

Molecular identification of bacteria in bronchoalveolar lavage fluid from children with cystic fibrosis

J. Kirk Harris^{*†}, Mary Ann De Groote^{*†}, Scott D. Sagel^{†§}, Edith T. Zemanick^{†§}, Robert Kapsner[§], Churee Penvari[§], Heidi Kaess[§], Robin R. Deterding^{†§}, Frank J. Accurso^{†§}, and Norman R. Pace^{*†¶}

^{*}Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, CO 80309; [†]Department of Pediatrics, School of Medicine, University of Colorado and Health Sciences Center, Aurora, CO 80045; and [§]Mike Morris Cystic Fibrosis Research and Treatment Center, The Children's Hospital, Denver, CO 80218

Contributed by Norman R. Pace, October 15, 2007 (sent for review August 27, 2007)

Culture of bronchoalveolar lavage fluid (BALF) is the gold standard for detection of pathogens in the lower airways in cystic fibrosis (CF). However, current culture results do not explain all clinical observations in CF, including negative culture results during pulmonary exacerbation and inflammation in the absence of pathogens. We hypothesize that organisms not routinely identified by culture occur in the CF airway and may contribute to disease. To test this hypothesis we used a culture-independent molecular approach, based on use of rRNA sequence analysis, to assess the bacterial composition of BALF from children with CF and disease controls (DC). Specimens from 42 subjects (28 CF) were examined, and ~6,600 total clones were screened to identify 121 species of bacteria. In general, a single rRNA type dominated clone libraries from CF specimens, but not DC. Thirteen CF subjects contained bacteria that are not routinely assessed by culture. In four CF subjects, candidate pathogens were identified and include the anaerobe *Prevotella denticola*, a *Lysobacter* sp., and members of the *Rickettsiales*. The presumptive pathogens *Tropheryma whipplei* and *Granulicatella elegans* were identified in cases from the DC group. The presence of unexpected bacteria in CF may explain inflammation without documented pathogens and consequent failure to respond to standard treatment. These results show that molecular techniques provide a broader perspective on airway bacteria than do routine clinical cultures and thus can identify targets for further clinical evaluation.

pulmonary microbiology | ribosomal RNA

Cystic fibrosis (CF), the most common autosomal lethal disease in Caucasians, is caused by mutation in the CF transmembrane conductance regulator gene, which results in a generalized exocrinopathy (1). Thickened secretions caused by improper regulation of airway surface liquid contribute to the accumulation of mucus in the airway and defective mucociliary clearance (2). Retained mucus plugs provide a niche for bacterial colonization and persistence. Most morbidity and mortality associated with CF is attributed to microbial infections in the airway and the persistent inflammatory response (3).

A small number of pathogens are recognized traditionally in CF airway disease. These include *Pseudomonas aeruginosa*, *Staphylococcus aureus*, *Haemophilus influenzae*, and the *Burkholderia cepacia* complex (4–7). Other bacteria, including *Stenotrophomonas maltophilia* and *Achromobacter xylosoxidans*, also are associated with CF, but there is controversy as to their roles as pathogens (8–10). This view of bacteria associated with CF is based on culture, however, which may not recover or identify all bacteria present in a specimen. Although known pathogens clearly are important in airway disease, a more comprehensive picture of the bacterial community in the CF airway potentially could lead to improved insight about the disease and thus better treatment. For instance, “normal” microbiota that are commonly present, but not considered pathogenic, may affect the disease course (11). In addition,

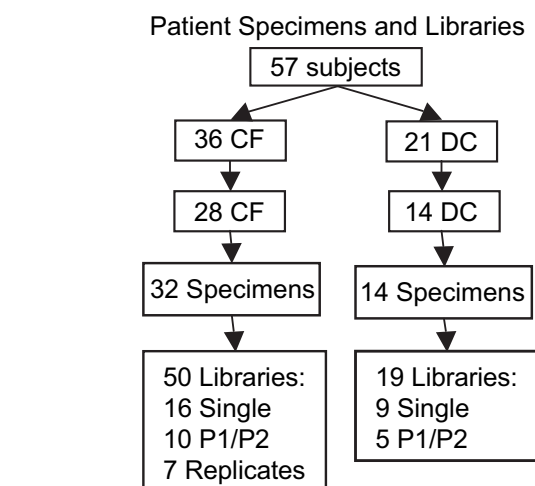


Fig. 1. Flow chart of specimens and PCR results for CF and DC specimens. Not all specimens were positive for PCR. For each specimen one to three processing types were examined. In 22 cases multiple libraries were constructed from the same specimen. Fifteen specimens were positive for PCR with both the hard- and soft-spin pellets and are labeled P1/P2. Replicate libraries were made in the remaining seven cases and are labeled replicates. Specimens with a single library constructed are labeled single.

there are documented cases of inflammation with no cultured pathogen (12–14).

