THE LUNG MICROBIOME IN CHRONIC OBSTRUCTIVE PULMONARY DISEASE

by

MARC ALEXANDER SZE

B.M.L.Sc., The University of British Columbia, 2009

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Experimental Medicine)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

July 2011

Abstract

Until recently the normal human lung was thought to be sterile below the larynx, but recent reports from other laboratories indicate that a diverse microbiome exists and becomes less diverse in smokers. These reports led naturally to the hypothesis that pathogens emerging from the abnormal microbiome in smokers could drive the innate and adaptive immune response that has been associated with the pathology of peripheral lung abnormalities observed in Chronic Obstructive Pulmonary Disease (COPD). The purpose of the present study was to examine this hypothesis in human lung tissue. This began with a preliminary experiment in which DNA isolated from 2 samples from a control lung were compared to DNA isolated from 5 different samples of a severe COPD lung, using 75 based pair-end tag sequencing (metagenomic sequencing). For bacteria, a weighted average genome size representing bacterial species identified was applied and the results validated using PCR and qPCR assays. This preliminary experiment was followed by a qPCR, T-RFLP, and targeted sequencing analysis of the bacterial 16S rRNA gene in DNA isolated from single samples of frozen lung tissue obtained from 8 non-smoking and 8 smoking controls, 8 COPD (GOLD 4), and 8 cystic fibrosis patients. The metagenomic sequencing conducted in the preliminary study showed that the 5 samples from a single COPD patient had an average of 2.4 ± 0.7 bacteria/1000 human genomes while the smoking control had 1.6 ± 0.8 bacteria/1000 human genomes. The qPCR results obtained from a single sample from 32 different subjects showed that on average the 8 samples/group of non-smokers, smokers, and COPD (GOLD 4) patients had 34.5 ± 21.8, 44.3 ± 47.0, and 24.1 ± 36.9 bacteria/1000 human cells, respectively, while cystic fibrosis patients had (20 ± 54) x 10^4 bacteria /1000 human cells. T-RFLP

analysis showed three distinct community compositions: smokers and non-smokers, cystic fibrosis, and COPD (GOLD 4) patients. These results confirm the presence of a small number of bacteria within the human lung of non-smoker and smoker controls and in COPD patients with a shift in bacterial composition in lungs of those with COPD (GOLD 4).

Preface

This research was approved by the UBC-Providence Health Care Research Ethics Board. The certificate number for this project is H10-00843.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

(Alphabetical Order)

3-D: Three Dimensional

BAL: Bronchoalveolar Lavage

BLAST: Basic Local Alignment Search Tool

bp: base pair

CF: Cystic Fibrosis

COPD: Chronic Obstructive Pulmonary Disease

CXC: C-X-C motif

DNA: Deoxyribonucleic Acid

dNTP: Deoxyribonucleotide Triphosphate

F: Forward Primer

FEV_1: Forced Expiratory Volume in 1 second

FVC: Forced Vital Capacity

GOLD: Global Initiative for Chronic Obstructive Lung Diseases

GWA: Genome Wide Association

HhaI: *Haemophilus haemolyticus* I

ICAM-1: Inter-cellular Adhesion Molecule-1

Ig: Immunoglobulin

IL: Interleukin

LB: Luria/Lenox Broth

Lm: Mean Linear Intercept

LPS: Lipopolysaccharide

MID: Multiplex Identifier

MMP: Matrix Metalloproteinases

MRPP: Multi Response Permutation Procedure

N/A: Not Applicable

NAPS: Nucleic Acid Protein Service Unit

NCBI: National Center for Biotechnology Information

NK: Natural Killer

NMS: Non-Metric Multidimensional Scaling

OCT: Optimal Cutting Temperature

OH: Hydroxide

PCR: Polymerase Chain Reaction

qPCR: Quantitative Polymerase Chain Reaction

R: Reverse Primer

RDP: Ribosomal Database Project

RPP40: Ribonuclease P Protein Subunit p40

rRNA: ribosomal Ribonucleic Acid

SD: Standard Deviation

SE: Standard Error

SmIgM: Surface Immunoglobulin M

SNPs: Single Nucleotide Polymorphisms

T-RFLP: Terminal Restriction Fragment Length Polymorphism

Th: T-helper

TNF-α: Tumor Necrosis Factor Alpha

Tregs: T regulatory Cells

Acknowledgements

I would like to thank both of my supervisors, Drs. Jim Hogg and Don Sin, for all their support, patience, and helpful advice that they have supplied during my time as a graduate student. I have truly learned more than I could have hoped for during my time as their graduate student.

Second, I would like to thank the members of my research committee, especially Dr. Bill Mohn. Without his helpful insight and suggestions this project may still be "spinning its wheels" so to speak. I would also like to thank the members of his lab and Dr. Pedro Dimitriu for his suggestions and assistance in the data analysis.

Third, I would like to thank the members of both the Hogg and Sin lab for all their encouragement. Finally, the hard work that both John Gosselink and Dr. Shizu Hayashi put into training me as a scientist was daunting and I would like to thank them for helping me learn how to approach problems critically and appropriately.

This project took the help of a great number of individuals and without their combined guidance, leadership, insight, and patience would not have been possible. It is because of the individuals mentioned that I have been able to complete this research project and thesis as quickly and as accurately as possible.

Dedication

To Tammy, Mom, Dad, and family

Chapter 1: Background

1.1 Natural History of COPD

1.1.1 Background

It is generally thought that chronic obstructive pulmonary disease (COPD) is preventable and can be treated with varying effect with bronchodilators and corticosteroids [1]. The disease itself is characterized by airflow limitation that is not fully reversible and is associated with an abnormal inflammatory response of the lungs to different types of harmful particles, of which cigarette smoke is the most prominent [1]. An international committee called the Global initiative for chronic Obstructive Lung Disease (GOLD) classifies the severity of COPD into four different levels ranging from GOLD 1 (mild) to GOLD 4 (very severe) COPD [2]. The diagnosis and classification of COPD is currently based on measurements of the forced expiratory volume in one second (FEV_1) and its ratio to the forced vital capacity (FVC) or total volume of air expired when no time limit is applied (measured by spirometery) [2]. The specific measurements of disease severity can be seen in the following table [Table 1].

Table 1: Breakdown of disease according to spirometry [1]

Severity	Postbronchodilator FEV1/FVC	FEV1 % Predicted
Mild COPD (GOLD 1)	≤0.7	≥80
Moderate COPD (GOLD 2)	≤0.7	50-79
Severe COPD (GOLD 3)	≤0.7	30-49
Very Severe COPD (GOLD 4)	≤0.7	<30

Historically, there were early indications that COPD was associated with smoking [3, 4]. The classic studies of the natural history of COPD by Fletcher, et al. [5] compared the decline in FEV_1 in non-smokers to that observed in current and former smokers to establish that only a minority of smokers developed the rapid decline in FEV_1 that was associated with the disability and death due to airflow limitation. Also, these studies determined that this rapid rate of decline could be slowed but not completely reversed by smoking cessation [5]. Kanner & Anthonisen, et al. [6] subsequently confirmed and extended many of Fletcher, et al's findings in the NIH sponsored Lung Health Study. These observations established that only a susceptible minority of smokers develop COPD and provided a clear indication that factors other than smoking such as air pollution, exposure to dusts in either the home or work place, childhood infections and genetic abnormalities such as alpha-1 anti-trypsin disease contribute to the pathogenesis of COPD [7]. Studies by Churg, et al. have shown that an exposure to asbestos, titanium particulate matter, and even particulate air pollution can increase fibrosis in the large and small airways [8, 9] which could lead to COPD within susceptible populations. Further studies have shown that these susceptible populations (e.g. coal miners) do indeed have worse disease and that smoking can have an additive or even synergistic effect [10]. Incidental and occupational exposure to cigarette smoke, dust, particulate matter, and air pollution are now well established risk factors for COPD. Currently COPD is projected to have an increasing prevalence world-wide and based on current projections will soon be the 4^{th} leading cause of mortality in the world by 2020 [11].

1.1.2 Phenotypes of COPD

1.1.2.1 Chronic Bronchitis

Chronic bronchitis is defined clinically by the presence of chronic or recurrent mucus hypersecretion and productive cough that occurs daily for 3 months of the year for three years [12, 13]. Its pathologic correlates include enlargement of mucus secreting glands and goblet cell metaplasia of the lining epithelium of the cartilaginous airways of the lower respiratory tract. This begins with the trachea and main stem bronchi and extends down to the smaller cartilaginous airways (i.e. bronchi) down to airways of approximately 3 mm in diameter [14]. This enlargement of the mucus secreting apparatus is associated with inflammatory immune cell infiltration, and remodeling of the muscle and connective tissues in the walls of the airways that make up the trachea-bronchial tree [15-17]. Both the increase in mucus production and mucus cell proliferation are caused predominately by cigarette smoke [18]. Mullen, et al showed that in approximately 50% of patients with chronic bronchitis the inflammatory immune cell infiltration and tissue remodeling observed in the cartilaginous airways extend down to the bronchioles (non-cartilaginous conducting airways) [19]. Moreover data initially presented by Niewoehner, et al, followed by several others, have identified very similar inflammatory immune cell infiltration and tissue remodeling in the gas exchanging tissue [20]. Although infections are implicated in the pathogenesis of the inflammatory immune cell infiltration and tissue remodeling observed in chronic bronchitis, they have mostly been associated with acute exacerbations of COPD [17, 21, 22], which are clinically defined by increased sputum volume and increased purulence in association with symptoms of dyspnea that are increased over baseline [23]. The precise role that

acute exacerbations play in the pathogenesis of COPD are poorly understood, but more frequent exacerbations are associated with greater morbidity and mortality in both non-hospitalized and hospitalized patients with COPD [24].

1.1.2.2 Emphysema

In contrast to the clinical symptoms that define chronic bronchitis, emphysema is defined by irreversible destruction of the lung tissue beyond the terminal bronchioles without associated fibrosis [25]. Prior to the introduction of computed tomography the early diagnosis of emphysema was restricted to pathological post mortem examination of lungs in the fixed inflated state [26, 27]. The pre-mortem diagnosis of emphysema was based on symptoms of dyspnea, wheezing, and signs of lung hyperinflation associated with advanced emphysematous destruction of the lungs with formation of emphysematous bullae. Further pathological studies of lungs resected for lung cancer showed that emphysema could be demonstrated in lungs from smokers without airflow limitation [28]. This observation of emphysematous destruction in the absence of airflow limitation was confirmed *in vivo* with the introduction of computed tomography [29-31]. From the classification of emphysema there are two major types that are commonly discussed: panlobular and centrilobular [32]. However a third type known as paraseptal emphysema was also characterized but has been mentioned less frequently than both panlobular and centrilobular emphysema. It is characterized by abnormal dilation of alveoli that extend to fixed structures such as the pleura, blood vessels, or the septum [33]. Panlobular emphysema is characterized by uniform enlargement of all acini contained within a lung lobule [32]. In contrast, centrilobular emphysema is characterized by

dilation and destruction of the respiratory bronchioles located in the centre of the lobules [17, 32, 34, 35]. The spatial distribution of disease also differs between centrilobular and panlobular emphysema. Centrilobular emphysema is more common in the upper regions of the lung near the apex [35-37] while panlobular emphysema affects the entire lungs [35] with a greater tendency to destroy the mid and lower portions [35, 37]. Subsequent molecular follow up studies have shown that panlobular emphysema is associated with a genetic deficiency known as alpha-1-antitrypsin [38, 39]. Although genetic components to the disease have been identified (e.g. alpha-1-antitrypsin) and other environmental factors such as dust, coal, and particulate matter have been associated with airway remodeling and emphysema [8-10], smoking is the leading risk factor for the disease [15].

1.1.2.2.1 Elastase/Anti-Elastase Balance

The destruction of the elastin frame work within the lung resulting in an imbalance between elastin degrading enzymes (elastase/proteases) and elastin preserving enzymes (anti-elastase/antiprotease) [15] is one of the leading hypothesis as to the cause of emphysema [40, 41]. The two major molecules that are studied in this area are neutrophil elastase and MMPs (extracellular matrix degrading proteins) [15]. It has been shown that smoking can influence neutrophil transit time as well as increase their retention within the lungs [42]. Furthermore, neutrophil elastase can be released in response to infection and can also inhibit and damage the ciliary activity of respiratory epithelial cells [43]. This build up of neutrophils within the lungs coupled with their response to cigarette smoke and infection could cause large amounts of tissue damage [44]. One of

the major points of the vicious cycle hypothesis [45] is that these defensive impairments may promote pathogenic microbial colonization which in turn activates the immune system [45], causing an additional release of neutrophil elastase and MMPs that further degrade the extracellular matrix protein structure within the lungs. Excess neutrophil elastase and MMP 8 can also be released in response to cigarette smoke [46, 47]. Essentially, there is an existing environment (either from dust inhalation, smoking, or other environmental exposures) as well as a possible genetic disposition (alpha (1)-antitrypsin) in which extracellular matrix degrading proteins are being released and actively destroying structural components of the lung. This process can then be further accelerated by pathogenic microbial infection which can cause an inflammatory response (both innate and adaptive) that can lead to a further increase in elastolytic activity [43, 45] which manifests in a worsening of COPD severity [45] in the form of a decrease in lung function and acute exacerbations.

1.1.2.3 Small Airways Disease

These airways measure less than 2 mm in diameter and are found from the fourth to twelfth generation of airway branching [48]. The small airways are made up of both ciliated and non-ciliated epithelial cells and have a thin layer of connective tissue and smooth muscle surrounding the epithelial cells [48]. In COPD, an increase in inflammation around the small airways and thickening of the airway wall can be observed [49-51]. Research has shown that the small airways play a very crucial role in individuals with COPD [48, 52-54]. For example, both the increased inflammation and fibrosis seen in small airways disease has been shown to correlate with a decrease in

FEV_1 [51]. These abnormalities in the small airways are not limited to just inflammation and fibrosis but also mucus plugging, which can be associated with the severity of airflow obstruction [55]. This obstruction of the small airways also causes decreases in FEV_1 as well as the ratio of FEV_1 / FVC [48, 55]. The reduction of elastic recoil in these airways are due to the large amount of fibrosis, inflammation, and mucus plugging, seen in susceptible smokers [48, 55]. Further , the small airways of individuals at different stages of disease has shown that airways thicken and the lumen narrows as the disease severity progresses [50]. A large amount of the limitation of airflow in COPD can be explained by a prolonged time constant for the lung to empty [56], which is mainly due to an increasing resistance found in the small airways and loss in elastic recoil pressure due to emphysematous destruction [56]. However, in disease, these two processes can occur both independently of each other as well as together. The main reason why the resistance increases with changes in the small airways is because when these airways are added together, they have a very large cross sectional area which enables them to have a lower total airway resistance [56]. Thus in healthy individuals the small airways do not contribute a great amount to total airway resistance [48]. However, large scale destruction of these small airways, mucus plugging, or increased fibrosis and remodeling can have a very drastic effect on the total airway resistance [56] which can ultimately lead to an impairment of lung function in disease. As a result the small airways play a significant part in total airway resistance in COPD [52].

1.1.3 The Dutch Hypothesis

The discussion of pathogenesis and natural history of COPD were dominated by two fundamentally different hypotheses in the latter third of the 20th century. The first of these came to be known as the Dutch hypothesis and is based on the notion that bronchial hyperreactivity is an underlying feature of both COPD and asthma [57]. Quite simply this hypothesis considers asthma and COPD as a single disease entity [58, 59]. Although it is unlikely that asthma and COPD are the same disease, there is evidence that concordance between the two diseases exist. This was best shown in a 20 year prospective study that found that the strongest risk factor for developing COPD, based on the hazard and attributable risk ratio, was asthma [60]. Although concordance exists there are still very clear differences between the two disease sets [58]. For example, in asthma there is commonly an increase in eosinophils, T-helper 2 lymphocytes (Th2), and activated mast cells [61]. However, in COPD, infiltration in to the lungs comes mostly from neutrophils, macrophages, and cytotoxic T-cells [61]. Interestingly, it has been noted that in severe asthma the cell infiltration begins to approximate that seen in COPD [58]. Additionally, COPD in asthmatics who smoke do have a more severe asthma response, an increased rate of admission to hospital, and a reduction in their general response to corticosteroid treatment [62]. Further, it has been found that the decline in lung function of smokers who are asthmatic are more severe but the effect seems to be additive [63]. A family history of asthma as well as smoking status can increase the incidence rate of COPD in a population [65]. It has also been shown that maternal smoking can increase the odds ratio of a child presenting with wheezing and coughing [64] which are common of both COPD and asthma. Additionally, there are some

encouraging results for the Dutch hypothesis from the candidate gene approach which has yielded interesting information. One such finding is with respect to SNPs (single nucleotide variation that occurs between individuals) in the *ADAM33* gene. Researchers have shown that *ADAM33* SNPs effect the development of low lung function (specifically airway resistance) in children ages 3-5 [66, 67]. Concurrently, *ADAM33* has also been shown to be associated with an accelerated FEV_1 decline in a select adult population [68]. The *ADAM33* gene mutations may play a role in the development of asthma in early life while in later life predispose select populations to an accelerated lung function decline. At the moment more studies need to be completed into the exact function of the *ADAM33* gene [66] to confirm this hypothesis. Although there are some promising areas of research regarding the Dutch Hypothesis, to date, not a single GWA study has been able to identify common genes between both asthma and COPD [66].

1.1.4 The British Hypothesis

The second hypothesis is known as the British hypothesis which states that recurrent infections of the bronchial airways are the driving force behind why some smokers develop COPD [69, 70]. These infections were thought to present themselves clinically as acute exacerbations [70]. The hypothesis was first established in the 1960's by British investigators and later Fletcher, et al. tested whether this hypothesis held true in a prospective study of working males [5] where they looked at the frequency of respiratory infections, sputum quantity and quality, and lung function decline. Fletcher, et al ultimately rejected the hypothesis that there was a natural progression from tobacco smoking to chronic bronchitis, repeated chest infections, progressive airflow limitation,

disability, and death. This was primarily based on his group's observations that many smokers that developed COPD never had symptoms of chronic bronchitis [5]. However, recent data shows that individuals who have an increased production of mucus and cough have a slightly accelerated decline in lung function and that the bacterial counts in the sputum from these individuals can be related to this decline in lung function [71, 72]. It has also been found in stable COPD, patients with identifiable pathogens have significantly different quality of life and inflammatory profiles than those that have bacteria but no identifiable pathogens [69]. Additionally, Kanner & Anthonisen, et al. showed that smokers visited their physician more often for lower respiratory tract infections than non-smokers [6]. When the smoking group was sub-divided, the continuous smokers had the highest visit rates for lower respiratory infections versus both sustained quitters and intermittent smokers [6]. Within this smoking group, those with the most physician visits for lower respiratory tract infections had a slightly greater decrease in their FEV_1 [6]. Further, Donaldson et al. found that individuals who had more frequent exacerbations also had an accelerated decline in FEV_1 [73]. This evidence would suggest that lower respiratory infections influence the decline of FEV_1 and the progression of COPD. The idea that bacteria could influence lung function was expanded on from research showing that over the course of a year both the change in bacterial count and absolute bacterial count correlated significantly with a decline in FEV_1 [74]: where the largest decrease in lung function could be found in patients who had a different bacterial species cultured at the start of the study versus the end of the study [74]. Not surprisingly, some of these studies also show that with higher bacterial loads there were higher measurable elastase activity and inflammatory cytokines [74, 76].

Additionally, Sethi, et al. showed that new strains of the same bacteria could be associated with a greater risk of exacerbations [21]. This research along with a recent longitudinal study on patients with COPD that found a positive association between increasing exacerbations and increasing COPD severity [75] provides strong evidence for the original British hypothesis. Sethi, et al. propose that a vicious cycle is in action, where continuous microbial insults amplify the inflammation and drive the remodeling process [45]. Although they believe that bacteria are the root cause there are others who propose a mechanism similar to the vicious cycle hypothesis but believe viruses are the important pathogens [77]. Essentially, the vicious cycle hypothesis posits that after an initiation factor (smoking) various impairments to normal lung function (e.g. mucociliary clearance) can lead to bacterial or viral colonization (e.g. *Haemohilus influenzae*) which can lead to airway epithelial injury [45] and progression of COPD. However, there is an alternative hypothesis that deserves attention. As mentioned in the Wilkinson paper a stable bacterial isolate or population that did not change over the course of a year meant less of a decrease in FEV_1 [74]. It is possible that a stable bacterial population, within the airways, could be protective and that constant smoke exposure could select for colonization of the airways by new species as well as new strain growth that then leads to the pathological changes associated with COPD. However, neither this hypothesis nor the vicious cycle hypothesis has been verified.

1.1.5 Innate Immune Defense Mechanisms

The major barrier defense in the lungs is the mucociliary transport system which consists of cilia and mucus. The cilia transport mucus that has trapped foreign material out of the lungs. Although smoking can cause a dysfunction in cilia [78] and cilia beat frequency [79] it has also been shown that bacteria such as non-typeable *Haemophilus influenzae* can also inhibit ciliary beat frequency [80]. Therefore infections can potentially cause an increased inhibition of mucus clearance in smokers by causing a further reduction in cilia functionality. Research has shown that mucus production and hypersecretion in the progression of severity of COPD is not certain [61]. Although there is evidence to suggest that mucus hypersecretion may be a risk factor for an increased decline in lung function [81], it has been reported by the same group that chronic cough and mucus production in smokers who have normal lung function does not predict a later development of COPD [82]. The barrier defense is supported by the innate immune system and its acute inflammatory cells. Previous data have shown that there is an increase in acute inflammatory cells, such as neutrophils, macrophages, and eosinophils as COPD severity increases [50, 83, 84]. However, some research suggests that the innate host defense may be impaired or suppressed by smoking [85]. It should be noted, that the authors of this study relate impairment of the innate immune system with increased susceptibility to microbial infection [85]. In general, the majority of findings show that there is an enhancement in the innate immune response, characterized by an increase in macrophages and neutrophils, that can help drive the inflammation and remodeling in COPD. Possible reasons for this stem from the secretions released by neutrophils: serine proteases (of major concern is neutrophil elastase), cathepsin G

(protease), proteinase-3 (serine protease), and matrix metalloproteases (in particular MMP-8 and MMP-9, which are a collagenase and gelatinase respectively) [61]. Increases in macrophages have been observed using different sampling techniques (BAL, airways, lung parenchyma, etc.) in patients with COPD [61]. Macrophages can correlate well with COPD severity [86] and can be activated by cigarette smoke to release a plethora of inflammatory mediators (e.g. TNF-α, IL-8, CXC chemokines, etc.) that can modulate both the innate and adaptive immune response in COPD [61]. Macrophages also use pattern recognition receptors (e.g. Toll-like receptor 4) in order to identify pathogen associated molecular patterns (PAMPs) and engulf foreign organisms. Therefore, upon exposure to microbial pathogens, the existing elevated inflammatory state of smokers could be further increased by an exuberant macrophage response which could lead to a manifestation of an acute exacerbation. The term macrophage polarization has recently come to the forefront of chronic inflammatory research and it has been identified that macrophages can be separated into two distinct subsets, either M1 or M2 macrophages [87]. The classically activated macrophages (M1) are induced by IFNγ (a soluble cytokine involved in innate and adaptive immune response) or with a microbial stimuli [87]. The M1 macrophages typically have high amounts of IL-12, IL-23, and a low amount of IL-10. They induce the production of reactive oxygen species and nitrogen intermediates as well as other inflammatory cytokines. They also promote a Th1 response and help to protect against intracellular parasites and tumors [87]. On the other hand, it has been found that IL-4 and IL-13 can induce an alternative type of macrophage, termed the M2 macrophage [88]. The M2 macrophages contain low IL-12 and IL-23, but a high expression of IL-10. Some M2 macrophages can produce

inflammatory cytokines while others cannot. They tend to have a high expression of scavenger, mannose, and galactose-type receptors. M2 macrophages are thought to participate mostly in the Th2 response and help promote tissue repair and remodeling [87]. The polarization and the various functions that macrophages have in inflammation and tissue remodeling can help explain the simultaneous loss in tissue in some areas and gain of tissue in others within COPD lungs [89]. More specifically the polarization of macrophages could help to explain why there is wall thickening in the airways that happens at the same time that there is airway or parenchymal destruction.

