MICROFLORA IDENTIFICATION IN CHRONIC OBSTRUCTIVE PULMONARY DISEASE EXACERBATED SPUTUM THROUGH CULTURE AND METAGENOMIC METHODS

DANIEL MINA
Spring, 2011

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Biochemistry and Molecular Biology
with honors in Biochemistry and Molecular Biology

Reviewed and approved* by the following:

Eric Harvill
Associate Professor of Microbiology and Infectious Disease
Thesis Supervisor

Ming Tien
Professor of Biochemistry
Honors Advisor

Wendy Hanna-Rose
Associate Department Head for Undergraduate Studies
Honors Advisor

* Signatures are on file in the Schreyer Honors College.
ABSTRACT

Microbial colonization by *Haemophilus influenzae, Moraxella catarrhalis, Streptococcus pneumoniae, Enterobacteriaceae, and Pseudomonas* species is speculated to be a factor in the progression of Chronic Obstructive Pulmonary Disease (COPD); however, the role of bacterial infection in COPD has not been completely understood (Brusse-Keizer, 2008). Researchers studying the bacteriology of COPD patients have varying conclusions on the most prevalent pathogens present in the sputum. Only a very small percentage of all bacteria can be cultured, so advanced molecular methods are needed for the diagnosis of infections (Staley, 1985). In this study, we used metagenomics based on 16S rRNA sequences to study the sputum microflora of 19 COPD patients admitted to The Penn State Hershey Medical Center. In comparison with the 454 metagenomics sequencing, culture methods were also used, further identifying cultured organisms by individual 16S rRNA sequencing. While we identified many of the pathogens known to contribute to COPD, we also identified *Actinomyces spp., Tannerella forsythia, Gemella morbillorum, Eikenella corrodens*, and other opportunistic pathogens in the respiratory tracts of COPD patients. As discovered for known COPD pathogens, these species may also be contributing to disease symptoms. Metagenomic approaches have a tremendous capability to generate compositional data, and it has provided us with novel insight into the bacterial colonization of the lower airways of COPD patients.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>7</td>
</tr>
<tr>
<td>Obtainment of COPD Lung Sputum Samples</td>
<td>7</td>
</tr>
<tr>
<td>Bacterial Isolation and Growth</td>
<td>7</td>
</tr>
<tr>
<td>16S rRNA Sequencing</td>
<td>8</td>
</tr>
<tr>
<td>Total Genomic DNA Isolation and 454 Sequencing</td>
<td>9</td>
</tr>
<tr>
<td>Results</td>
<td>11</td>
</tr>
<tr>
<td>Discussion</td>
<td>27</td>
</tr>
<tr>
<td>References</td>
<td>32</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 1: Media Type CFU for Eleven Patients .............................................................. 13
Figure 2: 16S rRNA Identifications ............................................................................. 16
Figure 3: Most Prominent Culturable Species .............................................................. 16
Figure 4: Taxonomic Tree from Metagenomic Sequencing .......................................... 21
Figure 5: 16S rRNA Gene Reads ............................................................................... 23
Figure 6: Heat Maps of 16S rRNA Gene Reads ............................................................ 25
LIST OF TABLES

Table 1: List of Patient Sputum Samples................................................................. 11
Table 2: List of Identified Culturable Bacteria. ...................................................... 18
ACKNOWLEDGEMENTS

I would like to give a very special thanks to Dr. Harvill for giving me the opportunity to conduct this research. I would also like to thank all the members of the awesome Harvill Lab. Thank you, Alexia Karanikas, for bringing me into the lab. Your initial advice as to how I should go about learning in the lab was essential for my growth. Thank you, Sara Hester, for giving me the chance to scare you several times in the middle of the night when you thought that no one else was in the lab. But seriously, thank you for being the one who pulled me out of the novice stage by providing me with my first project. I know it failed, but it was essential for me to learn firsthand that not everything works the way you want. Thank you, Olivier Rolin, for entertaining me with your arsenal of hilarious jokes while you were dissecting mice ten times faster than I was. Thank you, Jihye Park, for filling the Harvill lab with laughter. Sometimes I sat in front of a computer in the lab just so you could make me laugh. You will always be my hero. Thank you, Liron Bendor, for being a fellow Justin Bieber fan, and for being on my side when we were arguing with Jihye about him. Thank you, Sarah Muse, for having a very high IQ, and for being a highly resourceful guide to me while I was working on this thesis. Thank you, David Place, for being one of the only sane graduate students I’ve met. I seriously think you were the only normal guy in the lab during my stay. Sorry O-man. Thank you, Laura Goodfield, for providing the lab with that dream explanation book. I was having some seriously weird dreams during the period I was writing this thesis. I just couldn’t find an explanation for being attacked by elephant sized COPD lung sputum bacteria. Finally, thank you, Laura Weyrich, for being a gold mine of project guidance. Thank you for trusting me with this project, for teaching me the majority of the techniques I learned in the lab, for teaching me how to work on my own, for answering my never ending questions, for spending the time editing every aspect of this thesis, and for being the beastly graduate student machine that you are. I truly admire and respect all of you.
Introduction

Chronic Obstructive Pulmonary Disease (COPD) affects millions of American adults and is one of the fastest growing causes of mortality in the United States (Sullivan, 2000). It is estimated that by 2020, COPD will rank third as cause of mortality and fifth as cause of disability worldwide (Murray, 1997). While active or passive smoking has been identified as the major cause of this disorder, small irritants, pollutants, and genetic disorders can also affect the progress of this disease (Burge, 2003). Smoking, irritants, and infection lead to inflammation of the bronchial tubes, the alveoli, and most of the cells lining the inside of the lungs, causing extreme breathing discomfort for the victims.

