The Role of Bacteria in the Pathogenesis and Progression of Idiopathic Pulmonary Fibrosis

Phillip L. Molyneaux1,2, Michael J. Cox1, Saffron A. G. Willis-Owen1, Patrick Mallia1,3, Kirsty E. Russell1, Anne-Marie Russell1,2, Elissa Murphy4, Sebastian L. Johnston1,3, David A. Schwartz4, Athol U. Wells2, William O. C. Cookson1,2*, Toby M. Maher1,2*, and Miriam F. Moffatt1,2*

1National Heart and Lung Institute, Imperial College, London, United Kingdom; 2National Institute for Health Research Respiratory Biomedical Research Unit, Royal Brompton Hospital, London, United Kingdom; 3Imperial College Healthcare National Health Service Trust, London, United Kingdom; and 4Department of Medicine, University of Colorado, Denver, Colorado

Abstract

Rationale: Idiopathic pulmonary fibrosis (IPF) is a progressive lung disease of unknown cause that leads to respiratory failure and death within 5 years of diagnosis. Overt respiratory infection and immunosuppression carry a high morbidity and mortality, and polymorphisms in genes related to epithelial integrity and host defense predispose to IPF.

Objectives: To investigate the role of bacteria in the pathogenesis and progression of IPF.

Methods: We prospectively enrolled patients diagnosed with IPF according to international criteria together with healthy smokers, nonsmokers, and subjects with moderate chronic obstructive pulmonary disease as control subjects. Subjects underwent bronchoalveolar lavage (BAL), from which genomic DNA was isolated. The V3–V5 region of the bacterial 16S rRNA gene was amplified, allowing quantification of bacterial load and identification of communities by 16S rRNA quantitative polymerase chain reaction and pyrosequencing.

Measurements and Main Results: Sixty-five patients with IPF had double the burden of bacteria in BAL fluid compared with 44 control subjects. Baseline bacterial burden predicted the rate of decline in lung volume and risk of death and associated independently with the rs35709950 polymorphism of the MUC5B mucin gene, a proven host susceptibility factor for IPF. Sequencing yielded 912,883 high-quality reads from all subjects. We identified Haemophilus, Streptococcus, Neisseria, and Veillonella spp. to be more abundant in cases than control subjects. Regression analyses indicated that these specific operational taxonomic units as well as bacterial burden associated independently with IPF.

Conclusions: IPF is characterized by an increased bacterial burden in BAL that predicts decline in lung function and death. Trials of antimicrobial therapy are needed to determine if microbial burden is pathogenic in the disease.

Keywords: idiopathic pulmonary fibrosis; Muc5b; bacteria; microbiome

Idiopathic pulmonary fibrosis (IPF) is a progressive and fatal disease of unknown cause (1). It is increasing in prevalence (2), has a median survival from diagnosis of 3 years, and carries a worse life expectancy than some cancers (3). IPF occurs primarily in older adults, many of whom have been smokers, and polymorphisms in genes related to epithelial integrity and host...
At a Glance Commentary

Scientific Knowledge on the Subject: Idiopathic pulmonary fibrosis (IPF) is a progressive and fatal disease of unknown cause. Active infection in IPF is known to be associated with high morbidity and mortality, and immunosuppression is deleterious. Although the role of viruses in the pathogenesis and progression of IPF has been investigated, the role of bacteria has not been studied in detail.

What This Study Adds to the Field: Patients with IPF have an increased pulmonary bacterial load compared with matched control subjects, and the load at the time of diagnosis predicts rapidly progressive IPF and an increased risk of mortality. Clinical trials of antimicrobial therapy are needed to determine if microbial burden is causal or not in IPF progression.

We therefore investigated the role of bacteria in the pathogenesis and progression of IPF in a substantial prospective case-control study. As IPF often appears against a background of smoking-related lung disease, we included patients with chronic obstructive pulmonary disease (COPD) and matched numbers of smokers in our control groups. We used sequence-based culture-independent methodologies to quantify both the numbers of bacterial genomes and the community composition of specimens collected by bronchoalveolar lavage. Some of the results of these studies have been previously reported in the form of an abstract (18).