With respect to eosinophils some studies have found an increase in activated eosionphils while other studies have not [90]. There is evidence that due to neutrophil elastase release there is an increased number of degranulated eosinophils and that the elevated amounts of eosinophils can help predict whether or not a patient will be responsive to corticosteroid therapy [91, 92]. However, neutrophils and macrophages are the two major contributors towards disease with respect to innate inflammatory cells. Neutrophils are actively recruited by neutrophil chemotactic factors to the airways but what causes this recruitment aside from smoking is not well characterized [56, 61]. Similarly, an increased recruitment of macrophages to the lungs by cytokines could explain the increased number of macrophages observed, as COPD progresses [56, 61].

1.1.6 Adaptive Immune Response

It has been established that smokers have an increased inflammatory cell burden within their airways [52, 53] which is partially due to an active adaptive immune response [50]. This adaptive immune response consists mostly of lymphocytes that fall under the

category of T-cells or B-cells that target very specific antigens on foreign molecules. The most studied T-cells contain surface markers for CD4 (Th1 and Th2) and CD8. However, there have been discoveries of new subsets of T-cells that include Th17 (IL-17A), Tregs (FoxP3), Th3 (marker unkown), and natural killer T-Cells (TCRαβ) [93-96]. In contrast B-cells contain similar surface markers (e.g $CD45^+$ and SmIgM) but produce antibodies of different types (IgM, IgG, IgE, IgA, and IgD) [93]. However, there is evidence to suggest the existence of regulatory B-cells [97, 98]. Studies in COPD have found that B cells, $CD4^+$ T-cells, and $CD8^+$ T-cells are present in the airway tissue and that these cells are at a much higher prevalence in the most severe forms of COPD [50, 56]. It has also been shown that an increase in these cells can be associated with a decline in FEV_1 [83, 84]. This adaptive immune response, which can include class–switched B-cells [99] are coupled with an increased prevalence of antigen presenting cells (e.g. dendritic cells) of the innate immune system [100]. As disease severity progresses an increasing prevalence of bronchial associated lymphoid tissue (BALT) also known as tertiary lymphoid follicles can be found [50, 56, 101]. These tertiary lymphoid follicles are not seen in non-smokers, sporadically in smokers, and have a drastic increase in prevalence in GOLD 3 (severe) and GOLD 4 (very severe) stages of COPD [50]. A germinal centre can be found in these tertiary lymphoid follicles and upon immunohistochemical staining, these follicles have been found to contain the following immune cells: $CD4^+$ T-cells, $CD8^+$ T-cells, NK cells, B-cells, and dendritic cells [50]. The appearance of tertiary lymphoid follicles that are associated with the small airways and increase in prevalence as disease severity increases would suggest that an adaptive immune response may be actively involved in the pathogenesis of COPD. The Th1 adaptive immune response is thought to be the

prominent player in COPD [102]. However, more recent studies have shown that it is the Th17 side of the adaptive immune response that can have a larger role in various disease processes [103]. This Th17 population has been shown to play a role in other chronic inflammatory diseases such as, rheumatoid arthritis, multiple sclerosis, and inflammatory bowel disease [105]. With respect to COPD, Th17 cells are very good at producing IL-6 which in turn can upregulate Muc5AC and Muc5B (mucins) production by lung epithelial cells [104]. Thus there may be other important subsets, other than the classical Th1 response, active during the disease process. Regardless of what part of the adaptive immune response may or may not be up regulated, the target of this adaptive immune response has yet to be identified.

1.1.7 Antigen Targets

With respect to antigens from bacterial pathogens that may drive the adaptive immune response, some interesting observations have been made. New strains of the same bacteria have been shown to cause acute exacerbations in COPD patients [21, 106]. This would suggest that colonization by bacteria in the lungs of COPD patients occur and that in the majority of cases the bacteria do not illicit an immune response unless a new strain of the same bacterial species develops [106]. These findings suggest that the new strain antigens may drive the adaptive immune response characterized in the small airways [21, 50]. It may also be possible that in the most severe forms of the disease, antigen targets of infectious and pathogenic microbial species could cause exacerbations [21, 107-109] and result in disease progression. However, many of these studies only investigate single pathogenic infections (e.g. latent adenovirus infection [108, 109]) or changes in gene

expression of single pathogenic bacteria [21]. Since there has not been one clear pathogenic microbial species that has been identified as being associated with disease progression it is likely multiple microbial species are involved.

Alternatively, anti-elastin is a popular possible target for an autoimmune adaptive immune response. However, the literature is unclear as to whether or not a response against elastin is evident in COPD. For example, one study found an association with anti-elastin when looking at pack years smoked [110] while another study found no evidence of any difference in anti-elastin antibodies in the blood of patients with COPD versus controls [111].

1.1.8 Acute Exacerbations in COPD

1.1.8.1 Overview

As previously mentioned acute exacerbations present clinically in the form of increased sputum volume, increased sputum purulence, and increased dyspnea over baseline [23]. Recently, this definition has been expanded and a new working definition made: Acute exacerbations are defined as a sustained worsening in an individual's state away from what can be seen or expected from day-to-day variations which happens quickly and necessitates a change in the individuals regular medication for COPD [112]. Acute exacerbations results in approximately 500,000 hospitalizations per year in the United States alone [113] and severely influence the quality of life of patients that experience recurrent exacerbations [114]. Although the data collected by Fletcher, et al suggested that acute exacerbations had little effect on the subsequent decline in lung function, the Lung Health Study showed that the decline in FEV_1 was accelerated in those who

experienced exacerbations and continued to smoke [69]. Further evidence indicates that acute exacerbations accelerate the progression of COPD [21, 73, 74] and that bacteria, viruses, and other pathologic microorganisms are implicated in this process [115]. It is believed that bacteria account for around 50% of all acute exacerbations [116] while viruses account for approximately 30% [117] with the last 20% having an unknown associated cause.

1.1.8.2 Other Causes

Approximately 20% of all exacerbations cannot be attributed to a microbial agent. However, there has been growing evidence that pollution may be involved in this process. Anderson, et al. found that there was an increase in hospital admissions for COPD that correlated with air pollution [118]. Specific follow up studies, investigating both the sputum and bronchoalveolar lavage of individuals without COPD, have found that there is a generalized increase in both inflammatory cells and inflammatory markers when these individuals were exposed to air pollution versus filtered air [119, 120]. More recently Arbex, et al. have shown that acute exacerbations increased on the same day that particulate matter measurements increased [121]. These results need to be repeated, but if they hold true this would show a reproducible correlation between pollution and acute exacerbations in COPD.

1.1.8.3 Viral Infections

The current literature on the type of virus that is most prevalent during acute exacerbations has been varied [122]. The majority of research shows that the picornavirus (of which rhinovirus is a part of) is the most common virus [122, 123]; however, some research have shown that influenza or even RSV is the most common

virus in acute exacerbations of COPD [124-126]. Although rhinovirus may be important in exacerbations it has been shown that there is a seasonal variation in its prevalence amongst patients [123]. The highest prevalence was during the autumn season and the lowest prevalence was during the winter season [123]. Even though rhinovirus has been shown to be very important in acute exacerbations many other viruses may still have their own role to play in disease progression. For example, adenoviral infections along with cigarette smoke exposure can lead to an accelerated lung function decline, as shown by Fletcher and Peto [108] . It has been suggested that an integration of the adenoviral DNA into the human genome, in particular the E1A region, could predispose these cells to destruction by cytokines [108]. Extending these results PCR studies have shown that heavy smokers have adenovirus DNA in their lung tissue [108] and this can be confirmed using *in situ hybridization* staining methods of lung tissue [108]. What is fascinating about the E1A region integration of adenovirus is that upon addition of LPS, ICAM-1 expression was affected in individuals that had this adenovirus E1A region integrated into their genome [127]. This opens up the possibility that certain viruses that incorporate themselves into the host genome, that may not cause acute exacerbations, could cause dysfunction in host immune response.

1.1.8.4 Bacterial Infections

Acute infections with bacteria have been well documented to illicit various forms of pneumonia [128, 129]. However, bacteria also seem to be an important factor in COPD. Almost fifty percent of all acute exacerbations can be attributed to an acute infection with bacteria [117, 130]. Although commensual bacteria can cause chronic disease [131], within COPD, the most commonly studied bacteria are pathogenic bacteria such as

Haemophilus influenza, Moraxella catarrhalis, Streptococcus pneumonia, and *Pseudomonas aeruginosa* [21, 132]. Although acute exacerbations are important in COPD and a number of them are caused by bacteria [133], treatment with antibiotics has not been shown to be consistently advantageous to patients. This is because many studies on antibiotic effectiveness in COPD have come back with mixed results [134]. One possible reason for this could be attributed to a rise in antibiotic resistant respiratory pathogens [135, 136]. However, two recent meta-analysis studies of the available trials of antibiotics versus placebo have found an improvement in lung function in patients treated with the antibiotic intervention [137, 138]. Antibiotic usage has been demonstrated to have a very clear beneficial effect in COPD patients and mechanically ventilated COPD patients with severe exacerbations [134, 139]. Though pathogenic bacteria have been found and cultured in patients with acute exacerbations there is a significant proportion of bacterial cultures that do not grow any bacteria [139]. This finding is supported by the fact that approximately 20 percent of all acute exacerbations are caused by unknown factors [117]. This leads to the possibility that at least some of these exacerbations could be caused by unculturable bacteria. Subsequently, these bacteria may form a resident bacterial microbiome in the lungs and changes that occur in disease can facilitate the increased prevalence of pathogenic bacteria, both culturable and unculturable, resulting in acute exacerbations. Finally, there is some information that suggests coinfection, with both bacteria and virus and the interaction between the two could play a crucial role in COPD, in particular with RSV and rhinovirus infections [130]. Whether a persistent viral infection leads to acute bacterial exacerbations or vice versa is not clearly known. Some research has shown that bacteria can create a chronically

inflamed airway, that has a higher than average inflammatory profile (measured by IL-8, leukotriene B4, neutrophil elastase, and TNF-α) compared to individuals who did not have bacteria in them [140]. This larger inflammatory state could predispose an individual to coinfection [130], that is either viral or bacterial which could ultimately lead to progression of COPD.

1.2 The Microbiome

1.2.1 Introduction

Microorganisms can live and form active communities known as microbiomes in many different environments. A microbiome is a collection of different bacteria that come together to form their own unique ecological community within a larger environment. These microbiomes can be found in a variety of diverse environments from inert objects like soil or subterranean caves to living organisms such as coral, fish, and humans [141, 142]. However, it is the microbiome that inhabits various niches of the human body that is of great importance with respect to possible disease outcomes and mechanisms. Recent investigations have now started to show how these bacteria interact both with each other within their microbiomes and as a whole with the human host.

1.2.2 The Gastrointestinal Tract

With respect to the human bacterial microbiome, the largest amount of data that has been collected is on the gut. It is well known that we need the bacteria in our gut in order to properly absorb nutrients, such as vitamin K. It is also known that the diverse bacterial

flora can inhibit pathogenic bacteria from gaining a foot hold and causing infection. Recent reports indicate that this microbiome may even be protective against chronic inflammatory diseases like Crohn's disease [143]. Upon birth our gastrointestinal tract is almost immediately colonized and a microbiome starts to form [144, 145]. This primitive microbiome is relatively different from infant to infant [144]. However, each of these new microbiomes, as measured at one year of age, converges towards the bacterial make-up of an adult microbiome [144]. This adult microbiome that colonizes our gut also varies greatly from person to person [141, 146]. Some research has shown that this variance may be caused by how the differences between individuals (genes, gene environment interactions, etc.) affect the makeup of their gut microbiome [147]. However, other groups postulate that this variability can be explained by what is known as the neutral theory of community assembly, which states that species have a functional redundancy [141]. This theory states that even if the individual species are different for a given niche their functionality or what they do in this respective ecosystem would be similar [141, 148]. Thus even if there is a large difference in individual bacterial species between humans the different bacteria's functionality within this ecosystem or microbiome would still be the same [141]. Additional research has found that the unique bacterial species that live in the gut seem to be very resistant to changes but can still be affected with large dosages of antibiotics [149]. A clear shift in the microbial flora can be immediately measured after antibiotics are given to individuals [149]. A microbiome shift can also be seen in diseases, such as obesity. In mouse models it has been shown that by changing the gut microbiome of control mice to the microbiome of mice that had the ob/ob gene mutation, obesity as measured by an increase in body fat percentage could

be induced in the control mice [150]. However, it has not been conclusively demonstrated that the microbiome is what causes the obesity in these mice. There is very likely a complex interplay between genetics and environment that is occurring, where the predisposition of the ob/ob gene allows for a shift in the microbiome that is better suited for energy harvest [150]. Other researchers, through anecdotal observations, have noticed that some infants experience a large drop in their total number of bacteria and the type of bacteria present with antibiotic exposure [144]. Thus, one day, it may be possible to alter our initial gut microbiome to a particular make up of bacterial species.

1.2.3 Other Microbiomes

It has been shown that the microbiome on humans can vary by location [151]. For example, the microbiome of the gut will be different from what was sampled from the skin [151]. Even the sampling of the skin itself can show regional variation, such as differences in the microbial community between the arm pit and palm [151]. This would suggest that there are many different bacterial micro environments existing on and within the human body. This adds a further complexity to what has already been mentioned with respect to the bacterial microbiome and its role in disease. Thus not only are there differences in microbiomes between individuals but also differences in microbiomes within individuals. This section will give a brief overview of the microbiomes of other micro environments within and on the human body. The surface of our reproductive organs is another micro environment where it has been shown to have a microbial community that shifts in disease. A shift in the vaginal microbiome away from the normal flora, is associated with a change in pH [152] which can result in disease [152,

153]. More specifically an increase in community composition of the vaginal microbiota away from the normal bacterial inhabitants of *L. crispatus* and *L. iners* has been shown to be associated with bacterial vaginosis [152, 153]. The oral cavity also contains a large number of bacterial species and has been shown to contain a microbiome that is quite heterogeneous in nature [154]. Though biofilms are a natural occurrence and needed for health, in the oral cavity, their presence can help shield harmful bacteria from being recognized by the host immune system [154]. The nose microbiome has been shown to be very diverse in healthy subjects and less diverse within the hospital population, in particular this change has been most noticeable in inpatients [155]. This reduction in community compostion could be caused by an influx of pathogenic bacterial growth (e.g. *S.aureus*) and over abundance which may lead to disease [155]. From this literature one can surmise that disease can result from a shift in the microbiome away from normal. However, this shift is not always due to a drop in community composition like that seen in the nose microbiome. Instead disease can also come from a shift that leads to an increase in community composition like that seen in the vaginal microbiome. Although there is clear evidence that a microbiome can exist in the mouth, the nose, the skin, and the genitals of human individuals can a microbiome possibly exist in the lower respiratory tract of the lungs? Additionally, can changes in this microbiome composition have an impact on disease like in the gut and other areas of the body? Very recently the answers to these questions have started to emerge.

1.2.4 The Lung and Airborne Microbiome

It has generally been thought and taught that the lower respiratory tract is a sterile environment due to the mucocilliary transport system, which acts as a defense response against environmental insults. The mucus acts as an adherent, trapping various chemicals and organisms (bacteria, fungi, and viruses) while the cilia beat in a synchronized fashion to clear the mucus and move it into the oropharynx where it can be swallowed or spat out [156]. It is the impairment of this system and the subsequent exposure of the sterile lower respiratory tract to pathogens that is generally thought to be the major cause of infectious disease pathogenesis [156]. The impairment of this system clearly causes the buildup of bacteria and mucus in diseases such as cystic fibrosis, which ultimately leads to the colonization of the lower respiratory tract by bacteria [157]. However, what needs to be investigated is whether or not we may have a resident microbiome that already exists in the lower respiratory tract and whether this bacterial community is changed, destroyed, or affected by the introduction of foreign, possibly pathogenic, bacteria or other microorganisms. Current research now suggests that there could be a microbiome of low density that lives in the lower respiratory tract and shifts in diseases like asthma and COPD [158, 159]. Hilty, et al. performed qPCR on bronchial brushing samples from control, asthma, and moderate COPD individuals and found no significant difference in total number of bacteria [158]. However, after utilizing primers for PCR amplification from previous microbiome research [160] and cloning of the PCR products for sequencing, identification of the bacterial genera and even species was achieved [158]. They found that though the total number of bacteria did not change between controls and disease (asthma and COPD) there was a shift in the community composition of bacteria

present in control individuals and disease individuals [158]. This paper was followed by Huang, et al. which identified a microbiome in eight COPD subjects. There has now been a quick flurry of research into the microbiome in the lower respiratory tract, and the interaction that these two may have on each other with respect to health and disease [161-164]. Although many of the findings are preliminary they tend to agree with what was published earlier by the Hilty group, that there is a microbiome in the lower respiratory tract that exists in health, and shifts in disease.

Concurrently, there have been a number of papers that have shown that there is, in fact, a possible airborne microbiome that can exist indoors in urban settings and that this airborne microbiome can be relatively diverse [165, 166]. The exact source of this airborne microbiome has been postulated to come from numerous sources [165]. However, a popular theory that has been put forward is that the actual source of the bacteria from the airborne microbiome are from humans themselves [166]. The fascinating aspect of the airborne microbiome is that if the source is from humans then the lung microbiome may be a contributing factor. The lungs, mouth, and nose may play a larger part in the formation of niche communities within this airborne microbiome which may help to harbor harmful microbial pathogens that could subsequently be passed on to susceptible individuals. Alternatively, this also opens the possibility that shifted microbiomes in disease could have an adverse effect on the bacterial communities within the air and this could then be translated to new susceptible individuals. However, whether the lungs have a measurable microbiome that shifts in disease needs to be verified first before any subsequent studies are explored with respect to possible links between the airborne microbiome and the lung microbiome.

Chapter 2: Experimental Approach

2.1 Working Hypothesis

I hypothesize that a microbiome exists within the lungs and that changes in this microbiome are associated with the pathogenesis and progression of the severity of COPD.

2.2 Specific Aims

The experiments designed to test this hypothesis had the following specific aims:

1.) To validate recent reports that the lung contains a microbiome by utilizing PCR, qPCR, and metagenomic sequencing.
2.) To determine the variance of the severity of the lesions in different regions of lungs from patients with COPD is associated with changes in the microbiome.

Chapter 3: Methods

3.1 Introduction

Rather than utilizing sputum, broncheoalveolar lavage, and bronchial brushings [158, 159, 161, 162] as starting material the present experiments began with samples of human lung tissue. This was done in order to reduce the possibility of contamination from the sputum, lavage fluids, and bronchial brushings which have to pass through either the nose or mouth in order to collect samples. Moreover, the opportunity to measure the microbiome in tissue specimens provided the additional advantage that allows the tissue response to be measured in the same tissue as that used to demonstrate the presence of a microbiome.

3.2 Tissue Processing and Labeling

The James Hogg Research Centre Biobank was started in 1979. It contains over 30 years of patient information as well as tissue samples (cores). This unique resource allows for the parallel investigation of both molecular and pathological aspects of disease. The lung tissue from smokers, COPD (GOLD 4), and cysitic fibrosis patients were obtained from individuals who underwent either lung reduction surgery or transplantation. The non-smokers tissue was obtained from donors whose lungs were not used for transplantation. The patients whose tissue is held within the Biobank have all given their consent for their tissue to be used in research. Further precautions are taken by utilizing unique identifiers for every subject so that the patient's privacy is protected.

The unique processing of the lungs allows for optimal preservation of morphology and data extrapolation. Lungs were inflated using OCT, then frozen over liquid nitrogen, then sectioned into transverse slices. The sectioned lung tissue was cored as samples with a diameter of approximately 2 cm. Two different numbering methods were utilized to label the samples. The first method assigned a unique four digit identifier with an associated lung section position number and a tissue sample location letter (e.g. 1111-2AA). The unique four digit identifier corresponds to the individual from whom the lung the sample was taken, the associated lung section position number (e.g. -2) corresponds to the height within the lung the lung section was taken from (with lower numbers being closer to the apex and higher numbers being closer to the base) and the tissue sample location letter corresponds to the position on the lung section from which the tissue sample came (e.g. AA). Thus the smaller numbers (e.g. -1) corresponds to a higher lung section location relative to larger numbers (e.g. -5) and if both have the same four digit identifier (e.g. 1111) then they both came from the same individual but were taken at different heights. The second methodology retains the unique four digit identifier but the associated number after the identifier corresponds to tissue sample location within the lung as a whole (e.g. 1111-2). For example, if one associated number was low (e.g. -2) and another associated number was high (e.g. -5) and both had the same four digit identifier (e.g. 1111) then the low associated number (e.g. 1111-2) would be higher in the lung versus the higher associated number (e.g. 1111-5) which would be lower in the lung. An overview of the two different numbering methods can be seen in [Figure 1].