The syndrome of COPD consists of chronic bronchitis, bronchiectasis, and emphysema that combine in different ways for an individual patient. Bronchitis is defined as inflammation of the mucous membranes, bronchiectasis as irreversible dilation of part of the bronchial tree caused by muscle tissue destruction, and emphysema as destruction of the lung tissue. Because of these adverse effects, medical treatment is essential for improving the quality of life for these patients. The treatments that have been shown to improve survival in COPD patients are smoking cessation (Scannion, 2000) and oxygen therapy in subjects with respiratory failure (Nocturnal Oxygen Therapy Trial Group, 1980). The prognosis for COPD in the early stages is usually good if smoking is ceased and proper treatments are implemented by a clinician, but if the disease is permitted to progress to exacerbations through continued smoking, the prognosis becomes severely poor (Godtfredsen, 2002).
Patient exacerbations of COPD are frequently characterized by increased coughing, dyspnea, and sputum production with increased purulence in sputum. These exacerbations have been previously shown to be triggered by viruses, bacteria, and pollutants (Papi, 2006). Papi et. al. also found that during exacerbations, the inflammatory cellular pattern changes, with a further increase of eosinophils and/or neutrophils and various inflammatory mediators, for example: cytokines, chemokine receptors, adhesion molecules, and markers of oxidative stress. Worsening of inflammation is considered responsible for the deterioration of lung function and clinical status during exacerbations. Additionally, goblet cell hyperplasia and mucus hypersecretion are present in the airways of most COPD patients, especially in those with clinical features of bronchitis (Barnes, 2000). Furthermore, with inflammation leading to the activation of innate immune cells, infiltrating neutrophils and lymphocytes cause airway thickening and remodeling, resulting in reduced airflow in the lungs (Barnes, 2004). Treatments capable of reducing the frequency of acute exacerbations of COPD could potentially reduce mortality inherent to exacerbations.

Antibiotics may have beneficial effects in treating or preventing some episodes of exacerbations (Saint, 1995). In the United States, COPD treatment guidelines recommend antibiotic treatment for patients with purulent sputum and either an increase in sputum production or an increase in dyspnea, but the evidence is based on small, heterogeneous trials, few of which include hospitalized patients (Snow, 2001). However, a recent study of 84,621 hospitalized patients showed improved results for COPD patients with early antibiotic treatment regardless of disease severity (Rothberg, 2010). Antibiotics used included first-, second-, and third-generation cephalosporins,
quinolones, macrolides, tetracyclines, trimethoprim-sulfamethoxazole, and amoxicillin with or without clavulanic acid. Similarly, Murphy and Parameswaran recommend second/third generation cephalosporins, amoxicillin/clavulanic acid, azithromycin and respiratory fluoroquinolones as initial choices (Murphy, 2009). Even though these antibiotics are recommended, it is generally only recommended that they be administered to severely ill patients and those that are at high risk for respiratory failure and death (Murphy, 2009). Researchers in Denmark and Italy also recommend similar antibiotics (Larsen, 2009 and Luppi, 2010). Although these recommendations exist for severe cases of COPD, the role of pathogenic bacteria and the use of prophylactic antibiotics in progression of stable COPD are still understudied and inconclusive.