Developments in environmental microbiology have obviated, to considerable extent, the requirement to culture microbes to detect and identify them (15). Instead of culture tests, rRNA genes are harvested by PCR directly from samples, and their sequences are used to identify bacteria in the sample. In the current study, we have used this culture-independent molecular technique to determine the composition of bacterial communities associated with bronchoalveolar lavage fluid (BALF) from CF and disease control (DC)

Author contributions: J.K.H., F.J.A. and N.R.P. designed research; J.K.H. and M.A.D.G. performed research; S.D.S., E.T.Z., R.K., C.P. and R.R.D. contributed new reagents/analytic tools; J.K.H., H.K. and N.R.P. analyzed data; and J.K.H. and N.R.P. wrote the paper.

The authors declare no conflict of interest.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AY805987–AY806453, DQ188268–DQ188805, and EU111806–EU112454).

[¶]Present address: Department of Microbiology, Immunology, and Pathology, Colorado State University, Fort Collins, CO 80523.

^{¶¶}To whom correspondence should be addressed. E-mail: norman.pace@colorado.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0709804104/DC1.

© 2007 by The National Academy of Sciences of the USA

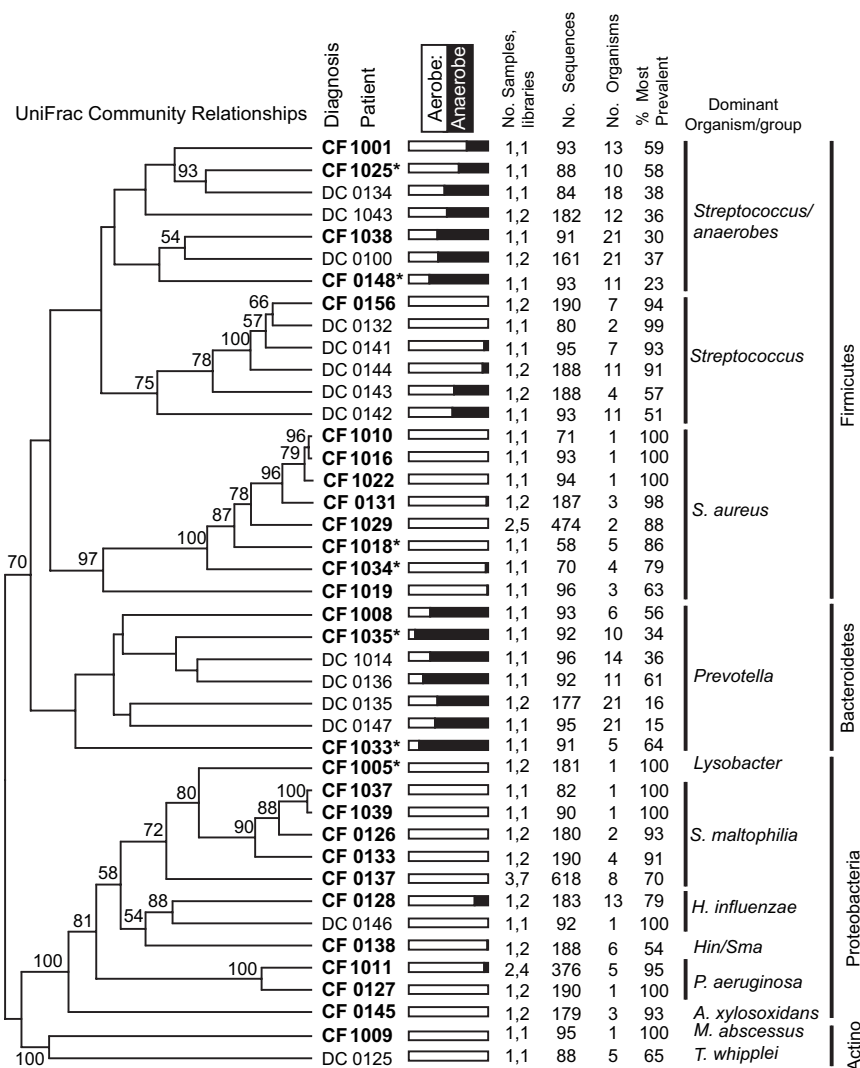


Fig. 2. Bacterial community comparisons of BALF-derived PCR libraries were determined with weighted UniFrac, and the relationships are displayed by the unweighted pair group method with arithmetic mean clustering algorithm. The numbers in the dendrogram represent the jackknife support for the labeled nodes. Cohort and patient numbers were used to label each subject. Asterisks are used to mark CF subjects with samples that did not yield any pathogenic isolates. The bars represent the percentage of clones indicative of aerobic organisms (open) and anaerobic organisms (filled). The number of samples, specimens, clones identified by sequence (including RFLP counts), organisms, and prevalence of the most common organism identified are also given. The tree is labeled by phylogenetic identity (genus and division) of the most common organism identified in each subject.