Methods

Study Design
Patients undergoing diagnostic bronchoscopy with bronchoalveolar lavage (BAL) for suspected but previously undiagnosed IPF were prospectively recruited between November 2010 and January 2013. A diagnosis of IPF was made after multidisciplinary team discussion. Only individuals fulfilling currently accepted international criteria were subsequently included in the study (19). Control subjects included nonsmokers and smokers with normal lung function (referred to throughout as healthy control subjects) and individuals with moderate (Global Initiative for Chronic Obstructive Lung Disease stage II) COPD and were recruited separately using the same protocols. Subjects were excluded if they had a history of self-reported upper or lower respiratory tract infection, antibiotic use in the prior 3 months, acute IPF exacerbation, or other respiratory disorders. Written informed consent was obtained from all subjects, and the study was approved by the local Research Ethics Committee (Ref 10/H0720/12, 00/BA/459E, and 07/H0712/138).

Bronchoscopy
Fiberoptic bronchoscopy with BAL was performed via the oropharyngeal route in accordance with American Thoracic Society guidelines (20). Aliquots of saline, to a total volume of 240 ml, were separately instilled in to a segment of the right middle lobe. The bronchoscopies were all performed according to a standard operating procedure. Post collection, an aliquot of unfiltered and unprocessed BAL was immediately placed on ice then frozen at −80°C by the bronchoscopist. Negative control samples were collected by aspirating buffered saline through the bronchoscope suction channel before bronchoscopy. BAL fluid was examined for the presence of macrophages to confirm access of the alveolar compartment, and the absence of ciliated epithelium was used to exclude large airway contamination.

DNA Extraction
Two 2-ml aliquots of BAL were centrifuged at 20,000 × g for 15 minutes, to pellet cell debris and bacteria, and genomic DNA was then isolated using the MP Bio FastDNA SPIN Kit for Soil (http://www.mpbio.com) as previously described (21).

Genotyping
Genotypes of the MUC5B SNP rs35705950 were determined using TaqMan assays (Life Technologies, Carlsbad, CA). Reactions were performed in 384-well plates, and fluorescence was read using an Applied Biosystems Viia7 Sequence Detection System.

16S rRNA Gene Quantitative Polymerase Chain Reaction and Pyrosequencing
The V3–V5 region of the bacterial 16S rRNA gene was amplified using the 357F forward primer and the 926R reverse primer for both 16S quantitative polymerase chain reaction (qPCR) and pyrosequencing as previously described (21). The barcoded pyrosequence reads were processed using QIME 1.7.0 (22). Initial denoising was performed to remove sequencing errors (23), and PCR-generated artifacts were removed using ChimeraSlayer (24). Sequences were clustered into operational taxonomic units (OTUs) at 97% identity (25), aligned to full-length 16S rRNA sequences (26), and assigned a taxonomic identity with the Ribosomal Database Project classifier using the SILVA reference database.

Statistical Analysis
Continuous variables are presented as means (±SD) and categorical variables as proportions. Metastats was used to perform nonparametric t test comparisons.
of microbial communities between groups (27), with P values corrected for multiple hypotheses testing using the FDR approach of Benjamini and Hochberg. We restricted testing to OTUs that had a differing mean abundance between cases and control subjects of more than 1% of the total. Shannon’s entropy (28) (α diversity index) and weighted and unweighted UniFrac distances (29) (β diversity) were calculated in QIIME. The time-to-event curves were calculated using the Kaplan–Meier method and compared with the use of the log-rank test. Differences between subject groups were evaluated with the use of the Mann–Whitney test for continuous variables and Fisher exact test for categorical variables. Spearman rho was used to calculate correlations between continuous variables. The statistical significance of association of variables with a diagnosis of IPF was assessed using a stepwise backward elimination logistic regression process to select among potential covariates for inclusion in the final model. The statistical significance of association of variables with genotypes of the rs35705950 (coded 0, 1, and 2) were modeled using multiple permutations of the data (bootstrapping). All analyses were performed with the use of SPSS (version 21) and R (http://cran.r-project.org/). A two-sided P value of less than 0.05 was considered to indicate statistical significance.

Results

Subjects, Sampling, and Sequencing
Seventy-five patients with suspected IPF and 44 control subjects (27 healthy control subjects and 17 subjects with moderate COPD) (Table 1) were enrolled in the study and underwent bronchoscopy. During the study period, nine further patients with suspected IPF were excluded from enrollment because they reported symptoms of a lower respiratory tract infection or antibiotic usage in the preceding 3 months. A final diagnosis of IPF was made after multidisciplinary team discussion; 10 of the 75 recruited patients did not fulfill the American Thoracic Society diagnostic criteria for IPF and were subsequently excluded from the study (19).