Despite evidence implicating bacterial colonization in the pathogenesis of airway diseases, there are very few systematic studies of bacteria and other organisms residing in the airways (Sethi, 2001). Although not to the same extent of cystic fibrosis, research suggests that the mucociliary clearance of bacterial pathogens may be impaired by the fact that respiratory epithelial cells lose cilia and undergo squamous metaplasia, which is a change in the shape of the cells (Hogg, 2009). Known bacterial pathogens have been found in around 50% of acute exacerbations of COPD, including *Haemophilus influenzae*, *Moraxella catarrhalis*, *Streptococcus pneumoniae*, *Enterobacteriaceae*, and *Pseudomonas* species (Murphy, 2000 and Murphy, 2009). However, high amounts of variation in the primary bacterial species plague researchers that study COPD sputum bacteriology. A study done in Hong Kong shows *S. pneumoniae* to be the most prevalent (Ko, 2007), Saudi Arabia- *M. catarrhalis* (Almoudi, 2007), and Germany - *H. influenzae* (Lode, 2007).
Out of all these bacterial species, the role of *P. aeruginosa* in COPD has been studied the most, perhaps because of the link between *Pseudomonas* infection and cystic fibrosis patients. Previous studies have shown that *P. aeruginosa* was isolated from sputum samples from between 4% and 15% of sputum samples of adults with COPD, which suggests that this bacterium contributes to infection (Groenewegen, 2003). The role of *P. aeruginosa* in COPD is not well understood, and it is not known if antibiotic therapy directed at this species is of any benefit (Murphy, 2008). Perhaps the most important step in treating patients with *Pseudomonas* colonization is establishing its role in exacerbations of COPD, a task which was undertaken by Martinez et al who found that patients with COPD were usually infected with one persistent *P. aeruginosa* clone without evidence of interpatient transmission. However, during chronic infection, they found that the clone was diversified, which led to the coexistence of isolates with different morphotypes and antibiotic susceptibility. *P. aeruginosa* was evolved with an increased mutation rate, increased antibiotic resistance, and reduced production of proteases in these patients. These results pioneered the hypothesis that *P. aeruginosa* contributes to chronic infections in COPD, but given that it only appears in at most 15% of patients with exacerbations, further research investigating other bacterial species is warranted.

Only small percentages of known bacteria can be cultured, so advanced methods are needed for the diagnosis of infections. Indeed, it has already been shown through advanced molecular methods that the microflora of humans is severely more complex than previously thought (Turnbaugh, 2008). Furthermore, humans have evolved relationships with their symbiotic bacteria that are essential for health, and ecological
changes altering this symbiosis can result in disease (Dethlefsen, 2007). Hilty et. al. recently performed a metagenomics study comparing the airway microflora in adult patients with asthma, COPD, and controls. The results of their 16S rRNA bacterial sequence analysis showed that the bronchial tree contained a characteristic microflora, and suggested that this microflora is disturbed in asthmatic airways (Hilty, 2010). When they made microflora comparisons between these three groups, they found that Bacteriodetes (particularly Prevotella spp.) were more common in control subjects than in five COPD patients, and that Neisseria, Haemophilus, Moraxella, and Staphylococcus spp. were present in excess in the airways of children with difficult asthma. Their results also suggested that Prevotella and Veillonella spp. are a distinctive component of the normal flora of the lung, and that they were not previously identified because of their requirement for anaerobic culture.

In this study, we have both cultured and used 16S rRNA metagenomics to study the sputum microflora of 19 COPD patients admitted to The Penn State Hershey Medical Center. We identified the six most culturable species in eight patients that can be isolated on Blood Agar (BA), Bordet Gengou agar (BG), Tryptic Soy Agar (TSA), nutrient agar (NA), Luria Burtani agar (LB), and MacConkey Agar (Mc). Six patients were additionally chosen for metagenomics analysis by 16S rRNA sequences via 454 sequencing technology. Here, we show that additional pathogens, such as Actinomyces spp., Tannerella forsythia, Gemella morbillorum, Eikenella corrodens, and other opportunistic pathogens are in the respiratory tracts of COPD patients and, therefore, may be contributing to their disease symptoms. Although the method is costly, metagenomic analysis of sputum samples reveals numerous complexities to the bacterial communities
in patients with COPD disease that cannot be shown with only the culture approach. This project shows that our previous understanding of the bacteriology of COPD lungs has been severely limited. Our results suggest that more advanced molecular research of this type is required for more conclusive analysis regarding the role of bacteria in exacerbations of COPD.
Materials and Methods

Obtainment of COPD Lung Sputum Samples

Nineteen COPD patient sputum samples analyzed for this study came from The Penn State Milton S. Hershey Medical Center, Hershey, Pennsylvania, USA. Nebulisation with 3% saline was performed using the Pari LC®D Nebulizer. After seven minutes of nebulisation, measurement of oxygen saturation and spirometric tests were performed and nebulisation continued if the forced expiratory volume (FEV$_1$) had not fallen by more than 20%. Sputum samples were collected during seven minutes and were placed in separate sputum pots. Samples were transported to The Pennsylvania State University, University Park, Pennsylvania, USA on ice, and later stored in -80 °C in the presence of an equal volumetric ratio of sample to RNA later after processing.

Bacterial Isolation and Growth

Sputum samples were taken and diluted to $10^{-1}$ and $10^{-4}$ concentrations using sterile phosphate buffered saline (PBS). Previously, it has been shown that COPD lung sputum bacteria grow on Blood Agar (BA) and MacConkey Agar (Mc) (Shahnawaz, 2003). In addition to BA and Mc, other types of media were used, including Bordet-Gengou Agar (BG), Nutrient Agar (NA), Tryptic Soy Agar (TSA), and Luria Burtani Agar (LB). Volumes of 100 ul of the diluted samples were plated on each media type bringing the sputum samples to a final concentration of $10^{-2}$ and $10^{-5}$. The plates were
kept at 37°C for one day. The experiment was duplicated for the 10⁻⁵ concentration, and these samples were incubated for one week at 37°C to allow for further growth. Freezer stock solutions of isolated bacterial colonies were kept at -80 °C in one ml of 20% glycerol in PBS. Colony forming units (CFU) were obtained for all plates, and graphs showing bacterial quantities were created on Microsoft Excel using a logarithmic scale. Plates that included a lawn of bacteria at the highest dilution were estimated at 1,000 CFU on the plate.