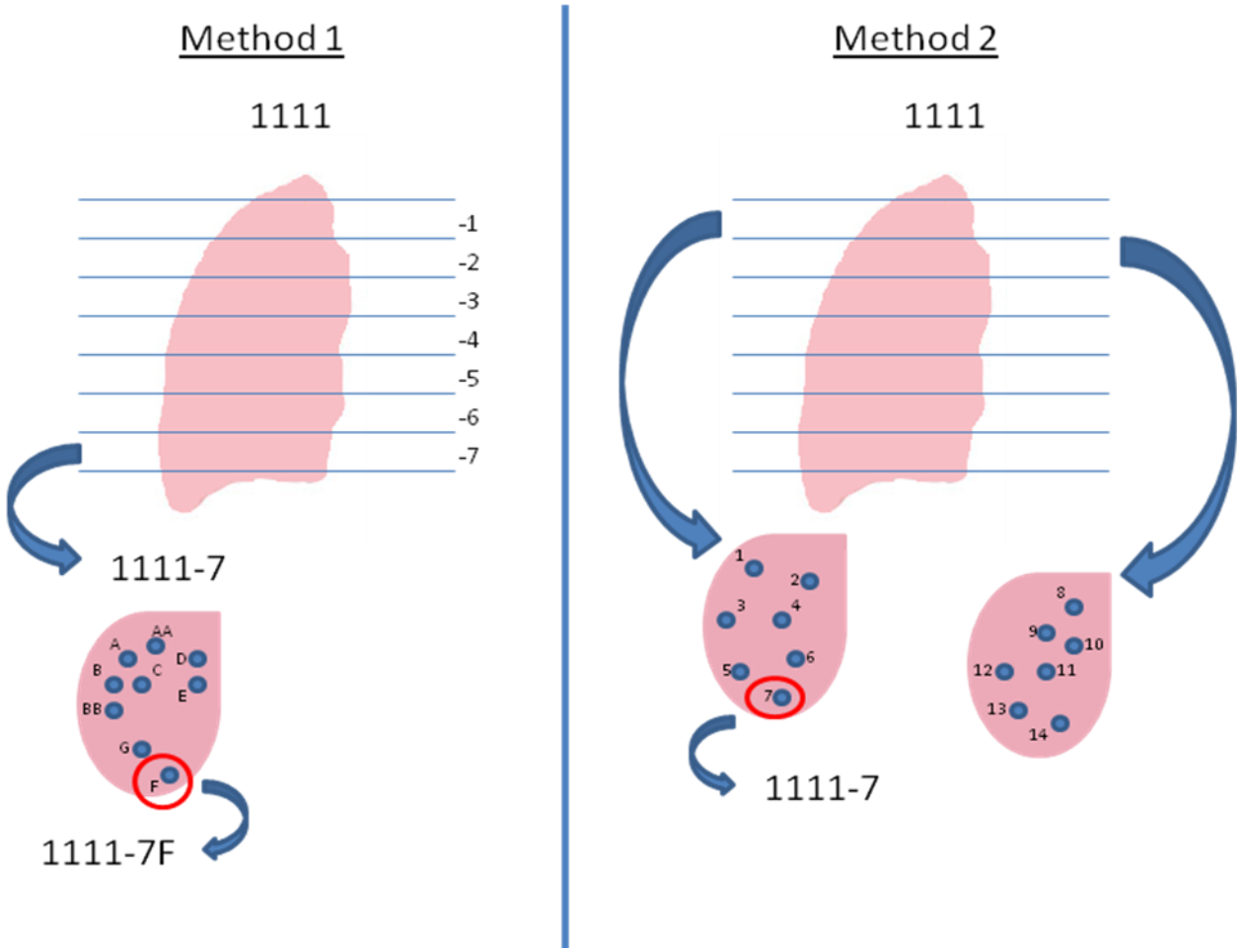

Figure 1: Diagram of lung tissue numbering
It should be noted that for method 2 the numbering was sometimes reversed so that the higher numbers were closer to the apex and lower numbers closer to the base. For method 2 numbering was also dependent on whether a whole lung or a piece of lung tissue was obtained.

These lung tissue samples were subsequently cut in a biosafety cabinet to give lung samples weighing between 10-30mg. DNA extraction was then performed using a DNeasy extraction kit produced by Qiagen (Maryland, USA). Once DNA was extracted, it was analyzed by Nanodrop (Delaware, USA) for quality and quantity using OD 260/230, OD 260/280, and OD 260 readings. Once quality and quantity had been assessed various downstream DNA applications could then be used (Ilumina Sequencing,

454 Sequencing, PCR, qPCR, etc.). Furthermore, these lung samples could be sectioned, placed on slides, and stained using different histological techniques if needed. These slide sections could then be analyzed using quantitative histology and compared to the various downstream DNA processes. An overview of these methods can be seen in [Figure 2].

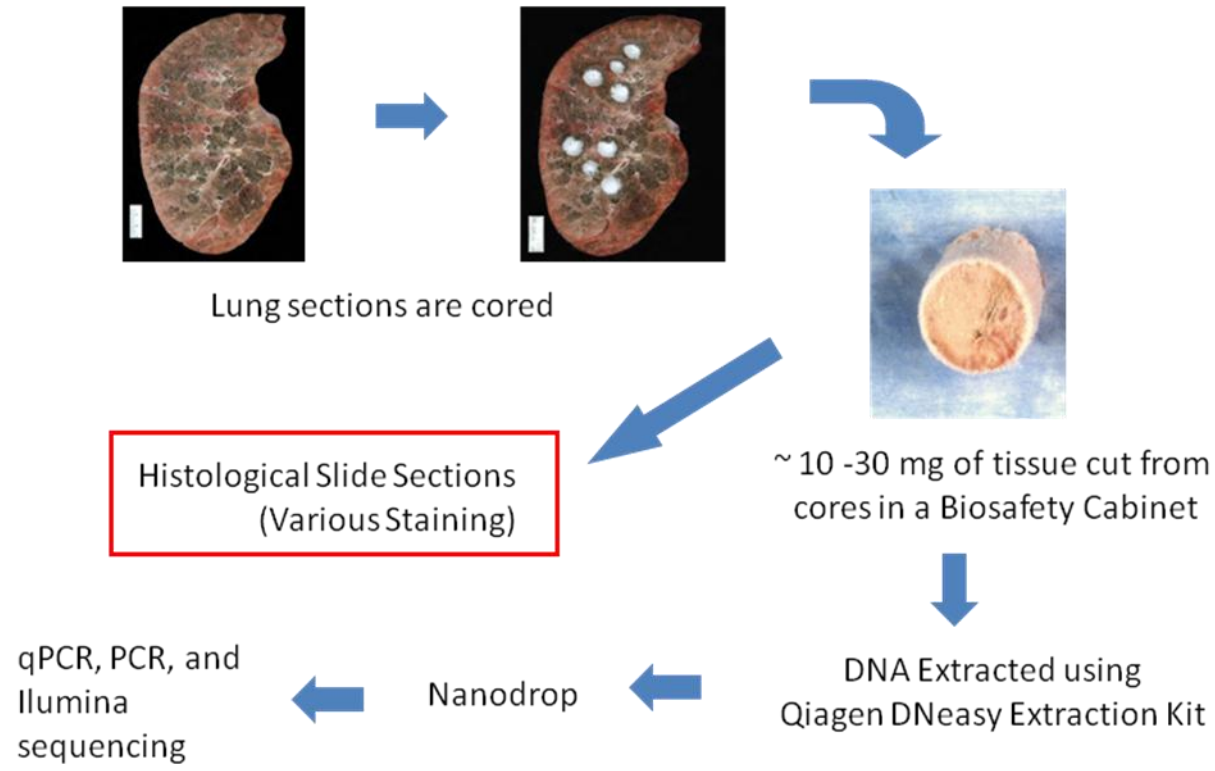

Figure 2: Diagram of tissue processing.

Additionally, extraction negative controls were included during the tissue extraction process. An extraction negative control was a sample put through the same tissue processing without any tissue.

3.3 DNA Extraction and Storage

For the DNA extractions from lung tissue the Qiagen DNeasy kit protocol (Maryland, USA) was followed with the following exceptions. First, the incubation at 55 ˚C lasted for 2 hours with intermittent vortexing, not necessarily every 10 minutes. Second, RNase treatment of samples was excluded. Finally, the final volume was 100 µL of AE buffer (Qiagen elution buffer) for all samples except those in the expanded sample set (see below) which had a final volume of 200 µL of AE buffer. All samples were originally stored in a -20 freezer. However, after September 10th, 2010 these extracted DNA samples were moved to long term storage in the -80 freezer and aliquots of either 1:5 (not bacteria or CF samples) or 1:10 dilutions were made and stored in the -20 freezer for short term usage.

3.4 Study Design

Characterization of the entire microbiome in the lung was first attempted by metagenomic sequencing (i.e., sequencing of the entire DNA present in the specimen, followed by the separation of microbial from human DNA and determining the composition of the possible microbiome present in the tissue). This preliminary study was completed in collaboration with Genome BC. Due to the high costs of this type of study it was limited to five samples of lung from one person with very severe COPD (GOLD 4) treated by lung transplant and two samples from a donor lung from a smoker without evidence of COPD. These seven DNA samples were sequenced on an Ilumina based platform. As these results suggested that the microbiome of the lung contained viral and fungal elements as well as bacteria, the costs of pursuing all of these

possibilities led to a decision to reduce the focus of the study and to concentrate on the bacterial component of the microbiome. This lead to the second part of the study which consisted of the same COPD (GOLD 4) individual and the same five samples from this individual and one cystic fibrosis patient of whom 6 samples of lung tissue were processed and DNA extracted. For this study we were interested to see if we could validate, using both PCR and qPCR techniques, what was discovered in the preliminary metagenomic sequencing study. The cystic fibrosis patient was used as a positive control since it has been well established that these individuals have a large number of bacteria within their lungs [157, 167]. Thus bacteria from these lung tissue samples should be detected by assays which utilize the universal 16S rRNA gene primers. After the second study was completed and the data analyzed, it was decided to expand our sample set to better address the second aim of our hypothesis. This third part of the study included an expanded sample set of eight non-smokers, eight smokers, eight COPD (GOLD 4), and eight cystic fibrosis patients. All groups were matched for sex as well as matched for the tissue sample location within the lung section. The tissue sample was taken from a middle region section of either a left or right lung. Additionally, the non-smoking, smoking, and COPD (GOLD 4) groups were all matched for exact age. We used qPCR to generate a relative total bacteria count in all the groups. We then utilized terminal restriction length polymorphism (T-RFLP) analysis to assess, on a generalized scale, the bacterial community composition contained within each individual and group. Finally, 454 targeted sequencing was used to identify the type of bacteria present in the samples and provide a link back to the community composition profiles and total bacterial numbers that were generated [Figure 3].

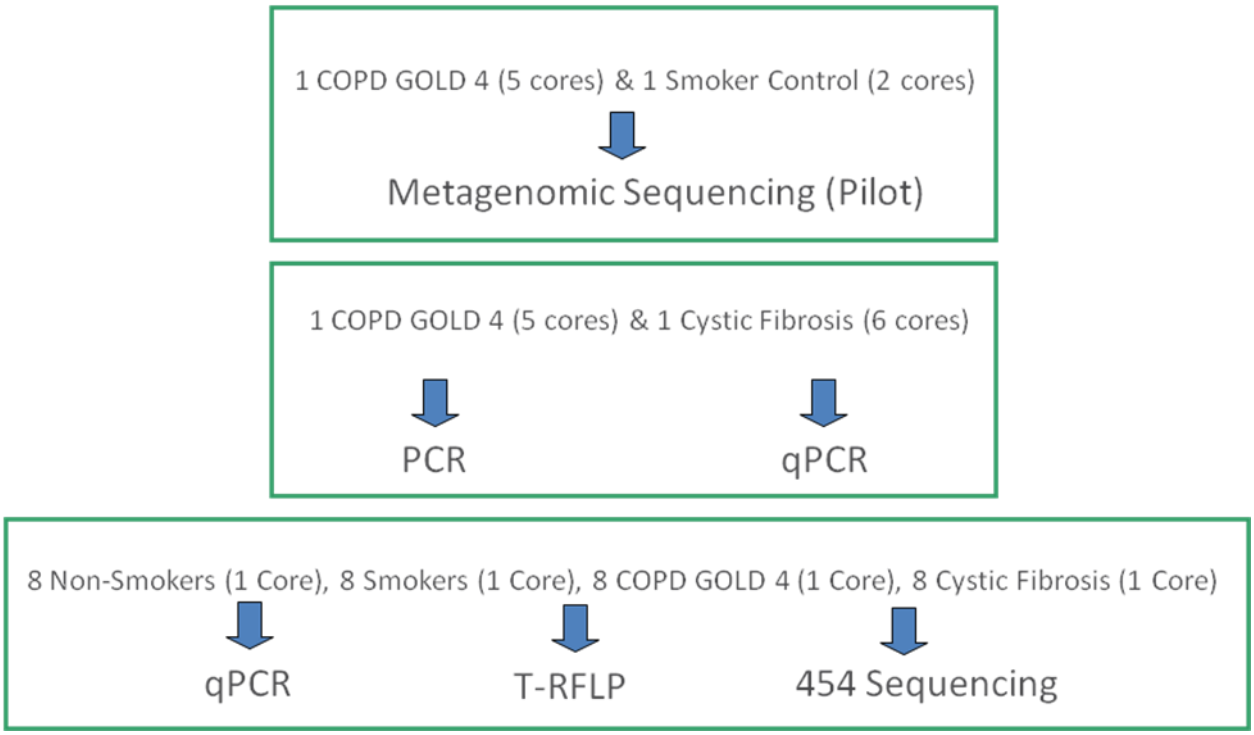

Figure 3: Diagram of the three different parts of the study.

3.4.1 16S rRNA Gene

This gene was chosen because it contains conserved regions in every bacterial genome plus variable regions that differ from genus to genus and family to family. The conserved regions can be used to design specific PCR primers that will target all 16S rRNA genes and allow quantification of total bacterial counts. However, correction factors need to be applied since many bacteria have genomes that contain different numbers of the 16S rRNA genes [168]. One method that was employed to get around this was the use of calculated total bacterial cells rather than copies of 16s rRNA genes in the qPCR assay per sample. Similarly, PCR primers can be designed to target different combinations of the nine different variable regions to determine the different community composition

profiles as well as to identify the bacteria within each DNA sample [169]. Both the T-RFLP analysis and 454 targeted sequencing utilized these variable regions in their analysis.

3.5 Experiment 1: Preliminary Ilumina Sequencing Project

3.5.1 COPD (GOLD 4) vs. Smoking Control

For the COPD (GOLD 4) individual the following samples were used: 6968-6AA, 9BB, 2BB, 8BB, and 11B. The smoking control individual had the following samples used: 6989-8B and 2C. For all these lung tissue samples approximately 10mg of tissue was used for DNA extraction [Table 2].

Table 2: Breakdown of tissue weight and DNA sent to Genome BC.
It should be noted that all these DNA samples are in a final volume of 100 µL of AE buffer.

Sample Group	Core ID	Tissue (mg)	DNA Concentration (ng/µL)
Smoker Control	6989-8B	10	96.1
Smoker Control	6989-2C	10	116.9
COPD GOLD 4	6968-6AA	10	87.7
COPD GOLD 4	6968-9BB	10	92.7
COPD GOLD 4	6968-2BB	10	94.7
COPD GOLD 4	6968-8BB	10	107.8
COPD GOLD 4	6968-11B	10	76.0

Samples were each given a unique number label once at Genome BC which could be traced back to that of the original sample. A 0.7% agarose gel of a representative sample (one sample and one of a λ DNA Hind III digestion) was submitted along with the genomic samples used for testing. The agarose gel was used to help assess the amount and quality of DNA that was being submitted.

3.5.2 Metagenomic Sequencing Platform

The Illumina Genome Analyzer IIx (California, USA) platform utilizes a form of sequencing known as massively parallel sequencing or shot-gun sequencing. DNA samples were taken and fragmented into approximately 75 base pair fragments. These fragments were then denatured and had a common adaptor ligated to them. These fragments were then added to the flow cell, where the actual sequencing reaction would take place (area that is coated with oligonucleotides that are exactly complimentary to the ligated adaptors on the fragments) [170-173]. A special PCR, known as bridge amplification, was performed to generate clusters for each individual fragment that was attached to the flow cell [171-173]. Next, the clusters were denatured and the sequencing primers, polymerase, and fluorescently labeled nucleotides (a different colour for each base) were added. The fluorescently labled nucleotides contain a chemically inactivated 3'OH thatn can be reversibly terminated [174]. This reversible termination allows for sequencing one nucleotide at a time to occur, since after each round the termination can be reversed for addition of the next nucleotide. After each round of a nucleotide addition, the surface of the slide is imaged and the colour of each cluster recorded [171-173].

3.5.3 Down-stream Analysis

Each sample was run on a separate flow cell and the DNA quality was assessed. First, all reads that were sequenced and that aligned to the human genome were subtracted out of the sequenced fragment pool. Second, those that aligned to *E.coli*, more specifically lab strains of *E.coli* were also subtracted out of the sequenced fragment pool. Finally, those that were left over were run through a BLAST search through a non-redundant database that was utilized by GenomeBC. From this BLAST search a list of up to the ten most likely bacterial species were given for each particular read that was identified as being from bacteria. This same BLAST search methodology also aligned a number of reads to fungi and virus. A final conversion was made on the information of the sequence reads from Genome BC so that the number of bacteria per 1000 human cells could be used. The conversion process consisted of the following steps. First the average total number of human reads (29.5×10^6) was corrected for genome size (6.4×10^9 base pairs). This gave 4.6×10^{-3} reads per human genome. Second, the average number of bacterial reads that were not *E.coli* was corrected for genome size using an average bacterial genome size of 4×10^6 base pairs as the reference. This gave a number of reads per bacterial genome. The number of reads per bacterial genome was divided by the number of reads per human genome to give a number that represents the number of bacteria per human cell. This number was then multiplied by 1000 to get the desired conversion of bacteria cells per 1000 human cells.

3.6 Experiment 2: Validation of Metagenomic Sequencing with PCR and qPCR

3.6.1 GenomeBC COPD (GOLD 4) vs. Cystic Fibrosis Samples

This group included the same five samples from the one COPD (GOLD 4) individual (6968-2BB, 6AA, 8BB, 9BB, 11B), which had been sequenced in the Ilumina project, compared to 6 samples from one cystic fibrosis individual (5928-2 to -7). The amount of tissue used and DNA concentration can be seen in [Table 3].

Table 3: Breakdown of the tissue weight and DNA used to verify metagenomic sequencing.
All DNA samples were in a final volume of 100 µL of AE buffer.

Sample Group	Core ID	Tissue (mg)	DNA Concentration (ng/µL)
COPD GOLD 4	6968-6AA	10	87.7
COPD GOLD 4	6968-9BB	10	92.7
COPD GOLD 4	6968-2BB	10	94.7
COPD GOLD 4	6968-8BB	10	107.8
COPD GOLD 4	6968-11B	10	76.0
Cystic Fibrosis	5928-2	20	30.0
Cystic Fibrosis	5928-3	20	44.0
Cystic Fibrosis	5928-4	20	10.0
Cystic Fibrosis	5928-5	20	13.0
Cystic Fibrosis	5928-6	20	52.0
Cystic Fibrosis	5928-7	20	142.0

3.6.2 PCR

3.6.2.1 Overview

Polymerase chain reaction (PCR) involves replication of a specified target in DNA. In our case we are using the 16S rRNA gene of bacteria as our target gene. We started with 6 different primers during the PCR troubleshooting phase. For the PCR assay reaction we used the following primer pairs 27F (5'-AGAGTTTGATCMTGGCTCAG-3') with 1390R (5'-GACGGGCGGTGTGTRCAA-3') which gave a product of 1364 base pairs [175, 176], 339F (5'-ACTCCTACGGGAGGCAGCAGT-3') with 907R (5'-CCGTCAATTCMTTTGAGTTT-3') which gave a product of 569 base pairs [158], and 63F (5'-GCAGGCCTAACACATGCAAGTC-3') with 355R (5'-CTGCTGCCTCCCGTAGGAGT-3') which gave a product of 293 base pairs [158]. The notation of a number followed by either F or R denotes the nucleotide in the gene sequence where the primer starts and whether it is the forward (F) or reverse (R) primer. The qPCR assay used the primer pair specifying the 293 base pair amplicon which is relevant to both the validation experiment and expanded sample set experiment. Furthermore a 100 base pair target of the human Rpp40 gene (forward primer: 5'-CGTAAGCAAGTTTAGTGAATACCTGAA-3' and the reverse primer: 5'-GCACAGCTTCCATCTTACTCAATC-3') was used as a control for all PCR reactions to insure that the samples contained human DNA. The Rpp40 subunit is part of the RNAse P gene family which is a single copy gene [177]. Thus bacterial numbers can be normalized to human cells since the Rpp40 gene is a single copy gene. The primers were chose and the specificity was checked by using the NCBI primer BLAST database. All primers were ordered from SIGMA Life Sciences (Ontario, Canada).

3.6.2.2 PCR Conditions

The following are the PCR cycling conditions that were used after the optimization and troubleshooting phase:

100 base pair product RPP40:

1 x (15 minutes at 95 ˚C)

40 x (15 seconds at 95 ˚C, 30 seconds at 60 ˚C, 30 seconds at 72 ˚C)

293 base pair product (63F and 355R):

1 x (15 minutes at 95 ˚C)

40 x (15 seconds at 95 ˚C s, 1 minute at 63 ˚C)

569 base pair product (339F and 907R):

1 x (15 minutes at 95 ˚C)

40 x (15 seconds at 95 ˚C, 30 seconds at 60 ˚C, 1 minute at 72 ˚C)

1363 base pair product (27F and 1390R):

1 x (15 minutes at 95 ˚C)

40 x (40 seconds at 94 ˚C, 30 seconds at 57 ˚C, 2 minutes at 72 ˚C)

1 x (5 minutes at 72 ˚C)

The Bio-Rad MyCycler Thermal Cycler was used for all sample runs after troubleshooting was completed. The final PCR reaction volume was 50 µL and the volume of each reagent in this total mixture was as follows: 5 µL of the Qiagen 10x PCR buffer, 1 µL of dNTP mix for a final concentration of 200 µM of each dNTP, 2 µL of the forward primer or reverse primer for a final concentration of 0.4 µM, 0.25 µL of

HotstarTaq DNA Polymerase (2.5units/reaction), 34.75 µL of RNase free water, and 5 µL of template DNA.

3.6.2.3 Experimental Sampling PCR Runs

The GenomeBC samples were run six separate times; three times with the 1363 base pair 16S rRNA gene product protocol, once with the 569 base pair 16S rRNA gene product protocol, and twice with the 293 base pair 16S rRNA gene product protocol. The cystic fiborsis samples were run five times in total once with the 1363 base pair 16S rRNA gene product protocol, once with the 569 base pair 16s rRNA gene product protocol, and twice with the 293 base pair 16S rRNA gene product protocol. The six JM109 *E.coli* serial dilution samples were run at the same time with the last run of the 293 base pair rRNA gene product for both the Genome BC sample and cystic fibrosis sample.

3.6.3 qPCR

3.6.3.1 qPCR Conditions

After a number of different trial runs it was decided that two separate plates would have to be run for each sample, one for the 16S rRNA 293 base pair product (primers sequence found in PCR section) and one for the RPP40 100 base pair product. The cycling conditions for the qPCR of the 293 bp target of 16S rRNA gene were the same as for the respective PCR described above (section 3.6.2.2). For the 100 bp target of the human RPP40 gene, they were the following:

1 x (15 minutes at 95 ˚C)

40 x (15 seconds at 95 ˚C, 1 minute at 60 ˚C)

In this case, RPP40 was used as an internal control and normalization factor to which the number of bacteria cells could be compared.

The qPCR was run on an ABI Prism 7900HT Sequencing Detection System (California, USA) using a 384 well plate. The total reaction volume was set at 10 µL and had the following volumes of reagents: 5 µL of 2x Quantitect SYBR green PCR master mix (Maryland, USA), 0.4 µL of the forward primer for a final concentration of 0.4 µM, 0.4 µL of the reverse primer for a final concentration of 0.4 µM, and 4.2 µL of template DNA. It should also be mentioned that every sample, for both the 16S rRNA and RPP40 assay, was run in triplicate including the serial dilutions for the standard curve

E.coli colonies were grown on LB medium plates and six medium sized colonies were pooled, DNA extracted, and DNA concentration measured using Nanodrop. Standard curves were created using a 6-fold serial dilution of this stock DNA solution of JM109 *E.coli* genomic DNA as well as human genomic DNA. An example of the standard curve for the 16S rRNA gene can be seen in [Figure 4] and for the RPP40 gene in [Figure 5]. For both sample dilutions the lowest serial dilution value had to be excluded since the curve flattened out and was not linear.