The remaining 65 subjects with IPF were predominantly men (77%), with a mean age of 68 years, and had moderately severe disease at enrollment (carbon monoxide diffusion capacity (DLCO), 44.7% predicted; FVC, 76.5% predicted). Nine of the 65 subjects with IPF had coexisting emphysema based on high-resolution computed tomography scan, and in all cases the extent of emphysema was less than the extent of fibrosis on computed tomography. The 44 control subjects were matched for smoking history and sex but were younger than the IPF cohort, with a mean age of 58.2 years (Table 1). None of the control group and only four of the subjects with IPF were using inhaled corticosteroids at the time of their assessment.

Bacterial Burden
First, we explored the differences in BAL bacterial load (burden) between cases versus control subjects. We found that on average, subjects with IPF had $1.9 \times 10^6$ copies of bacterial load.
the 16S rRNA gene/ml of BAL, which was more than twofold higher than the copy number in control subjects ($P < 0.0001$). Within the control subjects, there was no significant difference in bacterial load between subjects with COPD and healthy control subjects. The patients with IPF had a significantly higher bacterial burden than both control subgroups ($P = 0.006$ and $P = 0.0007$, respectively) (Figure 1). Sputum samples from 20 control subjects taken the day before bronchoscopy demonstrated a mean bacterial load of $3.1 \times 10^8$ copies of the 16S rRNA gene/ml, 32 times higher than the subsequent lavage ($P < 0.0001$). There was no correlation between the bacterial load and either the total or differential BAL cell counts. The negative control samples (sterile saline aspirated through the suction channel of the bronchoscope) yielded a bacterial burden close to or below the lower limit of qPCR quantification (1,000 copies/ml). Although BAL return differed between IPF cases and control subjects (124.4 ml ± 31 vs. 96 ml ± 41; $P < 0.001$), there was no relationship between bacterial burden and BAL yield (Spearman $\rho = -0.027$, $P = 0.78$). Importantly, no ciliated epithelial cells (which may have been indicative of large airway contamination) were seen in the BAL returns.

Within the subjects with IPF we found bacterial burden to be associated independently with the rs35705950 polymorphism in the promoter of the mucin gene MUC5B genotype ($P = 0.01$), with patients possessing a minor allele having a lower bacterial burden (see Figure E1 in the online supplement). There was no correlation between bacterial burden and baseline disease severity, measured by either FVC (Spearman $\rho = -0.11$, $P = 0.49$), total lung capacity (Spearman $\rho = -0.12$, $P = 0.49$), $DLCO$ (Spearman $\rho = -0.20$, $P = 0.21$), or the composite physiologic index (Spearman $\rho = 0.06$, $P = 0.70$).

To test for a relationship between bacterial load and disease progression, we partitioned the patients with IPF into groups with progressive or stable disease. As a 6-month decline of greater than 10% in FVC is associated with an increased risk of mortality in IPF (30), we defined disease progression as either a relative decline in FVC of greater than 10% or death. Individuals with IPF whose disease had progressed at 6 months ($n = 22$) demonstrated a significantly higher BAL bacterial burden when compared with subjects with stable disease ($2.35 \times 10^8 \pm 1.68 \times 10^8$ compared with $1.41 \times 10^8 \pm 1.40 \times 10^8$ copies of the 16S rRNA gene/ml of BAL; $P = 0.02$). To investigate this relationship further, we separated the subjects with IPF into tertiles based on the 16S rRNA gene copy number/ml of BAL. Individuals in the top tertile with the highest bacterial burden were at a substantially increased risk of mortality compared with subjects in the bottom tertile (i.e., those with the lowest bacterial burden) (hazard ratio, 4.59; 95% confidence interval, 1.05–20) (Figure 2). There were no significant differences in sex, age, smoking status, or disease severity between the patients with IPF within these tertiles of bacterial burden.