**16S rRNA Sequencing**

In order to determine the variation and identify culturable strains more specifically, 16S rRNA sequencing was completed on each individual isolate. Out of the 19 sputum samples, complete sequencing results for six isolated colonies were obtained for eight patients. Genomic DNA from isolated bacteria was added to PCR reactions containing rRNA gene primers 16SF AGAGTTTGATCATGCTCAG and 16SR AAGGAGGTGATCCAACC. The following conditions used for the primer set: 94°C for 5 min; 95°C for 30 s, 56°C for 1 min, 72°C for 1.5 min, 30 times; and 72°C for 8 min. These primers attached to the V2 variable region of the 16S rRNA encoding gene, and allowed for amplification of an approximately 1,500 bp product. Following amplification, the PCR products were separated on a 1% agarose gel, and bands corresponding to a target product length of 1,500 bp were excised under a UV machine (Foto Prep). DNA was extracted from bands using the published protocol from Qiagen MinElute Gel Extraction Kit (Valencia, CA). Samples were sent to the Penn State Genomics Core
Sequencing Facility for DNA sequencing. Forward and reverse sequences were concatenated into a MEGA file in order to identify consensus sequences. NCBI BLAST analysis was performed using the consensus nucleotide sequences (www.ncbi.nlm.nih.gov/blast/). BLAST results with greater than or equal to 90% sequence similarity were selected as the most probable species to which the sequence corresponded. If no significant BLAST hits were obtained, the bacteria were labeled as unknown. If BLAST hits with known species did not fit the criteria for use, and uncultured species were provided as a result, the bacteria were labeled as uncultured. Graphs were created on Microsoft Excel, which showed the species identification for all eight patients.

Total Genomic DNA Isolation and 454 Sequencing

Six samples were chosen for 454 metagenomics sequencing in order to identify a higher diversity of bacteria. These samples included patients 110, 115, 119, 121, 122, and 123. Isolation of total DNA was done using a Mo Bio Laboratories Kit (Carlsbad, CA) entitled Biostatic FFPE Tissue DNA Isolation Kit. Step 3 of the experienced user protocol was modified, in which the samples were left at 55°C overnight, instead of one hour. Library preparations for 454 sequencing were done by Gregory Grove at the Penn State Genomic facility. Samples were bar-coded with known DNA sequences to distinguish each sample, and sequenced on one fourth of a plate at the Penn State Genomics Core facility. Sequencing results (between ~22,000 and ~32,000 sequence reads for each sample) containing nucleotides ranged from 30 to 520 bp, and they were
put through Mothur metagenomic software in order to process the sequences based on length (longer than 100 bp) and quality. Sequences with good quality were blast run against the greengenes database (http://greengenes.lbl.gov), and then they were plugged into MEta Genome ANalyzer (MEGAN) which classified each read onto NCBI taxonomic trees.
Results

Lung sputum was provided by 19 patients from The Penn State Milton S. Hershey Medical Center located in Hershey, Pennsylvania. The study was conducted through the span of ten months beginning in April of 2010 (Table 1).

<table>
<thead>
<tr>
<th>Sample #</th>
<th>Patient #</th>
<th>Date</th>
<th>Media Analysis</th>
<th>Media Sequencing</th>
<th>Metagenomics Sequencing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>101</td>
<td>4/1/2010</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>102</td>
<td>4/1/2010</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>103</td>
<td>4/1/2010</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>104</td>
<td>5/3/2010</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>109</td>
<td>5/19/2010</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>110</td>
<td>5/26/2010</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>7</td>
<td>112</td>
<td>6/2/2010</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>113</td>
<td>6/7/2010</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>114</td>
<td>6/9/2010</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>115</td>
<td>6/10/2010</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>11</td>
<td>117</td>
<td>7/6/2010</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>118</td>
<td>7/26/2010</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>119</td>
<td>7/26/2010</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>14</td>
<td>121</td>
<td>8/16/2010</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>15</td>
<td>122</td>
<td>8/16/2010</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>16</td>
<td>123</td>
<td>8/17/2010</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>17</td>
<td>125</td>
<td>8/19/2010</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>126</td>
<td>11/4/2010</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>19</td>
<td>127</td>
<td>2/11/2011</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: List of all patient sputum samples received from Penn State Hershey Medical Center. All sputum was plated on bacterial media, as described above. Media sequencing was performed on eight samples that produced at least six different types of bacteria based on morphology observation. Metagenomics sequencing was performed on six samples of choice.
Patients who had six or more different culturable bacteria were characterized, and those that did not fit this criterion were not used for 16S rRNA sequencing analysis because they did not appear to have enough diversity in the cultured flora. Out of the 19 patients, eight were used for bacterial media sequencing, and six were used for 454 metagenomics sequencing. The eight used for bacterial media sequencing were patients 101, 102, 113, 115, 119, 121, 122, and 126. In order to be used for metagenomics analysis, more than 5 ng/ul of DNA was extracted from the remaining sample. The six samples used for 454 metagenomics sequencing were patients 110, 115, 119, 121, 122, and 123.