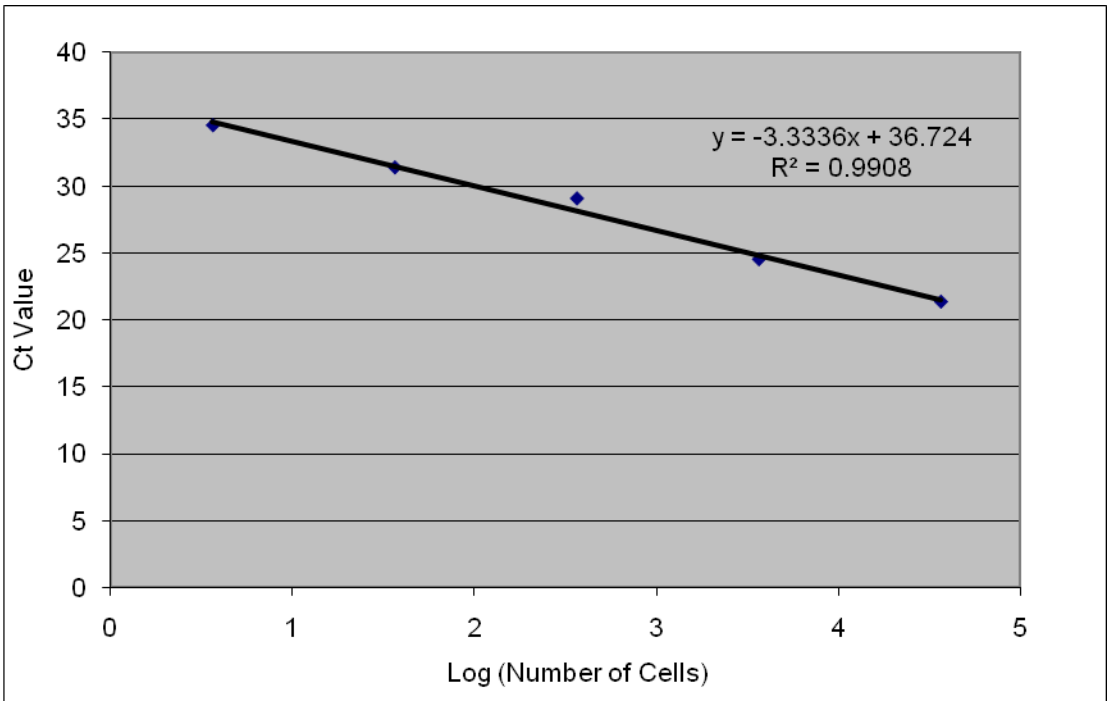

Figure 4: An example of the standard curve for the 16S rRNA gene qPCR assay. This assay was run under conditions specified in the methods section. The standard curve is derived from serial 10 fold dilutions of the *E.coli* DNA.

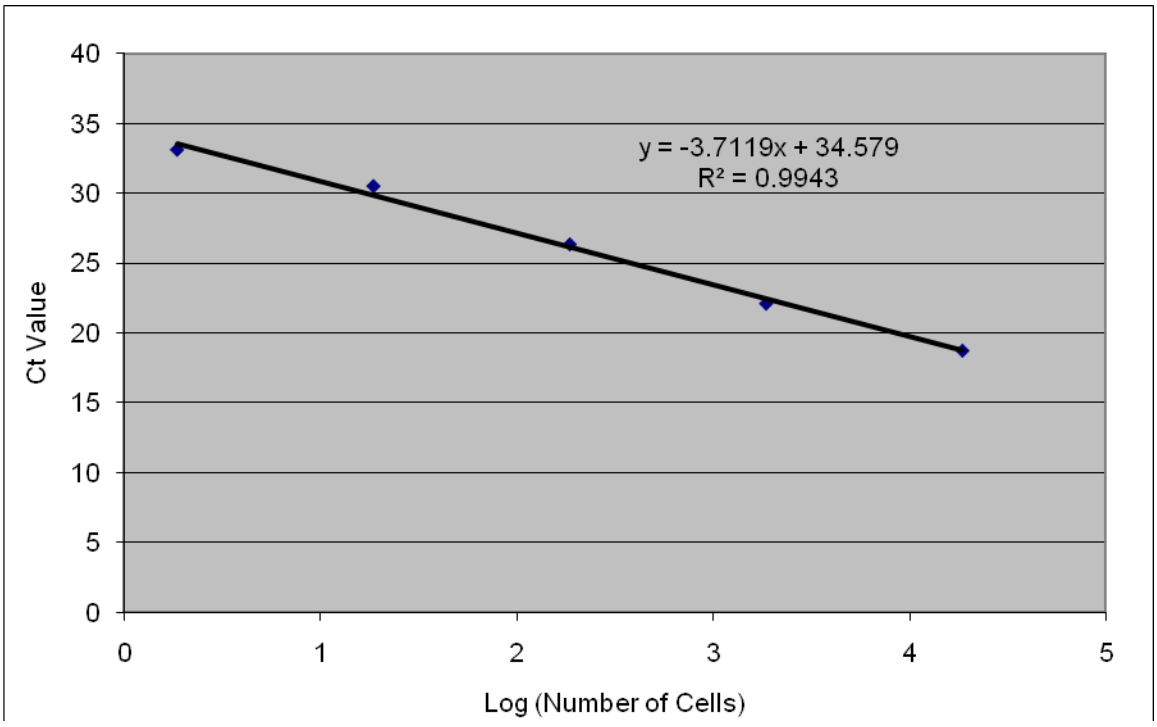

Figure 5: An example of the standard curve for the RPP40 gene qPCR assay.

This assay was run underconditions specified in the methods section. The standard curve is derived from serial 10 fold dilutions of the human DNA standard.

All runs included a dissociation curve run. The dissociation curve for the 16S product had a peak between 84 and 85 ˚C [Figure 6] while the dissociation curve for the RPP40 product had a peak at 78 ˚C [Figure 7]. After the first of the two qPCR runs the PCR products were run on a 1% agarose gel to verify that the product size corresponded to the dissociation curve peak for both the 16S and RPP40 product [Figure 8].

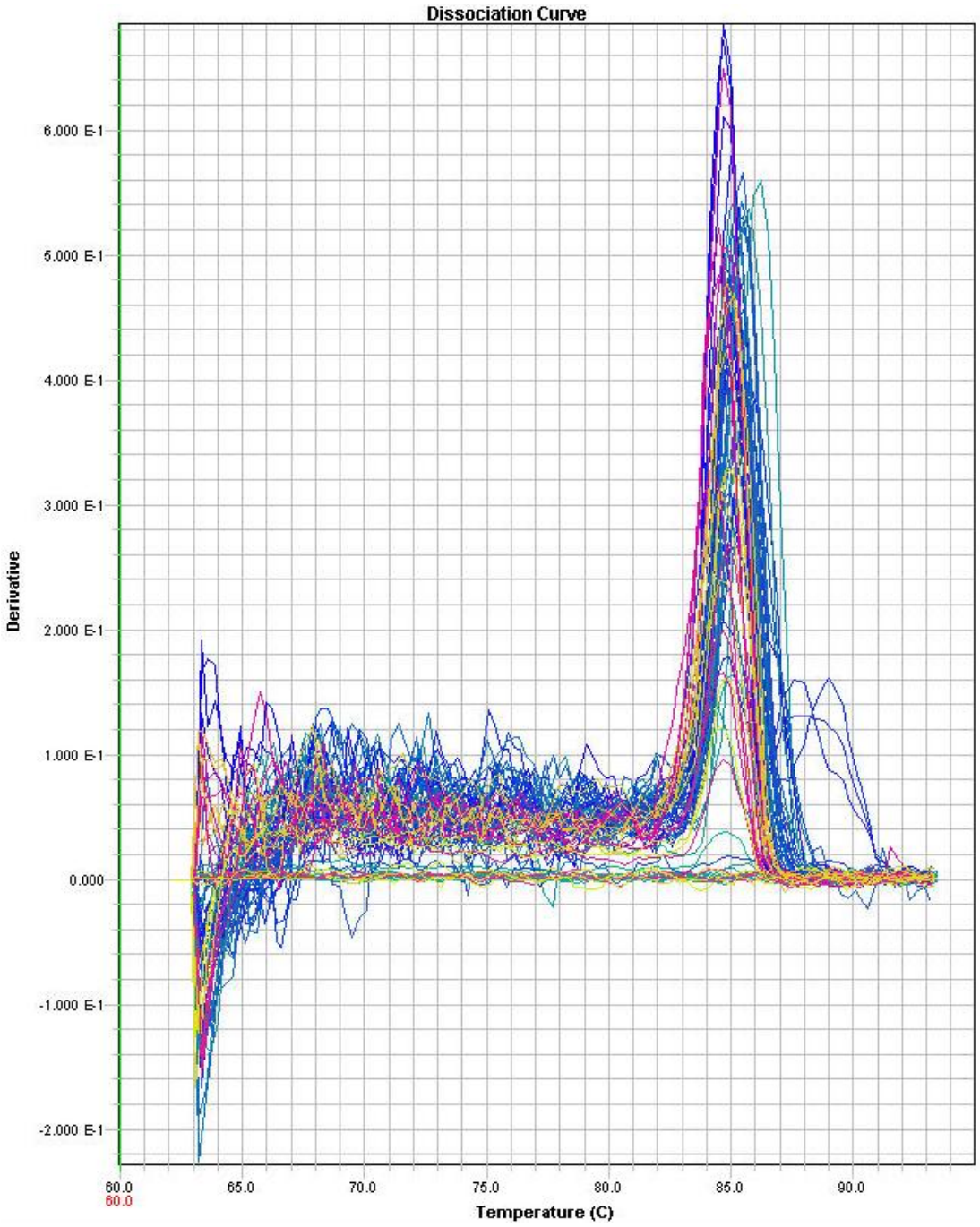

Figure 6: An example of the dissociation curve of the 16S rRNA qPCR product. There is one clear peak at approximately 84-85 ˚C.

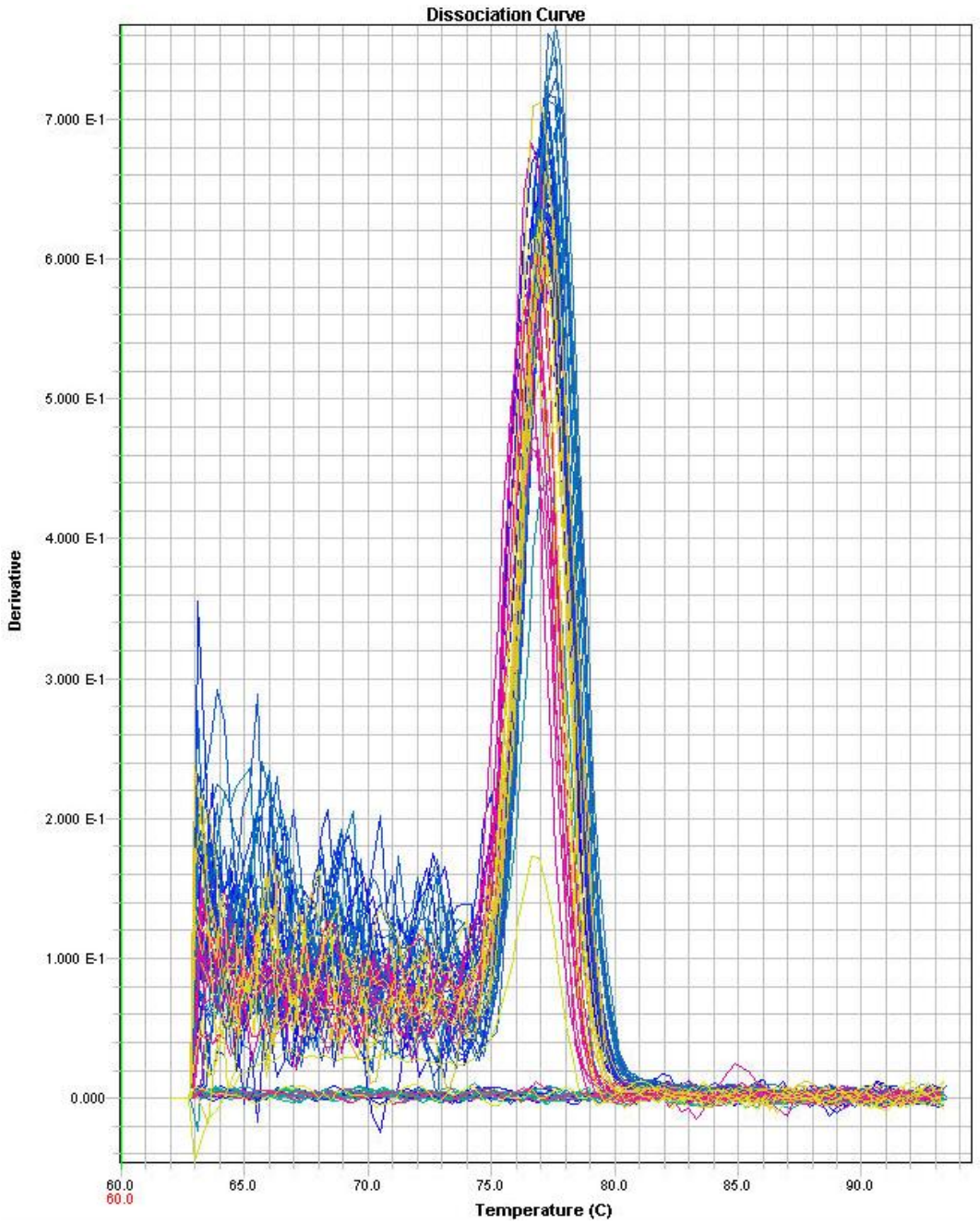

Figure 7: An example of the dissociation curve of the RPP40 qPCR product. There is one clear peak at approximately 78 ˚C.

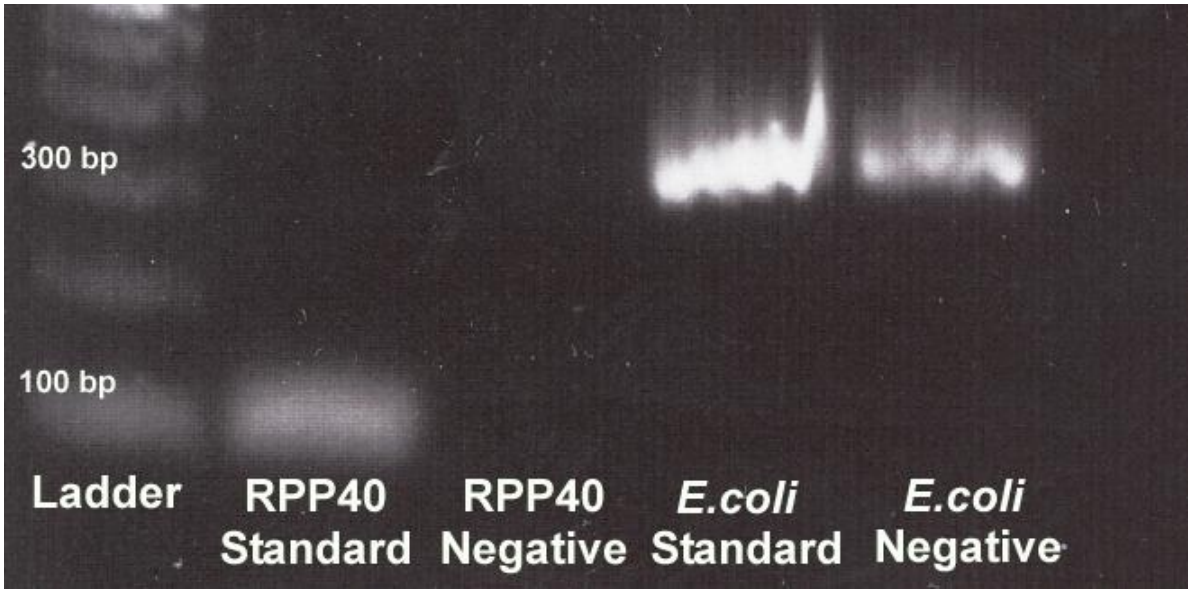

Figure 8: Verification of dissociation curve peaks after qPCR completion using a 1% agarose gel.

3.6.3.2 Cell Number Calculation for the Standard Curves

Each nucleotide is equal to 330 daltons. Therefore a single base pair is equal to 660 daltons. When converted to grams this works out to 1.0959×10^{-21} grams per base pair. The *E.coli* genome is 4.5 million base pairs. Thus 1.0959×10^{-21} grams per base pair multiplied by 4.5 million base pairs equals 4.932×10^{-15} grams of DNA per cell of *E.coli.* A similar conversion can be made for the human genome if 6 billion base pairs are used instead of 4.5 million base pairs. Thus after a similar amount of calculations it can be worked out that there is 6.58×10^{-12} grams of DNA per human cell. After obtaining a DNA concentration from the Nanodrop (Nanodrop, Delaware, USA) the ng/µL value was converted to a µg value based on the total volume. This µg value was then divided by grams of DNA per cell of *E.coli* and the resulting final number gave the number of cells per µL. Thus a serial dilution would result in 10 fold less cells each dilution. The same

process can then be applied for the number of human cells with the exception that the grams of DNA per human cell would be a different value.

3.6.3.3 The Correction Factor

All groups had two corrections performed on them. The first involved subtracting the average number of bacteria cells in the negative controls with no DNA added from the number of bacteria found in each sample (triplicate average). Once this was done this new value was divided by the number of human cells (determined using the RPP40 qPCR assay). Once these corrections were made all values represented bacteria cells per human cell. In order to work with whole numbers the value of bacteria cells per human cell was multiplied by a value of 1000 and the resulting value represented the number of bacteria cells per 1000 human cells.

3.7 Experiment 3: Expansion of Sample Size to Detect Bacterial Differences Between Sample Groups

3.7.1 Patient Information

This group included one lung sample from eight non-smokers (7180-2, 6788-2, 1977-10, 3037-6, 3263-7, 3480-7, 5909-10, and 6376-8), eight smokers (2014-9, 6894-15, 2431-9, 5771-2, 6651-9, 5882-7, 6043-3, and 6077-13), eight COPD (GOLD 4) (7013, 7014, 7015-31, 6968-30, 6971-26, 6967-5, 6969-30, and 6965-26), and eight cystic fibrosis individuals (2877-4, 5915-9, 5901-8, 5928-7, 5723-7, 5938-4, 6058-7, 5894-10). A single sample from a middle lung section slice was taken from all individuals and the

tissue sample was labeled according to tissue numbering method 2 (section 3.2 and Figure 1). The non-Smokers, smokers, and COPD (GOLD 4) groups were matched for age. Some samples in the different groups were reversed numbered so the smaller numbers are from tissue samples taken from a middle lung section slice. The COPD (GOLD 4) tissue samples from 7013 and 7014 were taken from a middle lung section slice that had not been sampled previously so do not have an associated number with the four digit identifier. Both the COPD (GOLD 4) and the cystic fibrosis groups include one individual each who was part of experiment 2 but a different lung sample from these individuals was used in this experiment. A table of the DNA concentration and amount of tissue used can be seen below [Table 4].

Table 4: The amount of tissue and DNA extracted in the expanded sample set.
All samples except the cystic fibrosis samples were in 100 µL of AE buffer. The cystic fibrosis samples were in 200 µL of AE buffer.

Sample Group	Core ID	Tissue (mg)	DNA Concentration (ng/µL)
Non-Smoker	7180	17	37.9
Non-Smoker	6788	28	46.9
Non-Smoker	1977	23	37.4
Non-Smoker	3037	44	134.3
Non-Smoker	3263	13	31.6
Non-Smoker	3480	29	75.0
Non-Smoker	5909	30	90.2
Non-Smoker	6376	11	34.2
Smoker	2014	18	122.3
Smoker	6894	25	71.6
Smoker	2431	12	12.6
Smoker	5771	11	18.2
Smoker	6651	28	74.4
Smoker	5882	21	93.6
Smoker	6043	17	36.6
Smoker	6077	27	33.6
COPD GOLD 4	7013	30	81.4
COPD GOLD 4	7014	40	122.2
COPD GOLD 4	7015	30	15.3
COPD GOLD 4	6968	30	74.2
COPD GOLD 4	6971	40	81.0
COPD GOLD 4	6967	40	80.0
COPD GOLD 4	6969	20	30.5
COPD GOLD 4	6965	30	116.7
Cystic Fibrosis	2877	30	157.2
Cystic Fibrosis	5915	40	151.8
Cystic Fibrosis	5901	30	127.4
Cystic Fibrosis	5928	30	172.7
Cystic Fibrosis	5723	40	200.6
Cystic Fibrosis	5938	30	221.3
Cystic Fibrosis	6058	30	203.8
Cystic Fibrosis	5894	30	333.3

3.7.2 qPCR

The standard curve generation and correction factor application were mentioned previously (Section 3.6.3). Thus they will not be mentioned any further in this section as the methodology was the same for the expanded sample set.

3.7.3 Terminal Restriction Fragment Length Polymorphism Analysis

3.7.3.1 Introduction

For the T-RFLP analysis a normal PCR is performed except that the forward primer is fluorescently labeled on the 5' end. The nucleotide sequence of the primer pair that was used is specified in section 3.8.3. Once the PCR is completed a restriction enzyme, *Hha*I, which cuts at a GCGC repeat, with the cut coming between the second GC was used to digest these products. These PCR fragments were then concentrated using a commercial kit (GE Health Care Ilustra GFX PCR DNA & Gel Band Purification Kit, Bukinghamshire, UK). The concentrated products were then run on a capillary gel electrophroesis machine. Since the products were digested only the products labeled at the 5' end would be visualized. The method itself takes advantage of the variation found between species to species and genus to genus in the sequence of the 16S rRNA gene. Output is shown with the y-axis representing fluorescence intensity while the x-axis represents fragment size. Thus, in general, the number of peaks observed at different sizes in each respective sample would give an indication of the total bacterial community composition, with the more diverse samples containing more total number of peaks. This

methodology is similar to that reported and used in other bacterial community composition studies [178].

3.7.3.2 T-RFLP Conditions

Samples from the PCR were run on a 1% agarose gel to check for product prior to T-RFLP analysis. The T-RFLP analysis was contracted out to the NAPS unit at the University of British Columbia.

The expanded sample set was run under the following conditions:

PCR Protocol-16S rRNA Gene, 881 Base Pair Product:

1 x (15 minutes at 95 ˚C)

40 x (40 seconds at 94 ˚C, 30 seconds at 57 ˚C, 1 minute 30 seconds at 72 ˚C)

Digestion Protocol-HHaI Restriction Enzyme:

Digestion conditions followed the protocol provided by New England BioLabs Inc. (Massachusetts, USA). However, the digestion time was modified to ensure complete digestion to the following:

4 hours at 37 ˚C

20 minutes at 60 ˚C

10 minutes at 4 ˚C

The PCR DNA purification kit utilized was the Illustra GFX PCR DNA and Gel Band Purification kit, from GE Healthcare, which can be commercially bought. No modifications were made to the provided protocol. It should be noted that only the expanded sample set, consisting of the eight Non-Smokers, eight Smokers, eight COPD (GOLD 4), and eight Cystic Fibrosis patients, had T-RFLP analysis performed on it.

3.7.3.3 The Correction Factor(s)

The intensities of all the different fragments in a sample were converted to a relative intensity versus the total intensity of the sample (relative intensity = (individual peak intensity) / (total intensity of all peaks)). After this correction the samples were analyzed and the positive (*E.coli* control) and negative (non-template control) peaks that corresponded to peaks in the samples were reduced to 0. The relative intensity was then recalculated and a new analysis using non-metric multidimensional scaling (NMS) ordination graphs were created to reflect this change. This removal was done to see if the positive and negative control products could influence the results that were observed.