subjects. We compare the results to those obtained by culture and conclude that molecular analysis provides a more comprehensive view of the microbiology of the lower airway than is obtained by standard laboratory culture.

Results

Molecular Summary. A total of 69 rRNA gene libraries (50 CF) were prepared with bacteria-specific PCR amplicons obtained from 42 subjects (28 CF; Fig. 1). Replicate libraries from the same sample were constructed from five CF subjects to examine the reproducibility of the procedures, and results were highly similar in all cases. Libraries prepared from three CF subjects and two DC subjects contained sequences of uncertain source, more similar to water samples (e.g., content of *Methylobacterium* spp.) than to other clinical samples and were excluded from the study. Overall, $\approx 6,600$ clones were screened by restriction fragment length polymorphism (RFLP) to identify unique types for sequence analysis and obtain a census of the different sequence types in each library. Approximately $\approx 1,600$ unique sequences were determined, each ≈ 800 nt in

length. Sequences identified 121 bacterial species (65 in CF) with 33 species observed in both groups.

The molecular microbiology among the specimen sets was complex in makeup, so we grouped the results according to the similarities of the sets by using UniFrac. Results are summarized in Fig. 2. The overall diversity observed in CF was distinct from that seen in the DC (Fig. 2). The most frequently encountered sequences in CF were members of relatively few species, generally less complex than clone libraries from DC subjects [summarized in Fig. 3 and [supporting information \(SI\) Table 1](#)].

Libraries from CF Subjects. Although variable in composition between subjects, libraries from CF subjects generally were dominated ($>85\%$ of total) by sequences characteristic of known CF pathogens (17/28, 61%). About half of the 28 CF subjects contained a single predominant pathogen ($>85\%$ of library; Fig. 1). Two CF pathogens dominated two additional subjects. The remaining subjects (11/28, 39%) yielded sequences indicative of organisms not routinely identified in the airway of subjects with

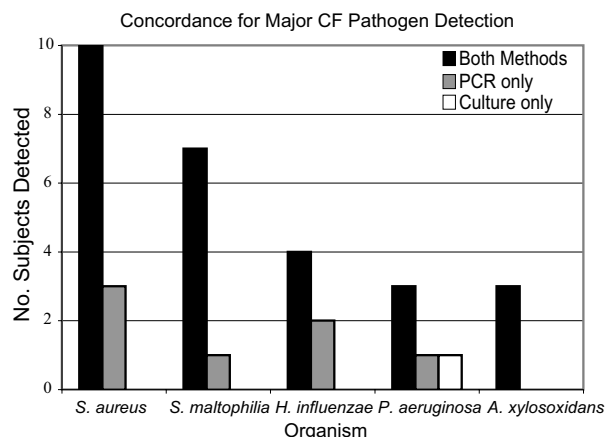


Fig. 5. Summary of major CF pathogens detected by both culture and molecular analysis with all-bacteria primers. Dark bars represent detection by both methods, gray bars represent detection by molecular methods only, and empty bars represent detection by culture only.

The remaining two subjects with low-complexity libraries from the DC contained unexpected sequences that may indicate microbial disease. In one subject the majority of the sequences from the library (66%) were indicative of *Tropheryma whippelii*, the causative agent of Whipple's disease. In addition, *Moraxella catarrhalis* and a novel member of the *Rickettsiales* (95% identification with sequences in Fig. 4D) were present in the library (Fig. 4E). The second subject contained sequences related to *Granulicatella elegans* and *Streptococcus mitis* group (Fig. 4F).

Comparison to Culture. Clinical culture results were compared with molecular results and are summarized in Fig. 5. Only 3 of the 14 DC subjects were culture positive for recognized CF bacterial pathogens. In contrast, 22 (79%) of the CF specimens were culture positive for such organisms. The molecular survey detected the cultured CF pathogen in 27 of 28 subjects (96%). In seven subjects (25%) the rDNA sequence analysis, but not culture, detected CF pathogens (Fig. 5). The relatively low numbers of subjects with *P. aeruginosa* was of note. We looked at the culture history in the year before bronchoscopy to determine why the incidence was lower than expected in these subjects. *P. aeruginosa* was detected in 12/28 subjects (43%) in the prior year (SI Table 2).