After plates were incubated for 24 hrs or one week, colony forming units (CFUs) were counted and recorded for each media type, and the results for eleven patients are shown (Figure 1a and b). Figure 1a shows log CFU for each media type after 24 hours of incubation, while Figure 1b shows log CFU for each media type after one week of incubation. This allowed for a comparison of each media type in regards to the efficiency of bacterial growth.
Figure 1: a) This graph shows the log 10 scale of CFU for BG, BA, NA, TSA, LB, and Mc after one week for patients 113, 115, 117, 118, 121, 122, 123, 125, 126, and 127. b) This graph shows the log 10 scale of CFU for BG, BA, NA, TSA, LB, and Mc after 24 hours for the eleven patients. The color bands represent a media type, while the size of a band corresponds to a log 10 CFU value. The limit of detection was set at 4.95, which corresponds to $9.99 \times 10^4$ CFU.
As inferred from Figure 1a, BA yielded bacterial growth for ten out of eleven patients (91%), BG- eleven out of eleven (100%), NA- six out of eleven (55%), TSA- five out of eleven (45%), LB- six out of eleven (55%), and Mc- zero out of eleven (0%). These results suggest that blood enhances bacterial isolation as more patient samples yielded colonies on BG and BA. Additionally, Mc was the least successful at isolating bacteria, as all eleven patients analyzed did not yield any detectable colonies for this media type.

Most colonies isolated from patients within one day of incubation were on BA (1.86 x 10^7 CFU), while no growth was observed on Mc. Similar numbers of bacteria were isolated on BG compared to BA with 1.66 x 10^7 CFU, while NA, TSA, and LB all yielded similar amounts of bacteria with 5.15 x 10^6 CFU, 7.8 x 10^6 CFU, and 4.3 x 10^6 CFU respectively.

After one week of incubation, BA yielded bacterial growth for eleven out of eleven patients (100%), BG- eleven out of eleven (100%), NA- ten out of eleven (91%), TSA- ten out of eleven (91%), LB- six out of eleven (55%), and Mc- four out of eleven (36%). As with the results after 24 hours of incubation, blood appeared to enhance colonization after one week, although the effect is weaker than that after 24 hrs of incubation.

Similar to the results of 24 hrs incubation, most colonies grew on BA (2.75 x 10^7 CFU), and the least amount of colonies were found on Mc (1.83 x 10^6 CFU) after one week of incubation, as shown in Figure 1b. Additionally, LB yielded the second highest amount of bacteria with 2.35x 10^7 CFU, while BG, NA, and TSA yielded 1.78 x 10^7 CFU, 1.69 x 10^6 CFU, and 1.08 x 10^7 CFU respectively.
In addition to counting the total CFU on each media type, six of the most prominent colony types were identified for eight patients (Figures 2 and 3). If the sequencing results provided the same bacterial species for more than one colony in a given patient, the data was combined.
Figure 2: These bar graphs show the six most prevalent culturable bacteria, as identified by 16S rRNA sequences from COPD patient sputum. The list of these species with their abbreviations is shown in Table 2.

Figure 3: This bar graph shows the most prominent culturable species found, and the number of patients in which they were found. The eight samples analyzed were patients 101, 102, 113, 115, 119, 121, 122, and 126.
The bacteria identification shown in Figure 2 was performed for patients 101, 102, 113, 115, 119, 121, 122, and 126. Several commensal respiratory bacterial species were identified by sequence homology in multiple patients, including *Streptococcus salivarius*, *Streptococcus parasanguinis*, *Rothia dentocariosa*, and *Staphylococcus epidermidis*. Interestingly, several pathogens known to cause respiratory disease, including *Neisseria meningitides*, *Neisseria flavescens*, and *Streptococcus pneumoniae*, were found in at least one patient, while known COPD patient pathogens, such as *Haemophilus influenzae*, *Moraxella catarralis*, and *Pseudomonas aeruginosa*, were not isolated from the patients analyzed. Out of the eight patients used for this experiment, four had at least one known respiratory pathogen present in their lung sputum. A total of 20 different bacteria were identified as a known species, and the list is shown in Table 2. Genera identified included *Arthrobacter*, *Bacillus*, *Corynebacterium*, *Lactobacillus*, *Micrococcus*, *Moraxella*, *Neisseria*, *Rothia*, *Staphylococcus*, and *Streptococcus*. As shown in Figure 3, *Streptococcus salivarius* was identified in all patients; *Neisseria flavescens* was identified in four patients; *Streptococcus parasanguinis*, *Rothia dentocariosa*, and *Rothia mucilaginosa* were identified in three patients; and the rest were identified in one patient. The following were the most prevalent species in each patient: *Rothia Dentocariosa* in patient 101, an unknown species in patient 102, *Staphylococcus epidermidis* in patient 113, *Streptococcus pneumoniae* in patient 115, *Neisseria flavescens* in patient 119, *Streptococcus parasanguinis* in patient 121 as well as in patient 122, and *Streptococcus mitis* in patient 126. Out of all the identified bacterial species, *Streptococcus pneumoniae*, *Neisseria meningitides*, and *Neisseria flavescens* are known to be involved in lung disease.
Table 2: This table provides a list of all identified bacteria from all patients. This list does not include unknown species.

