taxa *Alcaligenaceae* and *Achromobacter*.
Discussion

In this large, multicentre study of CF BALF, we found that airway microbiota differed substantially by patient age. *Streptococcus* was most prominent in young children with CF. Traditional CF taxa (*e.g.* *Pseudomonas*, *Staphylococcus* and *Stenotrophomonas*) comprised <50% of the bacterial community in children aged <2 years. In older patients, traditional CF bacteria were predominant, constituting >70% of the community in children aged >6 years and adults. Relative abundance of *Streptococcus*, *Veillonella*, *Prevotella* and *Haemophilus* was inversely associated with age, whereas *Pseudomonas* was positively associated with age. Diversity was lower in older age groups due to the predominance of a single taxon. Despite the difference in microbiota taxa and diversity, TBL did not vary significantly between age groups. Although traditional CF taxa most often dominated microbial communities in older age groups, bacterial taxa not traditionally associated with CF (*e.g.* *Prevotella* and *Streptococcus*) were the dominant community members in 20% of both paediatric and adult CF BALF.

The clinical impact of nontraditional taxa, particularly anaerobic bacteria, is still under investigation. Our group has previously shown that airway inflammation is increased in the presence of typical CF pathogens, particularly *Pseudomonas*, whereas anaerobic taxa appear less pro-inflammatory [40, 43]. In this study, we found increased airway inflammation in CF compared with disease control BALF, including in samples without pathogens detected by culture. Nontraditional taxa (*e.g.* communities with higher relative abundance of *Streptococcus*, *Prevotella* and *Veillonella*) were associated with less inflammation with lower total cell counts and fewer neutrophils. However, these taxa should not be construed to be entirely benign as inflammation was increased in all CF BALF compared with disease control BALF, suggesting that anaerobic infections may be inciting an inflammatory response even if blunted compared with those with typical CF pathogens. In addition, these taxa contain a diverse group of bacteria; thus, it is possible that some members of these groups may play a role in lung disease, particularly if present as the dominant bacteria [44, 45]. Communities dominated by mixed anaerobic taxa may be a marker of earlier and/or less severe disease that incites an inflammatory response that ultimately selects for typical, highly adaptive CF pathogens. The inflammatory potential of mixed anaerobic infections and other polymicrobial infections is supported by previous studies in CF and non-CF subjects [46–48]. Increased inflammation may also be related to underlying CF transmembrane conductance regulator dysfunction in the absence of infectious triggers. However, our data showing the presence of complex microbiota within airway samples suggest that these bacterial communities may be contributing, particularly in early disease, to the inflammatory response. We did not detect an association between *Pseudomonas* relative abundance and inflammation, although CF BALF samples with positive microbiological cultures had higher inflammation compared with those with negative cultures. This finding may be due to the large number of BALF samples with detectable but relatively low relative abundance of *Pseudomonas*.

Compared with non-CF paediatric disease controls, BALF from participants with CF displayed higher bacterial loads, lower diversity and distinct community structures. Comparing CF and disease control BALF, similar taxa were detected from both groups; however, there were distinct differences in the range of relative abundance of taxa within bacterial communities. These findings suggest that exposure of the airways to traditional CF taxa and anaerobes may be common in CF and non-CF patients, but that the host environment or exposure to antibiotics may alter the ability of these microbes to survive within the community. Although our numbers were relatively small, diversity was not associated with age in disease control patients as it was in the CF group. The inclusion of a control group provides important information about the relative bacterial load and community structure between children with CF and those with other diseases. However, interpretation of the results is limited by the heterogeneity of underlying diseases in our control group and the relatively small number of disease control BALF samples that sequenced successfully (likely related to the reduced bacterial burden).

Sequencing detected bacterial communities from CF and disease control BALF samples with negative cultures. The majority of these bacterial communities contained fastidious bacteria or bacteria typically considered part of the upper respiratory microbiota and thus not always speciated by clinical microbiology laboratories. However, *Haemophilus* was detected as the dominant taxa in two CF and one disease control sample with negative cultures, and low relative abundance of *Pseudomonas* was detected in four CF samples. Sequencing also detected *Pseudomonas*, *Burkholderia* and *Stenotrophomonas* from samples negative for these bacterial pathogens by standard culture technique. These findings suggest that a combined approach using culture and molecular approaches may improve microbiological diagnostics in the future, although additional studies are needed to understand the clinical importance of positive sequencing in the setting of negative cultures. Sequencing lacked sensitivity for detection of nontuberculous mycobacteria, consistent with other studies [49].

Most previous research on the airway microbiota in CF has focused on sputum or oropharyngeal specimens rather than lower airway BALF samples. COBURN et al. [50] studied sputum samples from 269
However, we found strong agreement between culture and sequencing status, with instilled systematically biased our results. BALF samples were also stored neat rather than processed; thus, the bacterial load, although we did not see an association with age that might be expected if the volume phylogenetic resolution. BALF collection was performed with differing volumes, which may have impacted CF flora may improve early treatment.

The authors thank the research coordinator, Meg Anthony, and database manager, Elinor Towler, at Children’s Hospital Colorado, Aurora, CO, USA, and site research coordinators: Colette Bucur (Rainbow Babies and Children’s Hospital, Cleveland, OH, USA), Karen Callahan (Johns Hopkins University, Baltimore, MD, USA), Lorrie Duan (Cincinnati Children’s Hospital Medical Center, Cincinnati, OH, USA), Elizabeth Hartigan (Children’s Hospital of Pittsburgh, Pittsburgh, PA, USA), Terri Johnson (Nationwide Children’s Hospital, Columbus, OH, USA), Vicki Kocida (St Louis University School of Medicine, St Louis, MO, USA), Caroline O’Connor LaFaye (University of North Carolina, Chapel Hill, NC, USA), Sharon McNamara (Seattle Children’s Hospital, Seattle, WA, USA), Valerie Eubanks Tarn (University of Alabama, Birmingham, AL, USA), Mary Teresi (University of Iowa, Iowa City, IA, USA), Noopur Singh and
Catherine Powers (Ann and Robert H. Lurie Children’s Hospital of Chicago, Chicago, IL, USA) and Trish Grover (University of Minnesota, Minneapolis, MN, USA), and the patients and families who contributed data and samples to make this work possible. The authors also thank Jill VanDalfsen and Bonnie Ramsey (Therapeutics Development Network Coordinating Center, Seattle, WA, USA) for support of this project.

Author contributions: E.T. Zemanick contributed to the study design, analysis and interpretation of data, and wrote the manuscript. B.D. Wagner contributed to the data analysis and statistical interpretation, and participated in writing and review of the manuscript. C.E. Robertson contributed to the data analysis and interpretation, and reviewed the manuscript. R.C. Ahrens, J.F. Chmiel, J.P. Clancy, R.L. Gibson, W.T. Harris, G. Kurland, T.A. Laguna, S.A. McColl, K. McCoy, G. Retsh-Bogart, K.T. Sobush and P.L. Zeitlin all served as site principle investigators, contributed to the study design and reviewed the manuscript. M.J. Stevens contributed to sample processing and analysis, and reviewed the manuscript. F.J. Accurso and S.D. Sagel contributed to the study design, data interpretation, and reviewed the manuscript. J.K. Harris contributed to the study design, sample analyses, data analysis and interpretation, and participated in writing and review of the manuscript.

References