Discussion

This and previous studies show that the sequence-based molecular approach to clinical microbiology provides microbiological perspective that can significantly enhance the standard clinical culture-based view (16–19). The use of sequences for identification in this study is a substantial improvement over electrophoretic patterns for the identification of microbes, because there is no ambiguity in identity. The two patient cohorts examined, CF and DC, were not age-matched, which may limit some comparisons between the two cohorts. Moreover, we acknowledge that the CF cohort is small and certainly does not represent the full spectrum of the disease. Nevertheless, this study is to date the most extensive sequence-based analysis of bacteria associated with the human airway to our knowledge.

Molecular results with most specimens matched the culture results with regard to conventionally recognized pathogens. In addition, several specimens from CF contained sequences that were unexpected and possibly identify new pathogens. In other CF specimens, likely pathogens (*P. denticola*, *Coxiellaceae* and *Rickettsiales*; Fig. 4) were observed along with known pathogens. Expanded studies will be required to test the specific association of

these potentially novel pathogens with CF and determine clinical manifestations.

One technical issue of concern in sampling the airway is the possibility of oral or upper respiratory contamination of the specimen as the bronchoscope is advanced into the lung. The sequence-based methods used here, with general bacterial PCR primers, sample the most abundant organisms present in specimens. Oral and upper airway suites of microbes are more complex and generally different phylogenetically from those we encountered in BALF specimens (20). We cannot rule out low-incidence sequences as upper airway contaminants, but the types and relatively low complexity of sequences in these BALF specimens both support their specificity to the lower airway.

Whether or not there are clinical consequences of anaerobic bacteria in the airways of CF subjects is an open question. It is not clear whether BAL consistently samples mucus from the airways; in many cases, the BALF remains clear without any significant mucus plugs present. Thus, anaerobic bacteria, which are expected to occur mainly in mucus plugs, may often be present but not sampled. There were no distinguishing features among the CF subjects in this study that contained high levels of anaerobic bacteria. A larger study will be required to elucidate any clinical features that predict involvement of anaerobic bacteria in exacerbation.

In addition to the results from CF specimens, several bacteria with potential clinical significance were detected in DC. In the single case of a lung abscess analyzed, culture had failed to identify any bacteria, but the molecular method detected *Granulicatella elegans* as 79% of the clones in the library prepared from the specimen (Fig. 4F). This organism previously has been associated with endocarditis (21). In a second case, diagnosed as pneumonia, culture results reported the presence of α -hemolytic *Streptococcus* (not *S. pneumoniae*). The sequence from this specimen, however, shared 99.5% sequence identity with those obtained from a CF specimen where *S. pneumoniae* was identified. This result demonstrates the utility of the molecular perspective in identification of organisms even in the context of uncertainty in culture-based identification. The result also demonstrates the utility of the molecular methods to address bacteria associated with pulmonary disease other than CF.

Tropheryma whippelii, the causative agent of Whipple's disease (22), was the most frequently encountered sequence in one of the DC specimens from an interstitial lung disease (ILD) subject. In addition, sequences belonging to the *Rickettsiales* group were present at lower levels in this specimen. Neither of these organisms is detected by standard culture. It is possible that the occurrence of one or both organisms may correlate with a particular subset of ILD patients and thereby provide diagnostic information relevant to therapy.

rRNA sequences obtained from environmental (including clinical) samples seldom are identical to previously encountered sequences. In this study, sequences were assigned to named taxa based on sequence similarities (approximately $\geq 97\%$ sequence identity). Although organisms with similar sequences are expected to share many physiological properties, we note that small differences in rRNA sequences may not reflect genomic-level differences with significant clinical considerations, for instance drug resistance vs. sensitivity. The primary goal of rRNA sequence analysis is to identify bacteria in the clinical specimens, and it does so unequivocally. Determination of physiological properties such as antibiotic resistance requires information beyond the specificity of rRNA sequences, either molecular- or culture-based results.

Methods based on bacterial gene sequences can complement standard clinical microbiology for characterization of bacteria associated with disease states and are increasingly economical. rRNA gene sequences are the most widely used molecular measure for bacterial identification. Consequently, an rRNA sequence represents an important first evaluation of an organism, its basic nature