In order to compare the metagenomic approach with this culture data, six patients were chosen for 454 metagenomics analysis. Sputum bacteria were identified using 16S rRNA database, and the data was incorporated into a taxonomic tree for each individual patient. The trees were then incorporated into a single taxonomic tree for patient comparison (Figure 4).
Figure 4: Bacterial species identified by 454 sequencing from patients suffering from COPD are compared on a taxonomic tree. Each color of bar represents an individual patient, and the height of the bar represents the number of reads found in each patient sample for a particular species.
Metagenomics data provided a much larger array of bacteria than culturable analysis as shown in Figure 4. All species identified through culture methods were also observed using 454 sequencing methods in at least the genus level. Additionally, we found several bacterial pathogens previously known to be present in COPD patient sputum in at least one of the six patients including *Neisseria, Haemophilus, Moraxella,* and *Streptococcus.* While these pathogens are known to have an effect on COPD exacerbation, several other species were identified in at least three of the six patients, including *Atopobium parvulum,* oral clones of *Actinomyces, Corynebacterium durum, Tannerella forsythia,* oral clones of *Prevotella,* *Bergeyella* clones, *Gemella morbillorum, Abiotrophia defectiva, Granulicatella para-adiacens,* oral clones of *Eubacterium,* oral clones of *Peptococcaceae,* oral clones of *Leptotrichia, Eikenella corrodens, Kingella denitrificans, Campylobacter curvus,* and *Campylobacter gracilis.*

Providing a more detailed figure of the amount of bacteria than shown in the taxonomic tree, a graph was created that showed the amounts of DNA hits for each class of bacteria (Figure 5). The number of DNA hits correlate to the amount of bacteria present in the sputum, and this analysis allows for the comparison of bacterial quantities present in the six patients.
Figure 5: This graph shows the number of 16sRNA gene reads for bacterial classes present in all six patients analyzed. Patients are represented by different shapes, while classes are assigned numbers and colors. While this does not provide the number of CFU present in the sputum, it can be assumed that the number of reads correlates to the number of CFU.

Figure 5 shows read counts based on classes, and we can actually compare number of bacteria present across patients. Four patients had over 10,000 Bacilli reads (contains Streptococcus), two patients had over 5,000 reads that were classified as Fusobacteria (contains species involved with periodontal disease), and two patients had over 5,000 Betaproteobacteria reads (contains Neisseria). It is also important to note that no patient had more than 5,000 Gammaproteobacteria reads, a class which contains *Moraxella, Pseudomonas*, and *Haemophilus* species. All other bacterial classes present in patients were under 5,000 reads.
In addition to this graph, heat maps were created allowing for the visualization of the most prevalent classes of bacteria in each of the six patients (Figure 6). These heat maps allow for a better comparison of bacterial quantities across patients, and they allow for the grouping of patients according to the bacteria present in their sputum.
Figure 6: 16s rRNA amplicon libraries from six patients (shown here as P110, P155, P119, P121, P122, and P123) were sequenced on a 454 sequencer. a) The pink heat map represents the percent of reads classified as a certain operational taxonomic unit (otu), represented here as bacterial orders. The darker the color, the higher the number of bacteria that were classified, although a white color does not necessarily indicate zero reads (shades of pink are chosen from a logarithmic scale). Here, the otus are clustered vertically or per patient, representing the most common species found in the patient as the darkest color. b) The green heat map represents bacterial reads compared among the patient sample. Darker shades represent a higher number of reads; however, white does not necessarily represent a value of zero. Question marks for both figures indicate that unknown bacteria were identified with the order, but could not be further specified, and may be novel species.

From the image in Figure 6a, the most commonly identified orders for each patient can be determined, as this heat map allows for comparisons across bacterial orders. When compared to the number of other bacteria that were present in their sputum, Patients 110, 115, 121, and 122 had higher proportions of Lactobacillales, patient 119...
had a higher proportion of Fusobacteriales, while patient 123 had higher proportions of Fusobacteriales and Mycoplasmatales. As indicated by the branches on the top of the image, patients 110 and 115 were the most closely related when comparing community similarity indexes, while patients 121 and 123 did not identify closely with the other patients, as indicated by the early separations in the tree, suggesting that patients may be grouped according to alterations in sputum microflora.