**Microbial Communities**

We then compared the baseline microbiota of the subjects with IPF and the control subjects (Figure 3). Sequencing revealed *Streptococcus*, representing 30% of total reads, was the most common genus in subjects with IPF, followed by *Prevotella* (10.9%) and *Veillonella* (10.6%). In the combined control subjects, *Streptococcus* also formed the most common genus (27.1% of total reads) followed by *Prevotella* (11.6%) and *Veillonella* (7.1%). We found no significant differences in the BAL microbiota between the healthy control subjects and subjects with COPD (Figure 3). The microbial communities of subjects with IPF were by contrast less diverse (Shannon diversity index, 3.81 ± 0.08 vs. 4.11 ± 0.10; $P = 0.005$) and contained fewer numbers of OTUs (44.89 ± 1.50 vs. 54.33 ± 1.86 OTUs; $P < 0.0001$) than the control subjects. We discovered in patients compared with control subjects a 3.4-fold increase in sequence reads of a potentially pathogenic *Haemophilus* sp. (OTU 739) (36.0 ± 7.5 vs. 10.5 sequences ± 1.9 sequences; $P < 0.001$); a 2.1-fold increase in a *Neisseria* sp. (OTU 594) (57.9 ± 9.4 vs. 27.5 ± 5.5 sequences; $P < 0.01$); a 1.4-fold increase in a *Streptococcus* sp. (OTU 881) (113.6 ± 11.4 vs. 82.2 ± 8.9 sequences; $P < 0.05$); and a 1.5-fold increase in a *Veillonella* sp. (OTU 271) (84.8 ± 5.7 vs. 56.6 ± 4.5 sequences; $P < 0.001$) (Figure 4).

There were no demonstrable changes in community structure or composition between patients with IPF with progressive or stable disease, despite the differences in overall bacterial burden. There were no significant differences in the BAL microbiota between the healthy control subjects and subjects with COPD, and exclusion of subjects with COPD from the control panel did not change the differences detected. When our data were analyzed with patients on corticosteroids excluded from the IPF cohort, the observed differences in the microbiota remained unchanged.

Standard microbial culture of BAL was positive in five cases of IPF (7.6%) but none of the control subjects. In each instance, the cultured bacterial species were also identified by the pyrosequencing data, although they were not always the most abundant species within the microbiome. We then examined, using a stepwise logistic regression, whether bacterial burden...
and the relative abundance of these OTUs were associated independently with a diagnosis of IPF. Age and smoking status were included in the model to control for possible confounding. Total bacterial burden and the numbers of three specific OTUs (Veillonella sp. [OTU 271], Neisseria sp. [OTU 594], and Streptococcus sp. [OTU 881]) all remained significantly associated with a diagnosis of IPF ($P = 0.001, 0.007,$ and 0.01, respectively). The overall $R^2$ was 0.66 with age in the model and was 0.51 without age. The abundance of Haemophilus OTU 739 was strongly correlated with Neisseria OTU 594 (Spearman $\rho = 0.42$). Removal of Neisseria from the model made Haemophilus OTU 739 significant ($P = 0.02; \beta = 0.038 \pm 0.016$) with minimal change in the overall multivariate $R^2$ (0.657).

**Discussion**

In this study we demonstrated that when compared with control subjects, patients with IPF have a higher bacterial load in bronchoalveolar lavage fluid and significant differences in the composition and diversity of their microbiota. We have also shown that an increased bacterial load at the time of diagnosis identified patients with more rapidly progressive IPF and a higher risk of mortality.

The baseline bacterial communities we observed in patients with IPF and control subjects contained organisms such as Streptococcus, Prevotella, Fusobacterium, and Haemophilus, which are commonly found in the airways of healthy subjects, subjects with asthma, and patients with COPD (31–34). We discovered differences in specific OTUs between cases and control subjects, notably the presence of more abundant Streptococcus, Haemophilus, Neisseria, and Veillonella spp. in the patients with IPF. This indicated that total bacterial load and the abundances of these specific OTUs provided independent predictors of IPF case status and suggested that potentially pathogenic OTUs may be acting synergistically within a context of an increased bacterial load.

The rs35705950 polymorphism in the promoter of the mucin gene MUC5B confers an increased risk of developing IPF (4) but paradoxically confers a survival benefit among patients with IPF (7). We found bacterial burden to be associated independently with rs35705950 genotype ($P = 0.01$), with patients possessing a minor allele having a lower bacterial burden, suggesting a direct relationship between host immunity and bacterial load.
more densely populated and less diverse bacterial communities of the lower airways in IPF may provide persistent stimuli for repetitive alveolar injury. The temporal and spatial heterogeneity observed in usual interstitial pneumonia (the histological lesion of IPF) speaks to the likely importance of repetitive injury as a major factor in the pathogenesis of the disease (35). The bacterial communities of the lower airways are a plausible candidate for this trigger, and regional differences in the bacterial microbiome (36) may help explain the distribution of the fibrotic lesions in IPF.