3.7.4 GS-FLX 454 Targeted Sequencing

3.7.4.1 Background

An initial PCR reaction was performed with modified primers to the 16S rRNA gene. There was an approximately 40 base pair extension to the forward primer and an approximately 20 base pair extension to the reverse primer. Some of these extensions were used to attach the amplicon products to the bead system that the 454 sequencing machines at Genome Quebec (Montreal, Canada) used. The forward primer also contained a unique nucleotide sequence that allowed for identification of different samples. This allowed for a pooling of PCR products from multiple samples to be sequenced at the same time and then sorted after by computer. Pooling allows for a reduction in cost by splitting the number of total available reads, in a given reaction area for a sequencing run, between approximately 1-100 different samples. The more samples pooled results in less total reads per individual sample. The 454 sequencing machine is

similar in some respects to the Illumina sequencing platform. That is to say that an amplification reaction is used to generate clusters. These clusters can then be sequenced by the sequential addition of nucleotides and visualized using the generation of light, normally via fluorescence [171]. However, the lengths of sequencing reads obtained are much longer than that of the Illumina platform (550 bp length for the 454 sequencing versus 75 bp for the Illumina sequencing). This, together with the fact that the sequencing is targeted to the 16S rRNA gene, allowed for greater identification power with respect to bacteria. In order to identify the sample source of the sequence that will be generated the forward primer for every sample is different in that it has a sequence unique to that sample. However, the sequence downstream of this unique identifier is the same for all samples and is targeted to the 16S rRNA gene starting at nucleotide 27 of the gene. The reverse primer which is the same for every sample has the following nucleotide sequence: 5'-

CCTATCCCCTGTGTGCCTTGGCAGTCTCAGGWATTACCGCGGCKGCTG.

The blue denotes the special sequence that was added to the reverse primer to allow the PCR amplicons to bind the beads in the 454 sequencing reaction. The dirty gray sequence is complimentary to the 16S rRNA gene up to nucleotide 519 of the gene. The table below shows the exact sequences of the 40 forward primers used [Table 5]. Two to three of the best PCR reactions (based on 1 % agarose gel band intensity) for each sample were chosen and pooled to be sent to Genome Quebec.

Table 5: The different forward primers used to identify the sample source in 454 targeted sequencing.

Not shown are which primer sequence was used for which sample. The blue color marks the sequence used to attach the PCR amplicon product to the beads sequencing system of the 454 machine. The red color marks the unique identifier tag which would be used after sequencing to identify which amplicon sequence belonged to which case and group in the pooled sample sequencing reaction. The green colour marks the 27F primer sequence.

Genome Quebec Identifier	Sample Group	Forward Primer Sequence (5'-3')
MID-1	Negative	CCATCTCATCCCTGCGTGTCTCCGACTCAGACGAGTGCGTAGAGTTTGATCMTGGCTCAG
MID-2	Non-Smoker	CCATCTCATCCCTGCGTGTCTCCGACTCAGACGCTCGACAAGAGTTTGATCMTGGCTCAG
MID-3	Non-Smoker	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGACGCACTCAGAGTTTGATCMTGGCTCAG
MID-4	Non-Smoker	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCACTGTAGAGAGTTTGATCMTGGCTCAG
MID-5	Non-Smoker	CCATCTCATCCCTGCGTGTCTCCGACTCAGATCAGACACGAGAGTTTGATCMTGGCTCAG
MID-6	Non-Smoker	CCATCTCATCCCTGCGTGTCTCCGACTCAGATATCGCGAGAGAGTTTGATCMTGGCTCAG
MID-7	Non-Smoker	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGTGTCTCTAAGAGTTTGATCMTGGCTCAG
MID-8	Non-Smoker	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTCGCGTGTCAGAGTTTGATCMTGGCTCAG
MID-10	Non-Smoker	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTCTATGCGAGAGTTTGATCMTGGCTCAG
MID-11	Smoker	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGATACGTCTAGAGTTTGATCMTGGCTCAG
MID-13	Smoker	CCATCTCATCCCTGCGTGTCTCCGACTCAGCATAGTAGTGAGAGTTTGATCMTGGCTCAG
MID-14	Smoker	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGAGAGATACAGAGTTTGATCMTGGCTCAG
MID-15	Smoker	CCATCTCATCCCTGCGTGTCTCCGACTCAGATACGACGTAAGAGTTTGATCMTGGCTCAG
MID-16	Smoker	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCACGTACTAAGAGTTTGATCMTGGCTCAG
MID-17	Smoker	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGTCTAGTACAGAGTTTGATCMTGGCTCAG
MID-18	Smoker	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTACGTAGCAGAGTTTGATCMTGGCTCAG
MID-19	Smoker	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGTACTACTCAGAGTTTGATCMTGGCTCAG
MID-20	COPD GOLD 4	CCATCTCATCCCTGCGTGTCTCCGACTCAGACGACTACAGAGAGTTTGATCMTGGCTCAG
MID-21	COPD GOLD 4	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGTAGACTAGAGAGTTTGATCMTGGCTCAG
MID-22	COPD GOLD 4	CCATCTCATCCCTGCGTGTCTCCGACTCAGTACGAGTATGAGAGTTTGATCMTGGCTCAG
MID-23	COPD GOLD 4	CCATCTCATCCCTGCGTGTCTCCGACTCAGTACTCTCGTGAGAGTTTGATCMTGGCTCAG
MID-24	COPD GOLD 4	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAGAGACGAGAGAGTTTGATCMTGGCTCAG
MID-25	COPD GOLD 4	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGTCGCTCGAGAGTTTGATCMTGGCTCAG
MID-26	COPD GOLD 4	CCATCTCATCCCTGCGTGTCTCCGACTCAGACATACGCGTAGAGTTTGATCMTGGCTCAG
MID-27	COPD GOLD 4	CCATCTCATCCCTGCGTGTCTCCGACTCAGACGCGAGTATAGAGTTTGATCMTGGCTCAG
MID-28	Cystic Fibrosis	CCATCTCATCCCTGCGTGTCTCCGACTCAGACTACTATGTAGAGTTTGATCMTGGCTCAG
MID-29	Cystic Fibrosis	CCATCTCATCCCTGCGTGTCTCCGACTCAGACTGTACAGTAGAGTTTGATCMTGGCTCAG
MID-30	Cystic Fibrosis	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGACTATACTAGAGTTTGATCMTGGCTCAG
MID-31	Cystic Fibrosis	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCGTCGTCTAGAGTTTGATCMTGGCTCAG
MID-32	Cystic Fibrosis	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGTACGCTATAGAGTTTGATCMTGGCTCAG
MID-33	Cystic Fibrosis	CCATCTCATCCCTGCGTGTCTCCGACTCAGATAGAGTACTAGAGTTTGATCMTGGCTCAG
MID-34	Cystic Fibrosis	CCATCTCATCCCTGCGTGTCTCCGACTCAGCACGCTACGTAGAGTTTGATCMTGGCTCAG
MID-35	Cystic Fibrosis	CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGTAGACGTAGAGTTTGATCMTGGCTCAG
MID-36	Negative	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGACGTGACTAGAGTTTGATCMTGGCTCAG
MID-37	Negative	CCATCTCATCCCTGCGTGTCTCCGACTCAGTACACACACTAGAGTTTGATCMTGGCTCAG
MID-38	Negative	CCATCTCATCCCTGCGTGTCTCCGACTCAGTACACGTGATAGAGTTTGATCMTGGCTCAG
MID-39	Cystic Fibrosis Control	CCATCTCATCCCTGCGTGTCTCCGACTCAGTACAGATCGTAGAGTTTGATCMTGGCTCAG
MID-40	COPD GOLD 4 Control	CCATCTCATCCCTGCGTGTCTCCGACTCAGTACGCTGTCTAGAGTTTGATCMTGGCTCAG
MID-41	Smoker Control	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAGTGTAGATAGAGTTTGATCMTGGCTCAG
MID-42	Non-Smoker Control	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGATCACGTAGAGTTTGATCMTGGCTCAG

3.7.4.2 454 Targeted Sequencing Conditions

The following PCR cycling conditions were used:

PCR Protocol-16S rRNA Gene, 550 Base Pair Product:

1 x (15 minutes at 95 ˚C)

40 x (40 seconds at 94 ˚C, 40 seconds at 61 ˚C, 1 minute at 72 ˚C)

1 x (10 minutes at 72 ˚C)

PCR products were run on a 1% agarose gel under similar conditions of the 1% agarose gel in the PCR for Experiment 2. Using the Qiaquick Gel Extraction kit the bands of interest, from each sample, were extracted and purified for Genome Quebec. Analysis was done using the open source mothur platform to filter out, screen, and perform general quality control of the sequencing reads (www.mothur.org) [179]. Identification of the filtered 16S amplicon reads was then performed using the RDP platform (http://rdp.cme.msu.edu/) which is constantly updated and maintained with up to date bacterial phylotypes. It has been published that a 50% confidence cut-off is used to produce accurate taxonomic identifications [164, 180]. However, due to the large number of reads that was returned from the targeted sequencing (55484 ± 7911 reads) a higher confidence cut-off of 80% was used.

3.7.5 Statistics

For most of the results, a simple T-test was used to compare differences between groups. Bonferroni correction was used where appropriate. For the qPCR results that involved comparing the cystic fibrosis group to other groups a non-parametric test (Wilcoxon/Kruskal Wallis Test) had to be used since the cystic fibrosis data was not

normally distributed. For the T-RFLP analysis a NMS graph was created to group the samples based on dissimilarity. It is an ordination method that takes into the account the number of peaks and the size of the fragments to group similar samples. Thus, the closer the samples or points are to each other the more similar they are. Finally, a Multi Response Permutation Procedure (MRPP), which is a type of multivariate analysis that involves pairwise comparisons, was run to compare whether the groups were significantly different from each other. The methodology, in particular the NMS and MRPP analysis, to analyze the T-RFLP data was similar to those previously used by others [181].

3.8 Troubleshooting

3.8.1 PCR

The first experiment run was on the GL Applied Research GTC-2 Genetic Thermal cycler machine followed by gel electrophoresis with a 1% agarose gel for visualization of the PCR products (0.5 grams of agarose in 50 mL of 1x TAE buffer solution and 2 µL of Gel Red). The primers first utilized were the 63F and 355R 16S rRNA gene product primers along with the 100 base pair product for the RPP40 gene. Oil was added to the reaction to avoid evaporation. The cycling conditions consisted of 15 minutes at 95 degrees followed by 40 cycles of 94 degrees for 15 seconds, 60 degrees for 30 seconds, and 72 degrees for 30 seconds (the final cycling conditions had this step omitted (section 3.6.2.2)) for the first attempt at the 16S rRNA gene PCR. A positive control was not used and a negative non-template control (water) was used. This experiment was

repeated under various different conditions to verify the results that were obtained. First, a 70% ethanol wipe down step of the bench area was introduced. Second, the number of negative controls was increased. Third, oil was not used. Fourth, the preparation was performed in a different location. Fifth, bleach was used as a wipe down agent along with 70% ethanol in two separate steps. Sixth, PCR of different water sources along with an RNaseZap (Ambion, Ontario, Canada) wipe down step was introduced. After these experiments a positive control was introduced. Our positive control consisted of JM109 *E.coli* that was grown and then extracted using the same DNA extraction conditions as stated previously. However, there was one exception that instead of tissue weight, six colonies of medium size were used for DNA extraction. A six fold serial dilution of the stock JM109 *E.*coli was created to help assess the sensitivity and specificity of the PCR assay and the cycling conditions used in section 3.6.2.2 were used. For these PCR reactions the Bio-Rad My Cycler Thermal Cycler and the PCR Express ThermoHybaid were utilized.

3.8.2 qPCR

The first run assessed the efficiency of the reaction for both the human RPP40 100 base pair product and the 16S rRNA gene 293 base pair product.

3.8.3 T-RFLP

Through collaboration with Dr. Bill Mohn (Life Sciences Centre, UBC) an aliquot of all samples were brought to the Mohn laboratory to work out problems related to the technology of T-RFLP. The results of the collaborations led to the conclusion that the

27F (5'-AGAGTTTGATCMTGGCTCAG-3') and 907R (5'-CCGTCAATTCMTTTGAGTTT-3') primers for the 16S rRNA gene, that gave an 881 base pair product, were the most suitable for our purpose. Using the normal protocol adapted in the Mohn laboratory the restriction enzyme *HHaI* was used (cut site at a GCGC repeat with cleavage between the second G and C). A PCR with a temperature gradient was run under 40 cycles to determine the optimal annealing temperature of these two primers in a PCR reaction together. It was decided that 57 ˚C would be the ideal temperature based on the results of the PCR temperature gradient experiment.

3.8.4 454 Targeted Sequencing

The first few rounds of PCR were with an *E.coli* positive control to determine an annealing temperature to use with these primers. One forward primer was chosen and the annealing temperature that was found was assumed to be the same for all other forward and universal reverse primer pairs. In order to attempt to get rid of the secondary products in the reactions both gel extraction and a modified PCR product purification protocol was used. The protocol for the modified PCR product purification was adapted from the Machery-Nagel Nucleo Spin Extract II kit (California, USA) protocol handbook. Various dilutions of the Buffer NT were made and the dilution that cleared the secondary product best was used (1:6 dilution with DNase and RNase free water). Gel extraction of the bands of interest and usage of the Qiaquick Gel Extraction kit (Maryland, USA) was ultimately chosen after Genome Quebec indicated that this modified PCR product purification protocol was unable to separate the secondary PCR products.

Chapter 4: Results

4.1 Experiment 1: Preliminary Ilumina Sequencing Project

The majority of the reads in the seven samples were classified as high quality (Q_{30} rating or above) meaning the probability of a wrong base call for the reads was 1 in 1000 or that there was a 99.99% accuracy that the base was assigned correctly. The total read range of 25-35 million reads with an average of 31.3±2.6 (SD) million reads per sample. Approximately 200,000 reads matched to other known databases with the large majority mapping to primate DNA. A further 1.3 million reads did not align to a known sequence in the non redundant database. The total bacterial reads in each COPD (GOLD 4) sample was as follows: 6968-2BB had 78 total bacterial reads, 6AA had 38 total bacterial reads, 8BB had 61 total bacterial reads, 9BB had 38 total bacterial reads, and 11B had 44 total bacterial reads [Figure 9]. The total reads in the two smoking control samples was as follows: 6989-2C had 18 total bacterial reads while 6989-8B had 44 total bacterial reads. The number of bacterial reads was graphed versus the relative lung height from where the sample was taken and also was separated based on whether the sample belonged to the COPD (GOLD 4) or the smoking control individual [Figure 9].

Sample ID	Total Bacterial Reads
6968-2BB	78
6968-6AA	38
6968-8BB	61
6968-9BB	38
6968-11B	44
6989-2C	18
6989-8B	44

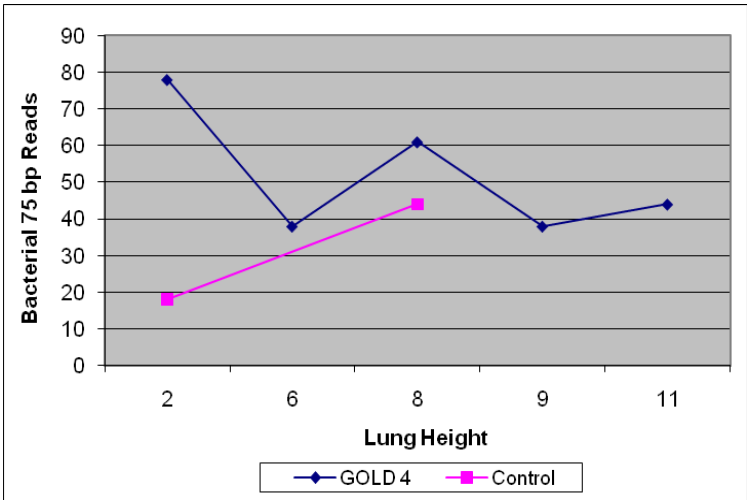

Figure 9: Total bacterial 75 nucleotide reads per sample and versus lung height. The table represents the total bacterial reads found in each core while the graph represents the bacterial reads with respect to lung height.

The average number of bacterial reads for the COPD (GOLD 4) individual from the five samples was 51.8. Upon correction with the conversion factor the COPD (GOLD 4) had 2.8 bacteria cells per 1000 human cells. In contrast the Smoking control individual from the two samples was 31 and upon correction gave 1.7 bacteria cells per 1000 human cells. The average total number of reads based for the COPD (GOLD 4) individual was not significantly different (P=0.32) then the Smoking control individual. Further the bacteria cells per 1000 human cells of the five samples from the COPD (GOLD 4) versus the two samples from the smoking control was not significantly different from each other (p=0.3) [Table 6].

Table 6: Metagenomic sample breakdown of the bacterial cells per 1000 human cells.

	Bacterial Cells/1000 Human Cells	
	COPD GOLD 4	Smoking Control
	4.2	1.1
	2.1	2.4
	3.3	
	2.1	
	2.4	
Average ± SD:	2.8 ± 0.9	1.8 ± 0.9

When the COPD (GOLD 4) samples and the Smoking normal controls were combined the average bacterial reads was 45.8 and 2.4 bacteria cells per 1000 human cells after conversion. When these corrected values of the number of bacteria reads per 1000 human genome were compared with Lm (a measure of the degree of emphysema, measured by John McDonough) of a corresponding sample for each sample analyzed there was a positive correlation with an R value of 0.74 ($p = 0.055$) [Figure 10].

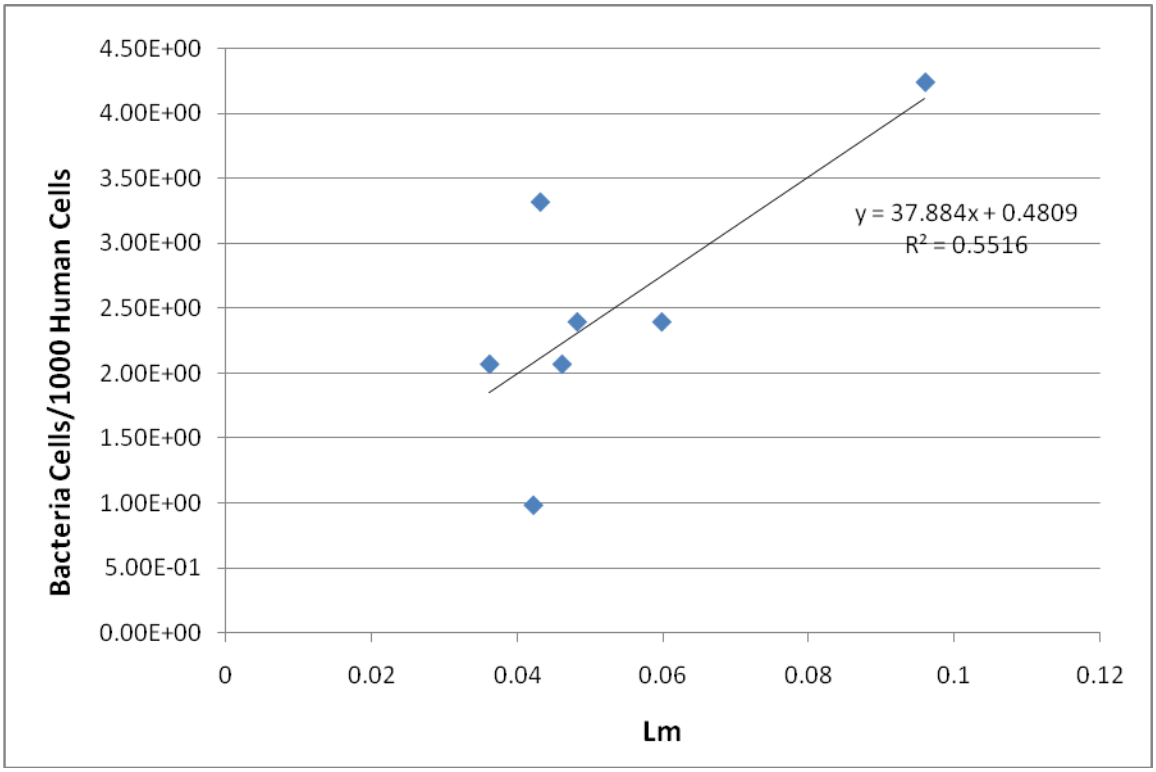

Figure 10: Graph of the bacteria cells per 1000 human cells versus the Lm.
The R^2 value was 0.55 and the R value was 0.74. The P value was equal to 0.055. There is a trend that with higher Lm there are more bacteria cells per 1000 human cells

When the individual reads were put through a BLAST search multiple species, genera, and phyla were identified for each 75 nucleotide read. Therefore definitive identification beyond assigning these reads to bacteria was not possible.

4.2 Experiment 2: Validation of Metagenomic Sequencing with PCR and qPCR

For this sample group both PCR analysis and qPCR analysis was performed. The COPD (GOLD 4) (6968-2BB, 6AA, 8BB, 9BB, 11B) PCR analysis yielded a faint band for all

the samples while the cystic fibrosis (5928-2, 3, 4, 5, 6, 7) yielded a variation in band intensity [Figure 11]. The band intensity of the cystic fibrosis individual was highest in the samples that corresponded to a lower lung height (closer to the bottom of the lung). There was a secondary product observed in all reactions that corresponded to approximately 120 base pairs in length.

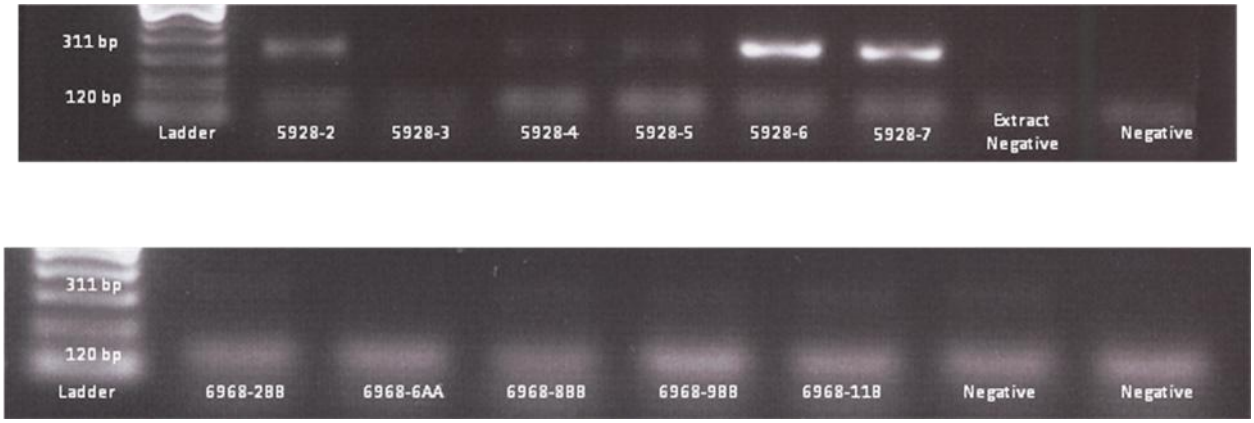

Figure 11: 1% Agarose gel of the cystic fibrosis versus the COPD (GOLD 4) individual.
The cystic fibrosis individual (6 samples) is shown above and the COPD (GOLD 4) individual (5 samples) below. The target amplicon was approximately 293 base pairs in length. There was a secondary product that is approximately 120 base pairs in length. The extract negative control was a DNase, RNase free water solution put through the DNA extraction process.