As shown in Figure 6b, the green heat map allows for a comparison across patients. Patient 122 had the most Lactobacillales (contains Streptococcus species), patient 115 had the most Neisseriales (contains Neisseria species), patients 115 and 122 had the most Pasteurellales (contains Haemophilus species), and patient 121 had the most Pseudomonadales (contains Moraxella and Pseudomonas species). Looking at the patients as a group, patients 119 and 122 had lower numbers of Lactobacillales, and patients 121 and 123 had lower numbers of Pasteurellales. As shown by the tree on the top of Figure 6b, patients 110 and 115 had similar numbers of all types of bacteria, while patients 123 and 119 had different numbers of bacteria when compared to the other four patients. It is important to understand that the heat maps provide different trees because the pink map finds relation based on similar ratios of bacteria, while the green heat map finds relation based on similar numbers of bacteria. Therefore, the green heat map is more applicable in grouping, as it allows for a comparison of bacterial quantities across the patients.
Discussion

Our results show that the bacteriology of COPD patient lungs is more diverse than previously thought. Culturable bacteria only accounted for a small number of the bacteria present in the sputum, while 454 metagenomics was able to identify a much greater diversity. As with the results of Hilty et. al., our results challenge the traditional medical teaching that the lower airways are sterile. Indeed, the techniques used in current literature on the microflora of COPD patients are based on culture techniques that are over 100 years old, which are now recognized to detect minimal numbers of bacteria in complex samples (Staley, 1985). Additionally, our results that identify respiratory pathogens in each of these patients challenge the recurring finding that bacteria only account for exacerbations in 50% of COPD patients, as this statement was based on culturable data alone. As shown in the taxonomic tree and the heat maps, we found that not only were there known COPD pathogens inhabiting the lungs of these patients, but there were also other opportunistic pathogens and bacteria, some of which are not culturable.

As shown in our media results, colonies grew best on BA after both 24 hours and one week. This is expected as BA contains a rich medium, which is often used to grow fastidious organisms, because it contains casein pancreatic digest, soy meal papaic digest, sodium chloride, and sheep’s blood (Atlas, 1997). Fastidious organisms with limited cultivation on BA were verified by our 16S rRNA sequencing of culturable bacteria, as much of the bacteria sequenced came from *Streptococci*, a genus known to grow well on
BA (Brown, 1919). Compared to BA, BG also allowed for the isolation of large bacterial numbers, as it produced bacteria from all patient samples after 24 hours and one week of incubation. BG is composed of glycerol, sodium chloride, casein pancreatic digest, animal tissue peptic digest, infusion potato solids, and sheep’s blood (Atlas, 1997). BG is usually used for the detection and isolation of *Bordetella* species; however, it has similar properties to BA, which explains the comparative numbers of isolation.

NA, LB, and TSA did not allow bacterial growth for as many patients as allowed by media containing sheep’s blood. NA contains Lab-Lemco beef extract, peptone, and sodium chloride (Atlas, 1997). NA is usually used for the isolation of heterotrophic bacteria, and it has the potential to allow growth for a wide range of bacteria. LB contains tryptone, sodium chloride, and yeast extract, and this media is usually used for the isolation and maintenance of bacteria for genetic and molecular studies (Atlas, 1997).

Lastly, we observed very little bacterial growth on Mc. Mc media is composed of peptone, lactose, bile salts, and sodium chloride. These components make a medium that is usually used for selective isolation and differentiation of coliforms and enteric pathogens based on the ability to ferment lactose (Atlas, 1997). Because Mc agar is meant to isolate enteric bacteria, it was expected that this media would not isolate bacterial organisms from the lungs; however, it was used as a negative control to investigate pathogens that may have found a niche in the lungs. TSA contains casein pancreatic digest, soybean meal papaic digest, and sodium chloride, which allows for the isolation of fastidious as well as nonfastidious microorganisms (Atlas, 1997). TSA has the same components as BA with the exception of sheep’s blood. Because the results after both 24 hours and one week showed BA to yield bacteria for more patient samples
than TSA, we can infer that the presence of blood allows for more growth of bacteria. Therefore, it is not surprising that media containing blood cultured greater amounts of bacteria, considering that the sputum samples came from the lungs.

Interestingly, several pathogens known to cause respiratory disease, including *Neisseria meningitides*, *Neisseria flavescens*, and *Streptococcus pneumoniae*, were found in at least one patient when using culture methods, while known COPD patient pathogens, such as *Haemophilus influenzae*, *Moraxella catarralis*, and *Pseudomonas aeruginosa*, were not isolated from the patients analyzed. Many species of *Streptococcus* were identified by 16S rRNA sequencing of cultured bacteria suggesting that this genus accounts for a significant portion of the lung microflora. *S. pneumoniae* was only present in one patient, *N. flavescens* was present in four patients and *N. meningitides* was present in one patient. A *Moraxella* species was also identified; however, this species was not *M. catarrhalis*, a species documented to have a role in the exacerbations of COPD patients.