We do not find any association between specific microbes and disease progression. Han and colleagues have recently presented data for the microbiota of individuals with IPF in a retrospective study (37). They demonstrate the most commonly identified bacteria in the lungs of individuals with IPF were Prevotella sp., Veillonella sp., and Escherichia sp., the presence of which we confirmed in our cohort. The authors found that the presence of a specific Streptococcus sp. or Staphylococcus sp. above a statistically modeled threshold was associated with a composite endpoint for disease progression. However, less than half of their patients had either of these bacteria in levels above the threshold, suggesting neither can fully explain disease pathogenesis or progression (38).

We are able to confirm the presence of Streptococcus sp. or Staphylococcus sp. in the IPF microbiome, and a longer follow-up period has allowed us to investigate for associations between the microbiome and survival. We demonstrate that although there are clear differences in the IPF microbiome compared with healthy individuals, it is the bacterial load that effects survival.

We cannot not draw any conclusions regarding the causal nature or not of this altered respiratory microbiome in IPF. Trials of antibiotic therapy may help elucidate this. There has recently been a large placebo-controlled multicenter study evaluating the use of Septrin in the broader category of fibrotic idiopathic interstitial pneumonias. This treatment was not well tolerated, and there was ultimately no detectable difference in the primary endpoint. Despite this, in subjects who tolerated Septrin therapy there was a reduction in infections and subsequent mortality. The authors hypothesize that this observation may be a result of Septrin’s antimicrobial effects, but it is difficult to draw conclusions because of its concurrent antiinflammatory actions. However, combined with our data here and the high mortality associated with bacterial respiratory tract infections in IPF, this observation does suggest a more targeted and well-tolerated antibiotic should be trialed in an IPF cohort.

There is a wide variation in the reported qPCR values for bacterial load in BAL, and the absolute quantification is dependent on multiple factors from sampling to processing, hindering comparisons between studies with differing primers and qPCR conditions. The bacterial loads reported here are comparable with other studies (36, 39), and our conclusions are based on relative values between cases and control subjects and those with stable or progressive disease. The inclusion of negative sampling control subjects (sterile saline aspirated through the suction channel of the bronchoscope), which yielded bacterial burden close to or below the lower limit of qPCR quantification, also demonstrates that the qPCR results are not simply artifact from a sampling or processing error.

Our study has a number of limitations. IPF is a disease of the lung parenchyma, whereas we used BAL fluid to sample the distal airways. Direct sequencing of lung
tissue may provide further information on the pulmonary microbiome (40), but lung biopsies are undertaken infrequently in IPF, and gathering truly healthy control samples would be extremely difficult. We used the oropharyngeal route to pass a bronchoscope into the lungs, and some secretions from the upper airways will have been carried on the tip of the bronchoscope. Any carryover, however, will have been heavily diluted by the 240 ml of BAL fluid. The high percentage of BAL return and an absence of ciliated epithelial cells in BAL fluid (indicative of large airway contamination) provided confidence that our BAL return was primarily derived from the distal airspace. We note that strong similarities have been seen between microbiota identified from BAL fluid and from surgically removed lung tissue in which upper airway contamination was not an issue (36).

Most importantly, we used the same protocols and procedures for cases and control subjects, so any contamination will not have systematically influenced our results. Although there is currently no universally adopted approach to minimize potential upper airways carryover, the use of protected BAL fluid, pro-BAL catheters, or protected catheter brushes could help avoid contamination.

Although the combination of 16S rRNA PCR with next-generation sequencing allows the parallel sequencing of large number of samples at relatively low cost and without the need for culture, there are limitations to the technique. The most significant are the biases introduced by primer design, which may select for, or against, particular bacteria (41). A further limitation to our study is that we did not assess for the presence or absence of other potential nonbacterial lung pathogens. Although a substantial study found no evidence of detectable viruses by PCR of BAL fluid from 40 stable patients with IPF (13), it is conceivable that some differences in progression could be due to the presence of respiratory viruses in our sample.

IPF is a fatal progressive disease with a median survival of 2 to 3 years. There are no effective treatments for the illness. We have demonstrated that increased bacterial load at the time of diagnosis identifies patients with more rapidly progressive IPF and a higher risk of mortality. It is not possible to conclude from our results whether the presence of an altered microbiome is the cause or a result of the destruction of the normal lung architecture. Although requiring further validation in larger prospective cohorts, our findings nevertheless provide a strong rationale for trials of antimicrobial therapy in IPF. Such trials will help determine the etiological role of bacteria in the progression of IPF while at the same time investigating a novel potential treatment for a currently intractable and fatal disease. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

Acknowledgment: The authors thank the late Dr Joseph Footitt, who was instrumental in the study of control subjects.

References