A qPCR analysis was performed and the results were similar. The standard curve for the bacterial 16S rRNA gene assay was $y = -4.2144x + 36.807$ with an $R^2 = 0.99$. The average efficiency of the 16S rRNA gene assay was 88.5±11.5 percent. The standard curve for the human RPP40 gene assay was $y = -3.6259x + 34.514$ with an $R^2 = 0.99$. The average efficiency of the RPP40 gene assay was 87.1±2.7 percent. The COPD (GOLD 4) average total number of bacteria was found to be 37 cells, the negative, non-

template controls was found to be 48 cells, and the Cystic Fibrosis was found to be 5049 cells. Analysis showed that there was no difference between the average total number of bacteria in the COPD (GOLD 4) group and the negative, non-template controls. However, the cystic fibrosis group was significantly different then both the COPD (GOLD 4) group and the negative, non-template control group ($p<0.0001$). A non-parametric test was used since the cystic fibrosis data were not normally distributed [Figure 12].

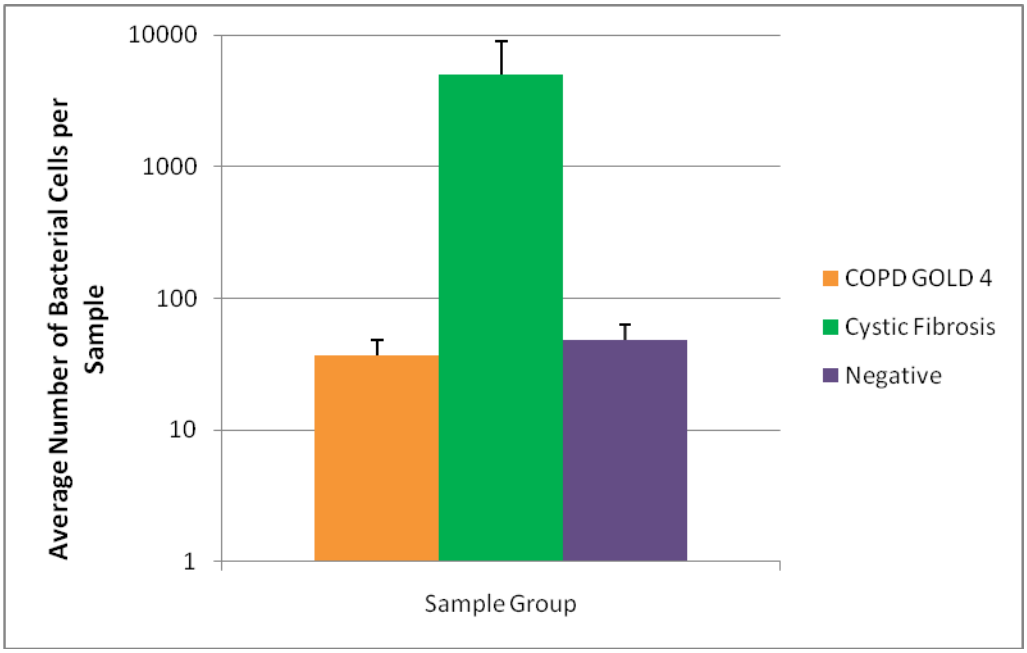

Figure 12: Graph of the average total number of bacteria measured by qPCR. The correction factor has not been applied. The five samples from the COPD GOLD 4 (36.8±11.8 (SE)) were not significantly different that than negative controls (48.4±14.5 (SE)). However, the 6 samples from the cystic fibrosis individual (5049±3843 (SE)) was significantly different then both the COPD (GOLD 4) group and the negative, non-template control group ($P<0.0001$) using nonparametric analysis since the Cystic Fibrosis data was not normally distributed.

When the average bacterial cells in the individual samples from the COPD (GOLD 4) and the Cystic Fibrosis individual were examined it was found that the COPD (GOLD 4) samples average bacterial cells remained relatively constant while that of the Cystic

Fibrosis individual fluctuated [Table 7]. This is before any corrections were applied to the data set.

Table 7: Difference in the bacterial loads of cystic fibrosis versus COPD (GOLD 4) patients

Sample Group	Core ID	Average Bacterial Cells	Standard Error
COPD (GOLD 4)	6968-2BB	83.34	51.45
COPD (GOLD 4)	6968-6AA	21.00	3.56
COPD (GOLD 4)	6968-8BB	33.43	22.03
COPD (GOLD 4)	6968-9BB	24.82	3.20
COPD (GOLD 4)	6968-11B	21.55	4.40
Cystic Fibrosis	5928-2	429.66	29.10
Cystic Fibrosis	5928-3	2.37E+03	3.64E+02
Cystic Fibrosis	5928-4	133.88	16.19
Cystic Fibrosis	5928-5	152.48	11.03
Cystic Fibrosis	5928-6	2.41E+04	1.51E+03
Cystic Fibrosis	5928-7	3.12E+03	1.55E+03
N/A	Negative	21.64	3.83
N/A	Negative	82.02	51.38
N/A	Negative	18.22	3.01
N/A	Negative	15.63	2.11

After the correction factor was applied only one COPD (GOLD 4) sample was above a value of zero and that corrected value was 3.26 bacteria cells per 1000 human cells. In contrast all the cystic fibrosis samples after correction were above zero [Table 8].

Table 8: qPCR results from cystic fibrosis and COPD (GOLD 4) patients after correction.

Sample Group	Core ID	Bacteria Cells/1000 Human Cells
COPD (GOLD 4)	6968-2BB	3.26
COPD (GOLD 4)	6968-6AA	-2.88
COPD (GOLD 4)	6968-8BB	-1.00
COPD (GOLD 4)	6968-9BB	-2.93
COPD (GOLD 4)	6968-11B	-1.07
Cystic Fibrosis	5928-2	102.09

Sample Group	Core ID	Bacteria Cells/1000 Human Cells
Cystic Fibrosis	5928-3	317.48
Cystic Fibrosis	5928-4	71.91
Cystic Fibrosis	5928-5	57.58
Cystic Fibrosis	5928-6	4836.56
Cystic Fibrosis	5928-7	1883.46

After correction the cystic fibrosis samples showed that there was an increase in the number of bacteria in the samples closer to the bottom of the lung [Table 8], which was similar to what was observed from the PCR analysis on the 1% agarose gel [Figure 11 & 13].

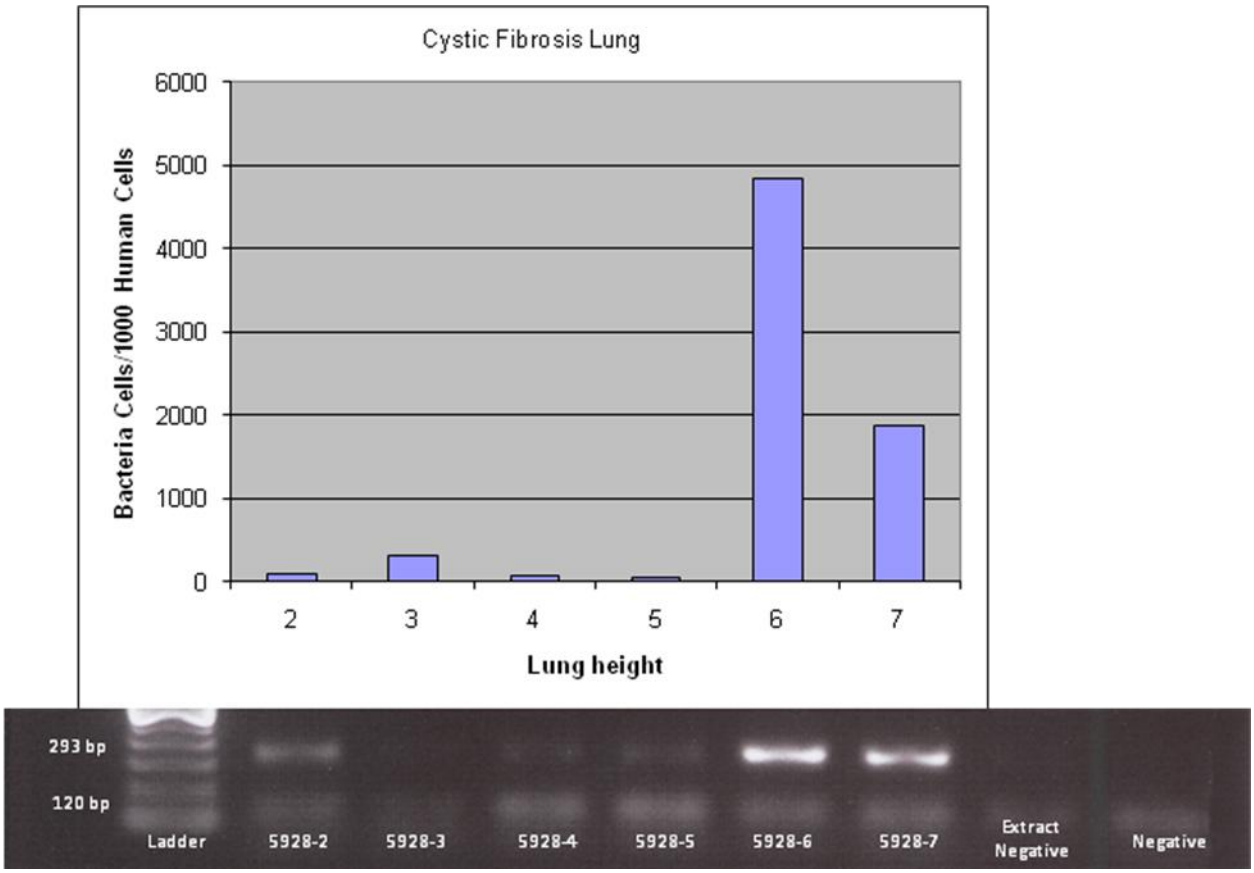

Figure 13: Cystic fibrosis PCR 1% agarose gel and qPCR SYBR green analysis.
A comparison of the results from the two types of analysis on the 6 different samples from one individual with cystic fibrosis. The top graph shows the number of bacterial cells per 1000 human cells determined by qPCR of the lung samples taken from different lung heights where the numbering starts at the top of the lung (method 2 from [Figure 1]).

The bottom represents a stained (gel red) 1% agarose gel of the products of conventional PCR from the same samples.

4.3 Experiment 3: Expansion of Sample Size to Detect Bacterial Differences Between Sample Groups

4.3.1 Patient Information

The clinical characteristics of the different patient groups are shown in the following table [Table 9]. The clinical data such as lung function for the cystic fibrosis individuals were either missing or not known. Further, at the time of this writing, lung function data for two individuals in the COPD (GOLD 4) group have not been sent to our research centre.

Table 9: Clinical breakdown of the four groups in the expanded sample set.
There were 8 patients/group and the star denotes a value that was significantly different then all others groups.

	Non-Smoker	Smoker	COPD GOLD 4	Cystic Fibrosis	P-Value
Age	56.25	56.875	58.75	32.63*	<0.015
Gender	5F/3M	5F/3M	3F/5M	3F/4M/1Unknown	———
FEV1	2.55	2.83	0.50*	N/A	<0.0001
FEV1pp	88.78	94.30	15.43*	N/A	<1x10-6
FVC	3.17	3.70	1.97*	N/A	<0.03
FVCpp	87.17	98.47	47.67*	N/A	<0.0001

4.3.2 qPCR

The standard curve for the 16S rRNA gene assay used to quantify bacteria in these groups of individuals was $y = -3.1638x + 36.373$ with a R^2 of 0.99. The corresponding

standard curve for the human Rpp40 gene assay was $y = -3.7119x + 34.579$ with an R^2 of 0.99. The average bacteria cells for each group, without the conversion factor, was 398 ± 239, 271±121, 353±547, 2363501±6052259, and 61±68 for the non-smoking, smoking, COPD (GOLD 4), cystic fibrosis, and negative (Non-Template control) group respectively. With the conversion factor the average number of bacteria cells per 1000 human cells was 35±22, 44±47, 24±37, and 205312±536960 for the Non-Smoking, Smoking, COPD (GOLD 4), and cystic fibrosis groups respectively. The data were not normally distributed and a non parametric (Wilcoxon-Kruskal Wallis Tests) analysis was used to compare these groups. The non-smoking ($P<0.0001$), smoking ($P<0.0001$), COPD (GOLD 4) ($P<0.002$), and cystic fibrosis ($P<0.0001$) groups were all significantly higher than the negative, non-template control group. The non-smoking, smoking, and COPD (GOLD 4) groups were not significantly different from each other ($P>0.05$) while the cystic fibrosis group was significantly different from these three groups ($P<0.0001$) [Figure 14]. The significance of the relationships stays the same with the addition of Bonferroni correction.

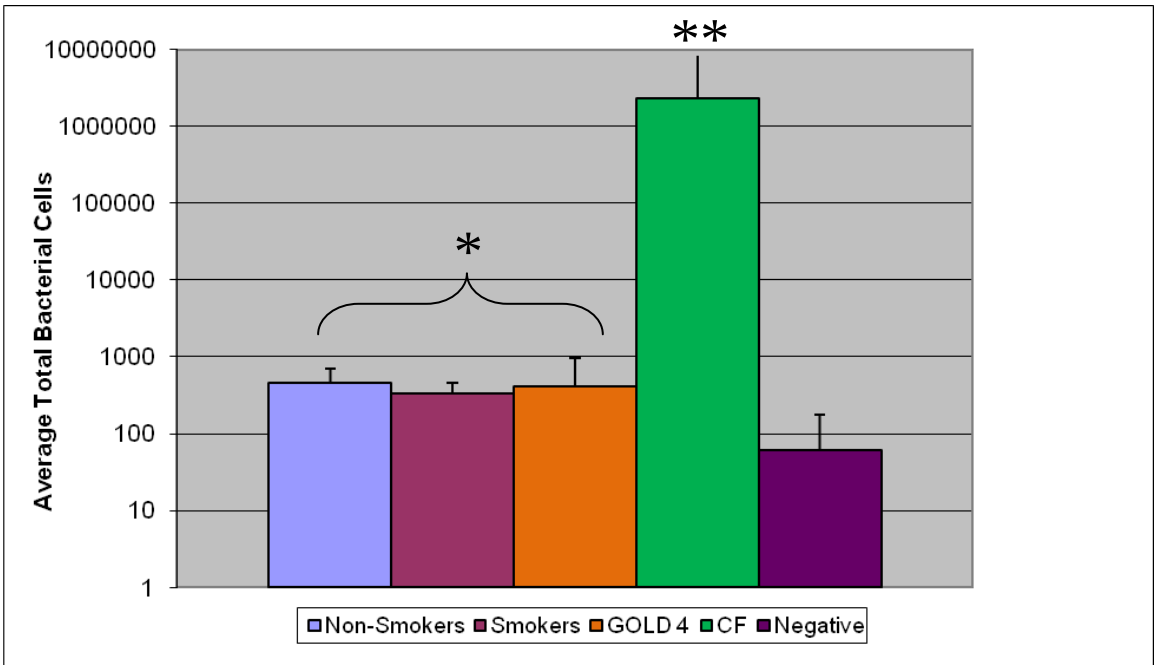

Figure 14: Average number of bacteria in the four groups of the expanded sample set.
The correction factors for normalization to 1000 human cells and subtraction of the average of the negative controls have not been applied. All the groups are significantly greater than the negative, non-template control group (*, $P<0.002$) shown in dark purple. The cystic fibrosis group was significantly different then all groups (**, $P<0.0001$) while the non-smoking group, smoking group, and COPD (GOLD 4) group were not significantly different from each other ($P>0.05$).

A breakdown of the individual samples in each respective group can help show that there was a great deal of variation both between groups and within groups themselves. Further there was a large variation between individual sample measurements using this method [Table 10].

Table 10: The breakdown of average total bacteria and variance within the individual samples (N=3) of each group.

Sample Group	Core ID	Number of Cells	Standard Error
Non-smoker Control	7180	293.34	167.81
Non-smoker Control	6788	808.51	436.19

Sample Group	Core ID	Number of Cells	Standard Error
Non-smoker Control	1977	307.66	126.47
Non-smoker Control	3037	5.20E+02	3.02E+04
Non-smoker Control	3263	399.27	164.79
Non-smoker Control	3480	818.66	292.69
Non-smoker Control	5909	339.97	160.06
Non-smoker Control	6376	180.13	152.61
Smoker Control	2014	549.09	410.26
Smoker Control	6894	152.02	81.65
Smoker Control	2431	440.24	78.72
Smoker Control	5771	308.21	34.32
Smoker Control	6651	263.06	42.94
Smoker Control	5882	260.07	71.49
Smoker Control	6043	358.74	212.29
Smoker Control	6077	322.67	157.04
COPD (GOLD 4)	7013	216.54	43.89
COPD (GOLD 4)	7014	184.51	165.25
COPD (GOLD 4)	7015	76.90	28.14
COPD (GOLD 4)	6968	137.14	65.94
COPD (GOLD 4)	6971	1.33E+03	5.95E+03
COPD (GOLD 4)	6967	23.78	12.09
COPD (GOLD 4)	6969	84.47	13.07
COPD (GOLD 4)	6965	1.25E+03	5.75E+02
Cystic Fibrosis	2877	1.73E+07	8.04E+06
Cystic Fibrosis	5915	2.10E+05	8.44E+04
Cystic Fibrosis	5901	3.98E+04	3.19E+04
Cystic Fibrosis	5928	2.71E+03	8.89E+02
Cystic Fibrosis	5723	8.54E+04	3.86E+03
Cystic Fibrosis	5938	1.71E+05	5.57E+04
Cystic Fibrosis	6058	1.08E+06	1.18E+05
Cystic Fibrosis	5894	7.54E+03	2.43E+02
N/A	Negative 1	48.91	17.03
N/A	Negative 2	60.87	55.59
N/A	Negative 3	21.75	11.35
N/A	Negative 4	81.71	54.37
N/A	Negative 5	0.91	0.55
N/A	Negative 6	40.74	40.57
N/A	Negative 7	1.60	1.41

Upon addition of the correction factor (normalization to 1000 human cells and subtraction of the average of the negative controls) as mentioned in the methods section the results remained the same with the cystic fibrosis group being significantly greater than the COPD (GOLD 4), non-smoking, and smoking groups. The cystic fibrosis group was significantly higher than the other three groups ($P<0.0001$) but the non-smoking, smoking, and COPD (GOLD 4) groups were not significantly different from each other ($P>0.05$) [Figure 15]. The significance of the relationships stays the same with the addition of Bonferroni correction.

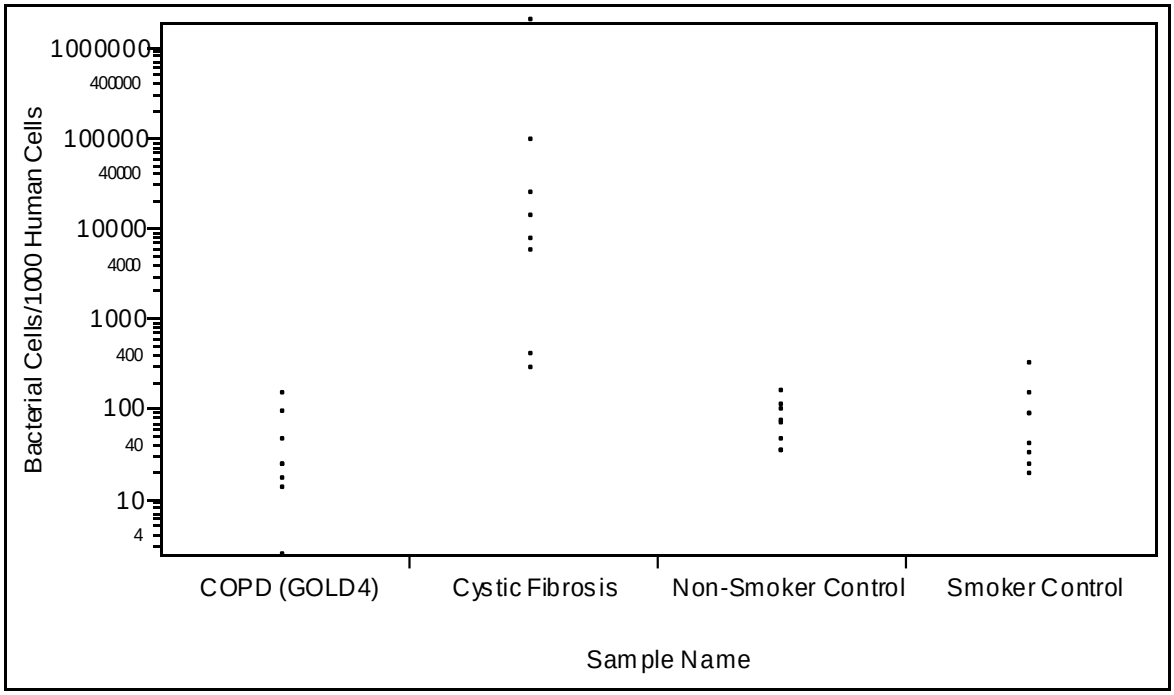

Figure 15: Total bacteria after correction in the four groups of the expanded sample set.
The bacterial cells per 1000 human cells were corrected for the negative controls and RPP40 human reference gene. The cystic fibrosis group contained more bacteria per 1000 human cells than all the other groups ($P<0.0002$). Although there was a large variance within the cystic fibrosis group there was no overlap with the results from the other groups. There was no significant difference between non-smokers, smokers, and COPD (GOLD 4) ($P>0.05$).

The individual breakdown of each sample can be seen in the table below. It can easily be seen that there is a large range in the cystic fibrosis samples despite controlling for location sampled. In contrast, there is not as much sample to sample change in the other three groups (Non-Smokers, Smokers, and COPD (GOLD 4)) [Table 11].

Table 11: Breakdown of the total bacteria in each sample after correction for negative controls and normalization to human cells.
The samples are organized from lowest bacteria cell per 1000 human cells to highest.

Sample Group	Core ID	Bacteria Cells/1000 Human Cells
COPD (GOLD 4)	6967	-2.46
COPD (GOLD 4)	6969	3.42
COPD (GOLD 4)	7015	4.24
COPD (GOLD 4)	6968	4.99
Smoker Control	6894	6.13
COPD (GOLD 4)	7014	7.10
COPD (GOLD 4)	7013	11.15
Smoker Control	5882	12.97
Non-smoker Control	3037	15.45
Non-smoker Control	5909	17.34
Non-smoker Control	6376	17.66
Smoker Control	2014	19.14
Smoker Control	6651	19.48
Non-smoker Control	1977	27.47
Non-smoker Control	7180	28.47
Smoker Control	6043	34.63
Non-smoker Control	3480	43.99
Smoker Control	6077	44.38
Non-smoker Control	3263	44.97
COPD (GOLD 4)	6965	66.99
Smoker Control	5771	67.44
Non-smoker Control	6788	80.33
COPD (GOLD 4)	6971	97.05
Smoker Control	2431	149.88
Cystic Fibrosis	5928	172.34
Cystic Fibrosis	5894	422.63
Cystic Fibrosis	5901	3617.11
Cystic Fibrosis	5723	5877.63

Sample Group	Core ID	Bacteria Cells/1000 Human Cells
Cystic Fibrosis	5938	10570.51
Cystic Fibrosis	5915	20154.49
Cystic Fibrosis	6058	68634.84
Cystic Fibrosis	2877	1533044.70

4.3.3 Terminal Restriction Fragment Length Polymorphism Analysis

Not every PCR reaction for T-RFLP analysis was successful. Only the product of the PCR with the highest intensity on a 1% agarose gel for each sample was used. A brief breakdown of the average number of positive reactions is presented [Figure 16]. What this graph shows essentially is that if 10 reactions were carried out for a particular group how many of them would turn out as positive. After Bonferroni correction four comparisons were still significant. The COPD (GOLD 4) and cystic fibrosis groups had a significantly higher percentage of positive reactions when compared to the non-smokers ($P = 0.0002$ for both the COPD (GOLD 4) and cystic fibrosis) and negative controls ($P = 0.001$ for the COPD (GOLD 4) and $P = 0.0015$ for the cystic fibrosis group). There was no significant difference seen between the Non-Smokers when compared to both the Smokers ($P = 0.01$) and Negative controls ($P = 0.84$). There was also no significant difference seen between the Smokers when compared to the COPD (GOLD 4) group ($P = 0.15$), the cystic fibrosis group ($P = 0.28$), and negative Controls ($P = 0.02$) [Figure 16].