While this data may not be taken as statistically relevant, we found that five out of the eleven patients had at least one COPD associated pathogen, a result which is close to the 50% of exacerbations stated to be caused by certain bacterial pathogens.

As shown by the array of the 454 16S rRNA sequencing data, metagenomics analysis provided a much larger diversity of bacteria than cultural analysis. The metagenomics analysis revealed the presence of the bacteria that were cultured, as well as unculturable species. While many of the bacteria present have not been identified as pathogens, several pathogenic microorganisms including *Neisseria*, *Haemophilus*, *Moraxella*, and *Streptococcus*, were identified for at least one of the patients, with each of the patients having at least one of the pathogens. However, sequencing matching to the
common COPD associated pathogen, *Pseudomonas aeruginosa*, was not identified in the taxonomic tree analysis, although other related Pseudonadales species were identified in two of the six patients. Our results introduced several more bacteria into the discussion of COPD exacerbation. These bacteria include *Actinomyces spp.*, *Tannerella forsythia*, *Gemella morbillorum*, *Eikenella corrodens*, and other opportunistic pathogens. As discovered for the known COPD pathogens, these species may also be contributing to disease symptoms.

Because this study did not include controls, we cannot make a comparison of our results with those of Hilty et. al. Their results suggested that *Prevotella* and *Veillonella* spp. are a distinctive component of the normal flora of the lung, and that they were not previously identified because of their requirement for anaerobic culture. In our data, we found that several *Prevotella* species were present in all six patients, while *Veillonella* was found in significant amounts in only one patient. Further research into this topic is required for conclusive evidence of an evolved microflora in COPD patients.

This data could be improved by incorporating more dilutions in the media analysis. Bringing the plates to concentrations lower than $10^{-5}$ may allow for more plates to yield bacterial colonies, and therefore, could result in an improved detection of species. Additionally, we plan to examine control sputum from healthy lungs, sputum from patients on antibiotics, and sputum from patients at different stages of COPD as classified by clinicians. Examining the microflora of healthy sputum would allow us to make an important comparison, which could help identify the differences that occur in the microflora of the lower airways in COPD patients. If patients are on antibiotics, we can examine the true effectiveness of the treatment on pathogens of the sputum, whether
certain antibiotics work better than others, whether certain dosages of antibiotics are more efficient or necessary, as well as the overall effects of antibiotics on the microflora of COPD patients at varying stages of the disease. These studies would allow us to further examine the relationship between microflora and COPD exacerbations which in turn would provide clinicians with the knowledge to improve the quality of life for these patients.
References


Barnes PJ. Alveolar macrophages as orchestrators of COPD. COPD 2004; 1:59-70.


Brown JH. "The use of blood agar for the study of streptococci." Monographs of the Rockefeller Institute for Medical Research 1919; No. 9.


Murphy TF. The many faces of pseudomonas aeruginosa in Chronic obstructive pulmonary disease. CID 2008; 47: 1534-6.


Sullivan SD, Ramsey SD, Lee TA. The economic burden of COPD. Chest 2000; 117(suppl),5S-9S.

Academic Vita

Daniel Mina

Education: The Pennsylvania State University, University Park, PA
Schreyer Honors College
B.S. in Biochemistry and Molecular Biology
Honors in Biochemistry and Molecular Biology
Expected Graduation Date: May 2011

Activities/ Service: The Pennsylvania State University
- Founding Father of Penn State Coptic Orthodox Youth Association
  - Vice President (Fall 2010- present)
  - Treasurer (Spring 2010-May 2010)
- Volunteer EMT for Penn State Emergency Medical Services (Spring 2009- Fall 2010)
  - Assistant TA for Emergency Medical Technician Course at the University (Spring 2010)
- Founding Father of Phi Kappa Sigma International Fraternity at The Pennsylvania State University (Spring 2008-Spring 2010)
  - Academic Chair- (Spring 2009-Spring 2010)
- Honor Society Inductee and Honors active member of Alpha Epsilon Delta Honor Society (2007- present)
  - Public Relations Officer and Service Chairman (2008-2009)
  - University Student Leader Scholarship
- Active member of Student Red Cross Club, 80+ hours volunteering (2007- present)
  - Residence Halls Coordinator (2008-2009)
  - On-Site Blood Drive Coordinator (2008-2010)
- Certifications: Emergency Medical Technician-Basic; Continuing education credits in Mass Casualty Incidents, Hazardous Materials, Emergency Vehicle Operations, and Emergency Care Instruction

Honors: - Dean’s List (All Semesters)
- Awards: Student Leader Scholarship, Phi Kappa Sigma Participation Based Scholarship, Taylor Community Foundation Scholarship, Cannon Memorial Scholarship

Work Experience: - Laboratory technician and undergraduate researcher for Harvill Laboratory (Fall 2009-present)
  - Prepared media, set up experiments, maintained laboratory equipment
  - Telemarketer for Person to Person Direct
  - Marketed Gevalia Coffee Products

Skills: - Basic fluency in Arabic language speaking