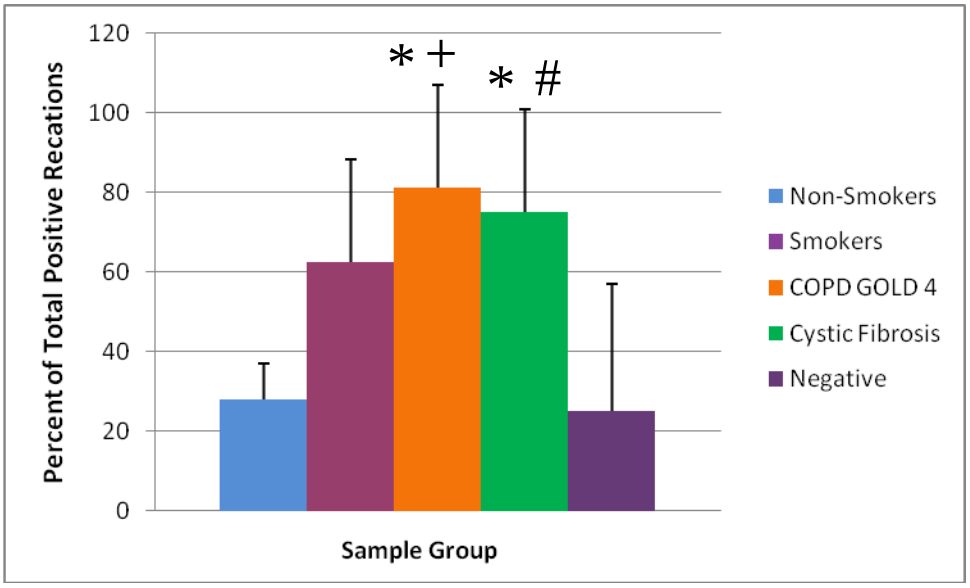

Figure 16: Percent of total PCR positive reactions befor T-RFLP analysis.
The bars show the mean ±SD of the number of positive reactions expected /100 attempts for each of the 4 experimental groups plus the negative controls. Following Bonferroni correction the COPD (GOLD 4) and cystic fibrosis groups were significantly different from both the non-smoking control group (*, P = 0.0002 for both) and the negative controls (+, P = 0.001 for the COPD (GOLD 4) group and #, P = 0.0015 for the cystic fibrosis group).

Upon completion of T-RFLP one COPD (GOLD 4) sample did not have enough material to be measured, so this group's total number went from eight to seven. When the relative peak intensity is graphed against the fragment length clear similarities between the non-smoking and smoking group can be seen [Figure 17]. Differences distinguishing the COPD (GOLD 4) group from the smoking, non-smoking, and cystic fibrosis group can also be visualized. Finally, the cystic fibrosis group can be seen as distinct compared to the other three groups that were analyzed.

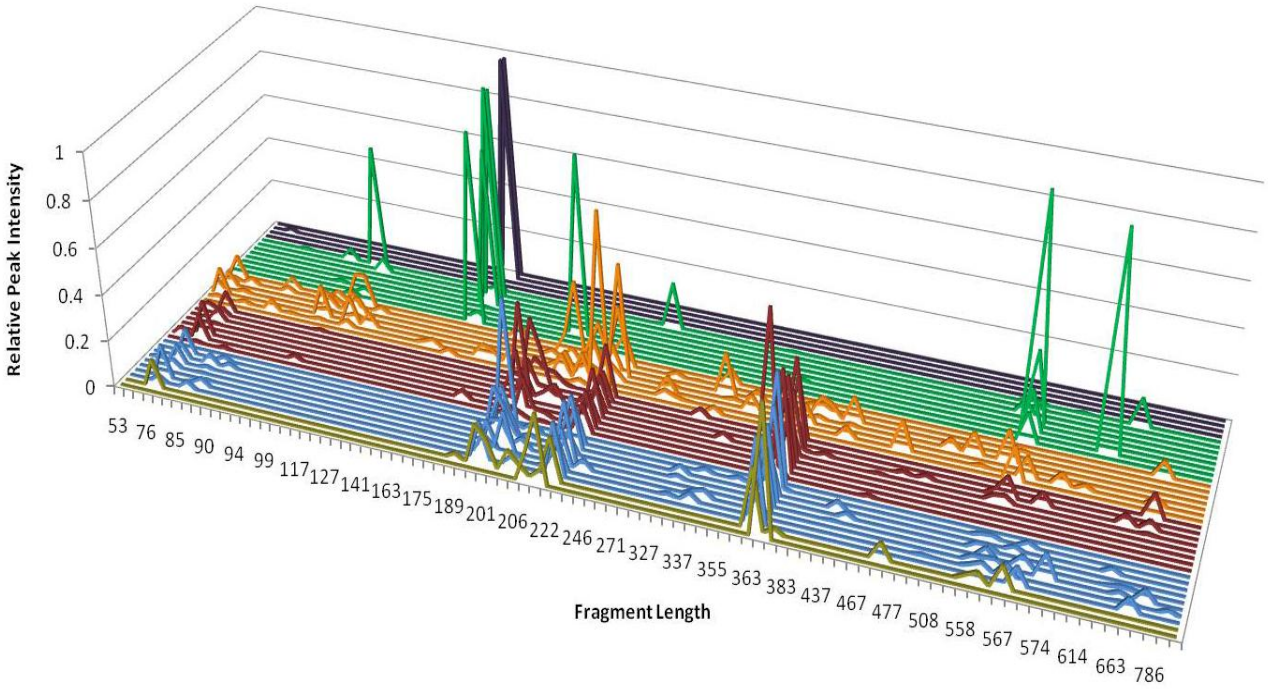

Figure 17: 3-D graph of relative peak intensity versus fragment length by group. The COPD (GOLD 4) group only had an N = 7 since one sample did not have enough product to be analyzed by T-RFLP. The dirty green represents the two *E.coli* positive controls, the light blue represents the non-smokers, the red are the smokers, the orange are the COPD (GOLD 4), the green are the cystic fibrosis, and the dark purple are the negative controls.

The negative controls that were submitted had two samples that were negative and two samples that had a single high intensity peak. The negative controls that were positive had their single peak correspond to a peak found in many of the cystic fibrosis samples. The positive controls which were *E.coli* DNA had more peaks and showed similarity to the non-smoking and smoking groups. However, both the non-smoking and smoking groups had additional peaks that did not correspond to the positive controls group. The COPD (GOLD 4) group had peaks that also matched to the positive controls group. However, as an observation, the COPD (GOLD 4) group had more peaks then the non-smoking, smoking, cystic fibrosis, negative control, and positive control groups.

Upon subtraction (as explained in section 3.7.3.3) of those peaks that were associated with the positive and negative control samples the cystic fibrosis samples lost their major peak. There are still similarities between the Smoking and Non-Smoking group, however, these similarities do not seem as strong as before subtraction. The cluster observed in the COPD (GOLD 4) group that was different then all the other groups remain upon applying the subtractions [Figure 18].

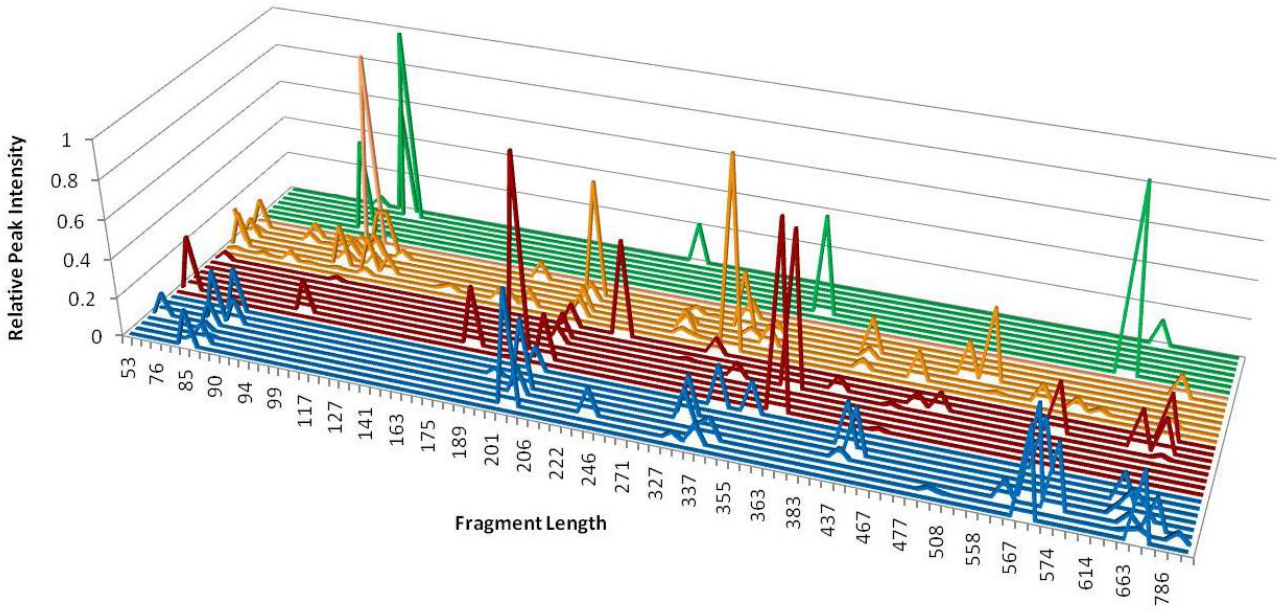

Figure 18: 3-D graph of relative peak intensity versus fragment length of each group after subtraction of the positive and negative controls.
The blue group is the non-smokers, the red group is the smokers, the yellow group is the COPD (GOLD 4) and the green group is the cystic fibrosis individuals. Since the peaks are relative intensity some peaks are more noticeable upon the removal of the dampening effects of the peaks associated with the *E.coli* positive controls. The COPD (GOLD 4) group shows a cluster difference versus the non-smokers and smokers that was also seen before subtraction of the negative and positive controls.

When the peak intensities of the T-RFLP results from the positive and negative control groups were not removed from those of the patient samples, the NMS ordination plot showed that the individuals in the smoking and non-smoking group clustered together in

a two dimensional graph [Figure 19]. The cystic fibrosis individuals clustered away from the smoking, non-smoking, and COPD (GOLD 4) groups. The COPD (GOLD 4) individuals clustered away from the smoking, non-smoking, and cystic fibrosis groups. In essence the NMS graph showed, in two dimensions, three distinct groups: one made up of the smokers and non-smokers, one made up of the COPD (GOLD 4), and one made up of the cystic fibrosis. The positive and negative control groups were not included in the NMS ordination plot since there were only two samples in each group. This was not enough to get an accurate profile of how these samples clustered.

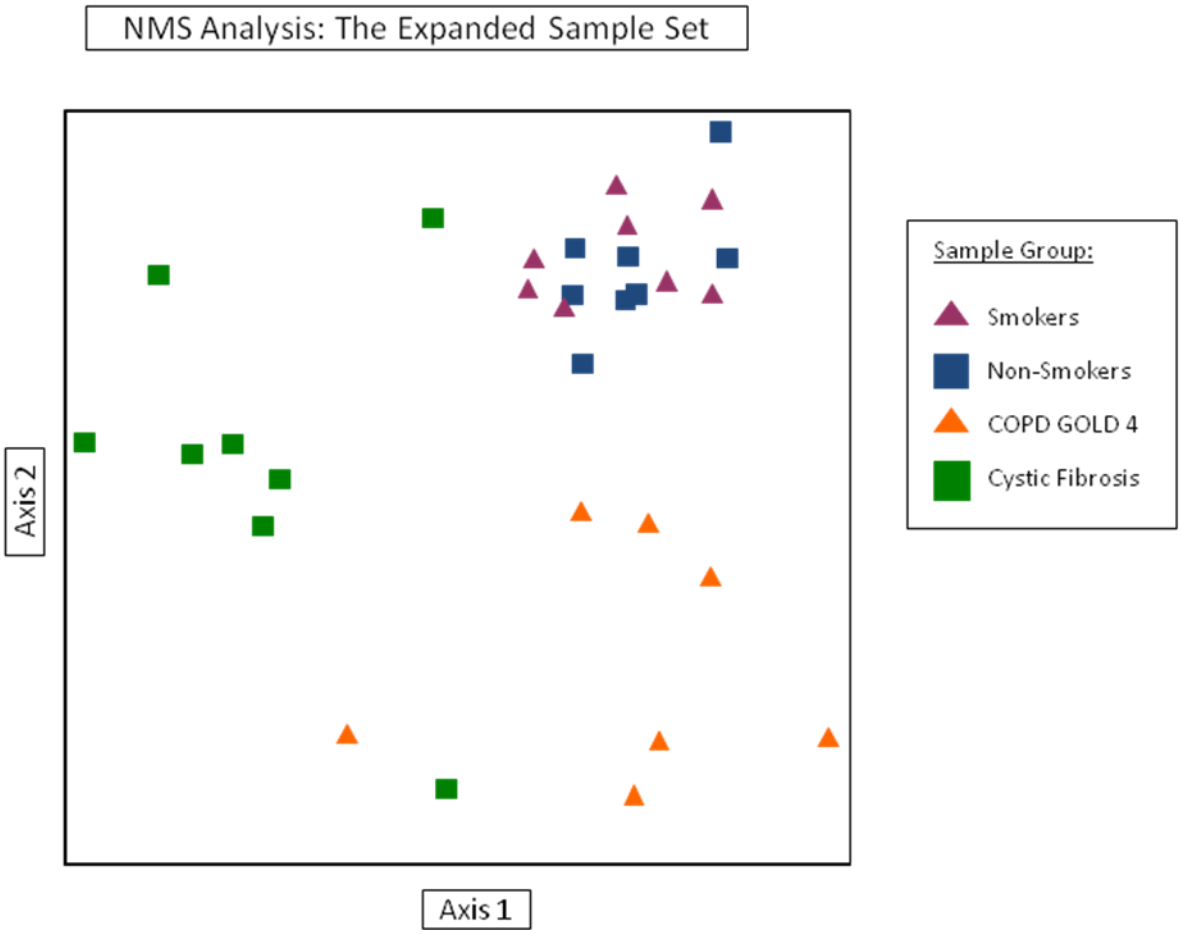

Figure 19: NMS ordination plot of the T-RFLP results not including the positive and negative controls.

Table 12: Multiple comparison MRPP analysis without positive and negative controls.
All values except the smoker versus the non-smoker group were significant after Bonferroni correction was applied.

	T-Value	A-Value	P-Value
Smokers vs. Cystic Fibrosis	-7.92	0.24	1.26E-05
Smokers vs. COPD GOLD 4	-7.80	0.35	0.000139
Smokers vs. Non-Smokers	1.04	-0.03	0.859892
Cystic Fibrosis vs. COPD GOLD 4	-5.68	0.19	0.000173
Cysitic Fibrosis vs. Non-Smokers	-7.37	0.21	8.86E-06
COPD GOLD 4 vs. Non-Smokers	-7.80	0.34	0.000126

An MRPP analysis was performed on the four different groups [Table 12]. This table shows a value called the A-value. This value shows the within group agreement of the different points when compared to another group. When the A-value is close to 1 or positive the individual points within the groups are similar to each other. If the A-value is exactly 0 then the differences between the individual points (the heterogeneity) within groups is equal to that expected by chance. Finally when the A-value is less than 0, that is negative the lack of difference (the heterogeneity) between the individual points between the two groups is greater than that expected by chance [181]. The T-value, on the other hand, shows the difference between groups with the larger negative values showing groups with a larger difference from one another [181]. The P-value shows whether these differences can be attributed to chance or not, with a lower P-value illustrating a lower probability that the event happened by chance. All group comparisons, except that between the Smoking and Non-Smoking groups, had a P-value that was less than 0.0002. The result from this MRPP analysis agrees with the previous MRPP analysis with when both the negative and positive controls are included. That is to

say that the comparisons that showed a significant difference remained regardless of whether the positive and negative controls were included.

When peak intensities of the negative control and positive control groups were removed from that of every patient sample and revised relative intensities were generated, the cysitic fibrosis samples, due to their close relation to the negative controls could not be analyzed with NMS. Thus only the non-smokers, smokers, and COPD (GOLD 4) groups could be analyzed using the NMS ordination [Figure 20]. The differences between groups were less than in the previous NMS analysis and there was more overlap between groups.

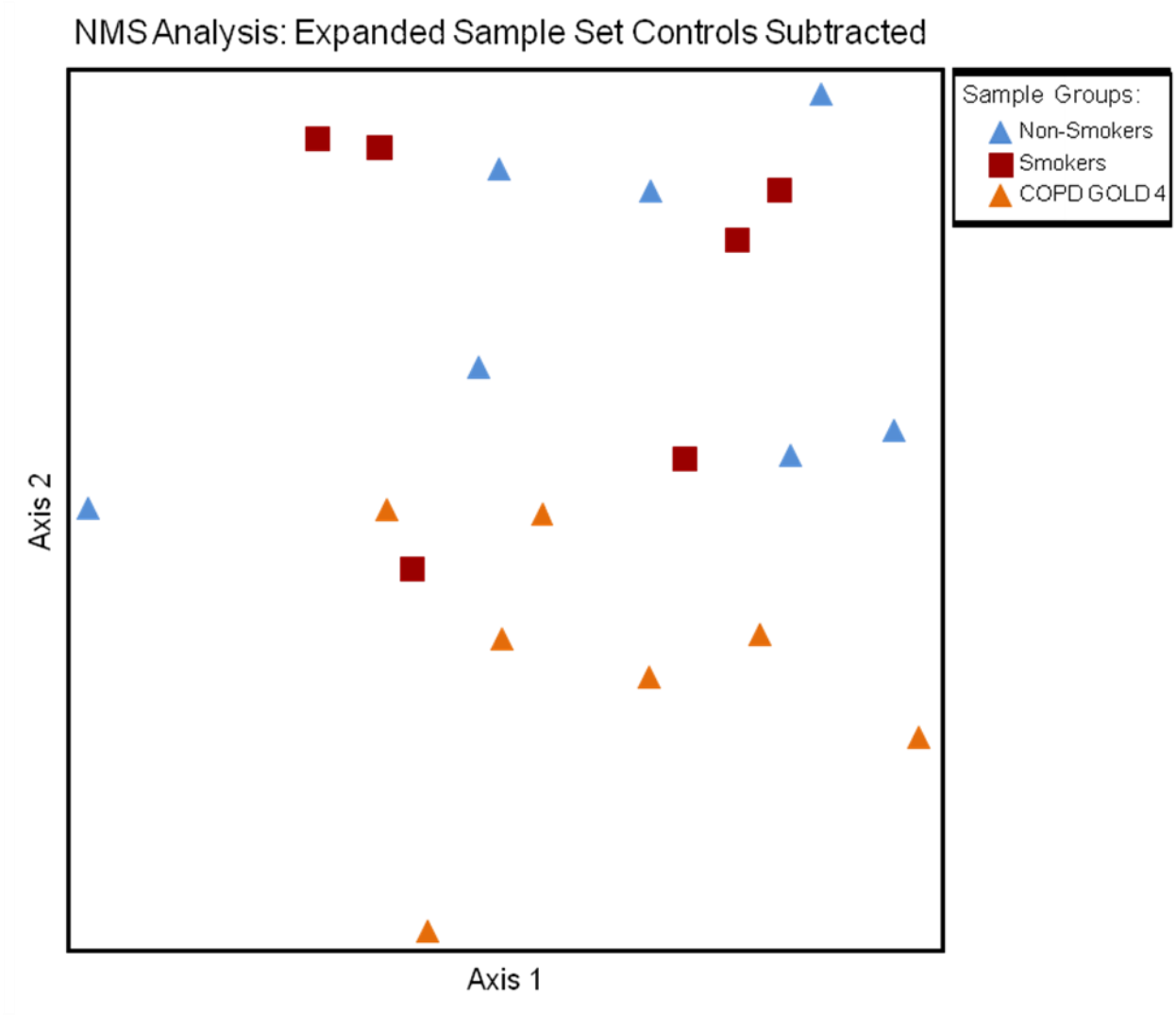

Figure 20: NMS ordination after the peak intensities of the negative and positive controls were removed from those of the samples.

When the MRPP analysis was performed on the NMS ordination with the controls removed it was found that even though there was more crossover between groups there was still a significant difference between the COPD (GOLD 4) group and both the non-smokers and smokers [Table 13].

Table 13: Multiple comparison MRPP analysis with negative and positive controls removed.
All values except the Smoker versus the Non-Smoker group were significant after Bonferroni correction was applied.

	T-Value	A-Value	P-Value	Significant
Non-Smokers vs. Smokers	-1.06	0.05	0.15	
Non-Smokers vs. COPD GOLD 4	-3.46	0.16	3.54E-03	*
Smokers vs. COPD GOLD 4	-3.32	0.14	3.52E-03	*

4.3.4 GS-FLX 454 Targeted Sequencing

As with the T-RFLP analysis the original PCR reactions to generate PCR products for 454 targeted sequencing were not positive every time [Figure 21].

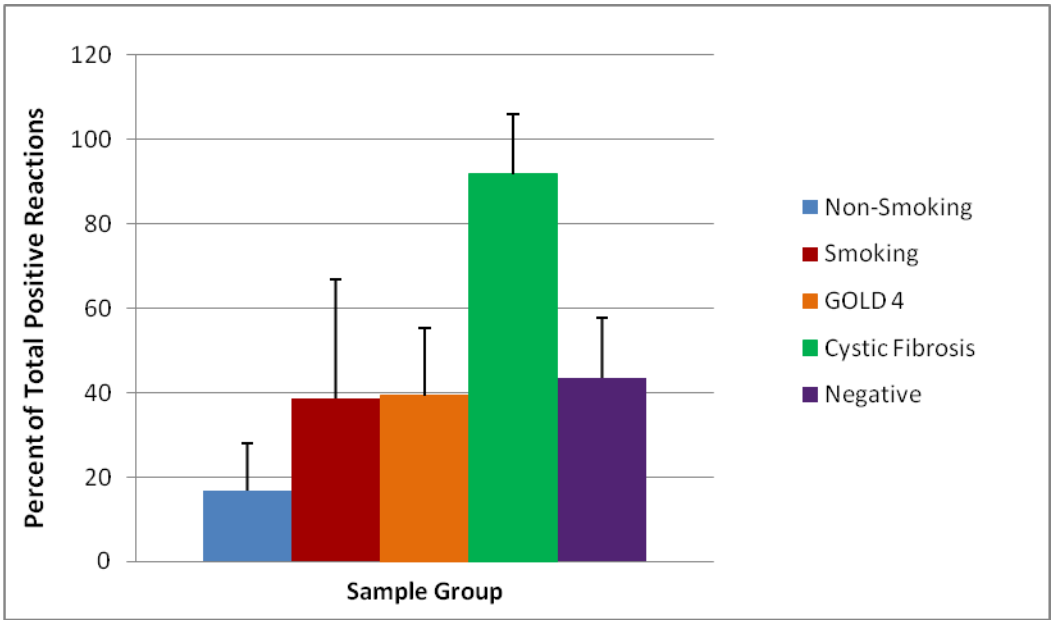

Figure 21: Percent PCR positive reactions before 454 targeted sequencing.
The bars show the mean ±SD of the number of positive reactions expected /100 attempts for each of the 4 experimental groups plus the negative controls

Further, there were not very many strong positive reactions, as visualized on a 1% agarose gel. Thus not every reaction could be pooled together. The Cystic Fibrosis

group was significantly greater than all other groups ($P < 0.05$) [Figure 21]. All other group comparisons showed no significant difference ($P > 0.05$). Upon correction with Bonferroni only the cystic fibrosis versus the COPD (GOLD 4) and the cystic fibrosis versus the smoking group remained significant. During the troubleshooting process, described in section 3.8.4, a secondary product was observed besides the primary product (regardless of whether human DNA was present or not). This can be seen in a PCR experiment using serial dilutions of *E.coli* cells with the modified 454 sequencing primers (MID-1 with the universal reverse primer) [Figure 22].

Due to this secondary band both a modified PCR purification protocol and a gel extraction protocol were tested. The gel purification protocol did not yield an adequate amount of DNA amplicon material required for sequencing. After following the dilution guidelines for the Buffer NT, to remove the 100bp product, both a 1:5 and a 1:6 dilution with DNase and RNase free water, successfully removed the 100bp products during the PCR purification [Figure 23].

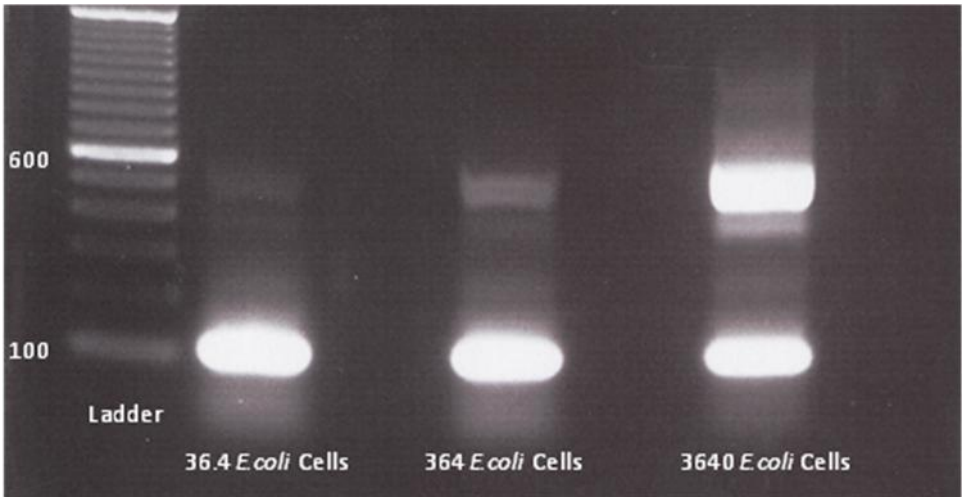

Figure 22: 1% Agarose gel of 454 targeted sequencing PCR products from serial dilutions of an *E.coli* positive control.
The primary product can be seen that is slightly smaller than the 600bp ladder marker. The secondary product is a very strong intensity band that is slightly larger than the 100bp ladder marker.

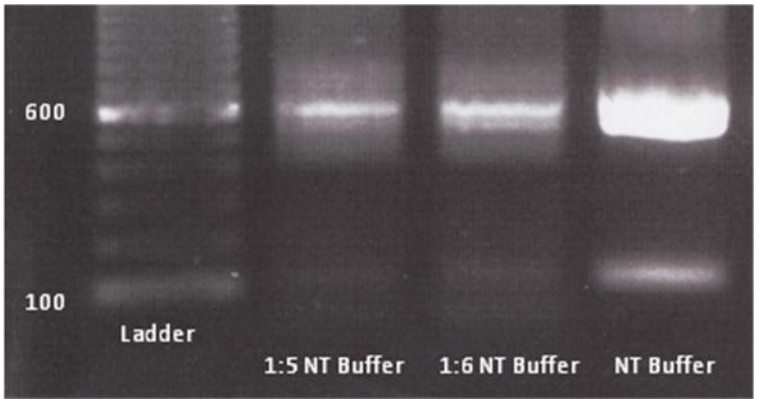

Figure 23: 1% Agarose gel of 454 targeted sequencing PCR products from serial dilutions of NT Buffer.
The PCR reaction was done on genomic DNA from approximately 3640 *E.coli* cells. The dilutions were 1 part NT buffer followed by either 5 or 6 parts DNase and RNase free water.

Further experimentation with the cystic fibrosis samples and the 1:6 dilution of the NT buffer showed that it is possible to remove the secondary product but there was a large reduction, as in the gel purification, in the yield and thus concentration of DNA amplicon product obtained [Figure 24].

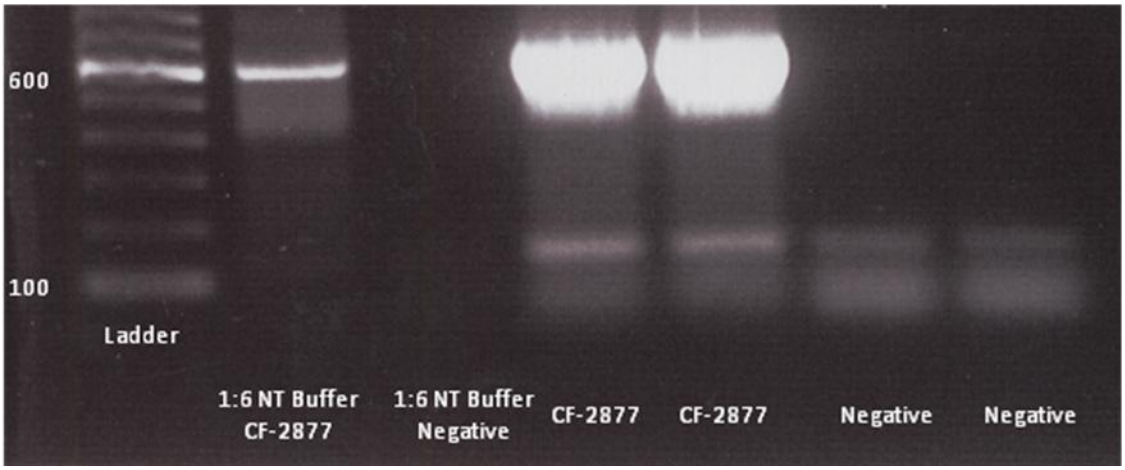

Figure 24: 1% Agarose gel of 454 targeted sequencing PCR products from a cystic fibrosis sample.
The negative controls are PCR reactions done on water without any template DNA.

Even with gel extraction, purification, and pooling of samples there was not enough amplicon DNA that could be used for sequencing except with the cystic fibrosis group. Thus, using the standard GS-FLX 454 targeted sequencing approach, it was not possible to sequence any of the other products from the expanded sample set.

It was found that the cystic fibrosis sequencing results showed relatively good agreement with the bacterial community composition demonstrated by T-RFLP analysis [Figure 25]. That is to say that samples that had different bacteria as identified by 454 sequencing clustered apart from the other cystic fibrosis samples in the T-RFLP analysis.

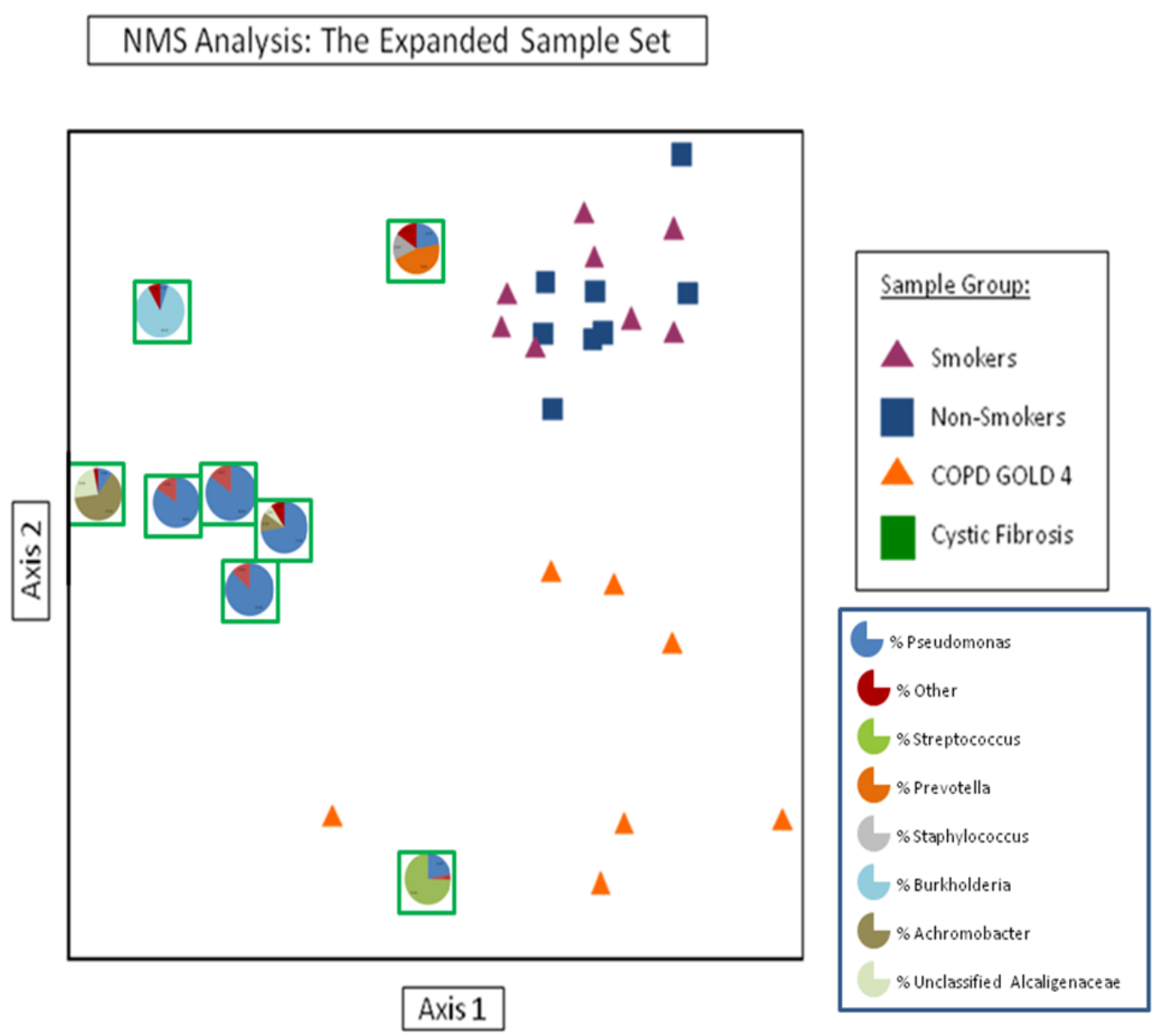

Figure 25: Targeted sequencing results superimposed onto the T-RFLP analysis graph.
The samples from the cystic fibrosis patients are represented by pie charts that are colour coded to represent the percentage of the different genus of bacteria found in each sample.

Chapter 5: Discussion

5.1 Summary

Using the Illumina platform for metagenomic sequencing we were able to quantify that a small number of bacteria were present in the samples from two individuals, one with COPD (GOLD 4) and the other, a smoker without airway obstruction, but we were unable to unequivocally identify the bacteria. Using both PCR and qPCR methods, we successfully confirmed the results of the metagenomic sequencing. A small number of bacteria could be identified in lung tissue samples. However, with the addition of both positive cystic fibrosis controls and negative non-template controls to our assays we found that the total bacteria quantified was close to what could be detected in the negative controls, possibly from random contamination, but significantly less than the total bacteria found in cystic fibrosis samples. An expanded sample set was created to determine if these results held up on a larger scale and to explore Hilty, et al's statement that total number of bacteria stays relatively constant while the community composition shifts in disease [158]. The data we generated showed that with more samples, total numbers of bacteria in the lung samples were significantly higher than the negative non-template control background. Further, there was no significant difference among our non-smoking and smoking control groups and our COPD (GOLD 4) group. This agrees with Hilty et al's finding that the number of bacteria does not change [158]. Using T-RFLP analysis to assess the bacterial community composition we found three distinct groups: one with the cystic fibrosis patients, one with the non-smoking and smoking controls, and one with the COPD (GOLD 4) patients. Using this method we have partially confirmed the other aspect of the Hilty et al's findings that bacterial community

composition shifts in disease [158]. This is a partial confirmation since we were unable to sequence these samples to identify phyla, class, or genus of bacteria that may be associated with this shift in bacterial community composition. However, the cystic fibrosis patient samples that we were able to sequence showed good agreement with our T-RFLP analysis for bacterial community composition.

5.2 General Comments

An advantage the present sample acquisition method has over previous studies of the lung microbiome [158, 159] is that samples are not exposed to the risk of contamination involved in traveling through the distinct microbiomes that inhabit the nose and mouth. However, the difficulties involved in eliminating all possible contamination from a variety of random sources within the laboratories in which the lung tissue samples were processed remained a concern. The fact that the negative controls often showed low levels of products from PCR, qPCR, and T-RFLP (see results section) analyses was interpreted as contamination, because they were consistently lower in quantity and community composition. Indeed, it is not unreasonable to attribute the amount and difference in community composition in the negative control samples to contamination because it is so small when compared to other microbiomes, like the gastrointestinal tract, [141, 147, 182] which contains between 10-100 trillion bacteria cells [141]. Previously published studies on the lung microbiome have shown that a large and diverse population of bacteria exist in the lungs [158, 161]. Yet this is different from our present study which demonstrates a relatively small bacterial community in the lungs of both donor controls (smokers and non-smokers without airway obstruction) and individuals with

very severe COPD. This disagreement between previous studies may be due to the differences in methodology used. Further, with the demonstration of a microbiome that was four orders of magnitude greater in lungs from patients with cystic fibrosis than in either the control or COPD cases examined strongly suggests a fundamental difference in the nature of the microbiome in cystic fibrosis compared to COPD and controls [Figure 14]. Although it is very unlikely that eliminating all sources of contaminating organisms can account for the very high levels of bacteria found in cystic fibrosis it is possible that this type of contamination might explain some of the differences between this study and the levels previously reported for the microbiome in COPD.

5.3 Experiment 1: Preliminary Ilumina Sequencing Project

Although the cost of metagenomic sequencing has decreased substantially in recent years it remains the major barrier to studying the lung microbiome. The preliminary studies reported here show that the vast majority of the sequenced reads were human DNA (greater than 99% of reads for all 7 samples). Indeed the relatively small numbers of microbial reads made it impossible to separate the bacterial reads according to class, phyla, genus, etc. due to the fact that the length of the reads were too short. Therefore metagenomic sequencing was only able to provide the total number of bacterial reads per sample. The Illumina platform showed that the five COPD (GOLD 4) samples from one individual were generally higher (2.8±0.9 bacterial cells / 1000 human cells) than the two samples from the smoking control individual (1.8 ± 0.9 bacterial cells / 1000 human cells). However, this difference was not significant ($P = 0.32$). This could be for a number of reasons. First, our sample size was relatively small and it is possible that with

a larger sample size we may be able to show a significant difference. Second, not only are we comparing one COPD (GOLD 4) to one smoker but also are only comparing the average of five samples to two samples. Again the sample size is too small to draw any definitive conclusions. Third, there may not be a measurable microbiome present in lung tissue and the bacterial sequences that showed up could have been due to contamination introduced from the reagents used during DNA extraction or during the processing of the DNA for sequencing. Though [Figure 9] and [Figure 10] show sample-dependent properties, with respect to bacterial counts and either lung height or Lm, caution should be shown in trying to interpret these results. Both graphs in the figures are highly influenced by the one high bacterial read result from the COPD (GOLD 4) patient, more so for [Figure 10]. Finally, the assumption that bacterial reads are an adequate estimate of the total number of bacteria present in the lung tissue is not necessarily correct. Since the number of reads is not only affected by the numbers of genomes of the organisms of interest that are present, but also is influenced by their genome sizes, that is larger genomes yield relatively more reads than smaller ones, corrections for genome size, as discussed in section 3.5.3 should be applied. Since 75 base pair reads are not long enough to allow for proper differentiation of the bacteria from the respective reads, an accurate correction was not possible. Therefore, it is possible that the difference we observe between the one COPD (GOLD 4) individual and the one smoking control individual may not actually reflect a real difference in total bacteria present within the tissue. It is possible that the total number of bacteria in both samples is equal even if the number of reads associated with bacteria is not. The other major problem with using this method was that we were unable to generate enough coverage of the human genome to be

able to accurately sequence the genome of the two individuals that were being compared. In order to adequately and accurately sequence the human genome using this approach a 3-fold coverage of the genome is needed, on average we reached a 1-fold coverage for the five COPD (GOLD 4) samples and the two smoking control samples. It may be worthwhile to expand the sample size tested if this method is to be explored in more detail. It would also be important for future studies to incorporate enough DNA so that the coverage is large enough so human DNA can be used for further analysis. In general, as a proof of concept the experiment showed that it is possible to detect small numbers of bacterial reads in a large pool of human reads in DNA extracted from human lung tissue.

5.4 Experiment 2: Validation of Metagenomic Sequencing with PCR and qPCR

5.4.1 PCR

The major problems with the PCR assays were background noise and reproducibility of the results. Even under the cleanest conditions within our laboratory the negative controls (RNase and Dnase free water) were often as positive as the tissue from the COPD (GOLD 4) individuals and the smoking and non-smoking controls. This suggests that either there was random contamination of the samples or reagents or that the bacterial DNA in the sample was different in bacterial composition but near background levels, with respect to total bacteria present. Although all three primer pairs that were used generated similar results (data not presented) the 27F with the 1390R primers generated a secondary product that was approximately 120bp in length. This secondary

product was only observed when sample DNA, which contained human genomic DNA, was used as a template and was not observed in the positive control (*E.coli* genomic DNA) that yielded the expected 1364bp product (data not shown) and in the negative controls to which DNA was not added. This particular primer pair had not been previously used in a background of human genomic DNA [175, 183, 184]. Further evaluation of this primer pair by one of my supervisors (Dr. Shizu Hayashi) showed that the 1390R primer had a 100% homology to a sequence in a similar region of the human 18S rRNA ribosomal gene and that the 27F forward primer had 7-8 mismatches with a sequence about 120 bp 3' of the 100% homology sequence. This suggests that PCR reactions using these primers can yield a secondary product from the human 18S rRNA gene that might explain some of the problem with reproducibility with these primers in situations where the ratio of human to bacterial DNA was very large and why the human lung samples came up negative for bacterial DNA multiple times. Using other primer pairs generated a similar result, with respect to contamination [Figure 11]. However, when the PCR mix was changed to a new batch this secondary product only persisted in the 27F and 1390R primer pairs and vanished in all other primer combinations used, as evidenced by [Figure 8] (although this was a 1% gel after a qPCR reaction the result is the same as PCR with new mix).

When PCR was completed on the multiple samples from the one cystic fibrosis individual the samples consistently came back as positive. However, there was a substantial difference with respect to band intensity for the different samples within a single individual cystic fibrosis patient. This would tend to suggest that the bacteria were not evenly distributed within the lung of a cystic fibrosis patient and that some areas

remained relatively free of bacteria. In contrast, multiple samples from the COPD (GOLD 4) individual did not yield bands with different intensity when they were positive. Instead, all COPD (GOLD 4) bands were evenly weak. This agrees with the metagenomic results that show only a small difference in bacterial reads among samples from the same individual.

5.4.2 qPCR

The qPCR results from the single lung affected by cystic fibrosis confirmed the PCR experiments on the same samples [Figure 13]. The qPCR on the samples of the COPD (GOLD 4) individual that were sent to Genome BC for the metagenomic sequencing, before correction, showed a small total number of bacteria that was similar from sample to sample. This agrees with the metagenomic sequencing data that found a low number of total bacterial reads in the same COPD (GOLD 4) samples. What is interesting is that, upon application of the correction factor, all the COPD (GOLD 4) qPCR results for the samples sent to Genome BC were effectively zero. Two possible reasons for this arise. First, we may be sampling the random background noise in our Genome BC COPD (GOLD 4) samples. Second, the bacteria that were found within the lungs were so few in number that they were equivalent to what could be found from sampling background or random contamination.

An important consideration that should be made when quantifying the bacteria utilizing the 16s rRNA gene is that different bacteria contain different numbers of the 16s rRNA genes and rarely do bacteria only contain one copy. For example, the *E.coli* that was used for the standard curve generation in the qPCR experiments contain seven copies of

the 16s rRNA gene. Thus, quantifying an exact number of bacteria using this method may not be possible. However, it is likely that, even with this variation in 16s rRNA gene copies among bacteria the estimated bacteria/1000 human cells are the right order of magnitude.

Despite problems with the qPCR assay it is quite clear that two different methods, the Illumina sequencing and the qPCR assay agree with respect to the relative amount of bacteria that can be found in a COPD (GOLD 4) individual.

5.5 Experiment 3: Expansion of Sample Size to Detect Bacterial Differences Between Sample Groups

5.5.1 qPCR

It was found that there was no significant difference between the non-smoking, smoking, and COPD (GOLD 4) group with respect to total number of bacterial cells as measured using the 16s rRNA gene as a target. In contrast, the cystic fibrosis group showed a multiple order of magnitude difference from all other groups as well as much greater variability. This could be expected since PCR experiments had shown that in a single individual the bacterial loads of the cystic fibrosis varied from sample to sample. Thus even though we selected samples from a similar location in the lung in different individuals with cystic fibrosis it is likely that each individual's distribution of bacteria in their lungs varies. Another interesting result was that all of the groups were significantly higher than the negative control group. However, the non-smoking, smoking, and COPD

(GOLD 4) groups were relatively close to the negative control group while the cystic fibrosis group was clearly much larger than the negative controls.

Re-sampling of the same cystic fibrosis sample (5928-7) was done since it was also used for the expanded sample set and the numbers compared relatively well for total bacteria, (3.12x10^3) ± (1.55 x10^3) [Table 8] versus (2.71x10^3) ± (8.89 x10^2) [Table 10]. There is a rather large standard deviation but this was also seen in other studies of the lung microbiome [158]. The amount of bacteria found in the qPCR assay was relatively robust, as shown by the close agreement of two independently extracted and sampled genomic DNA samples from 5928-7. Upon application of the correction factor, the non-smokers, smokers, and COPD (GOLD 4) groups had a drastic drop in their bacterial cell/1000 human cells. The non-smoking group had 35±22, the smoking group had 44±47, and the COPD (GOLD 4) group had 24±37 bacteria cells/1000 human cells. However, the bacteria cell/1000 human cells in the cystic fibrosis group were drastically larger at (2.05x10^5) ± (5.37x10^5). This data agrees with the metagenomic sequencing data that found only a small number of bacterial reads in the different samples of the one COPD (GOLD 4) individual (5 samples) and the one smoking individual (2 samples). However, the estimated bacteria/1000 human cells from the metagenomic sequencing were smaller than that found in the qPCR experiments with the expanded sample set ([Figure 9] vs. [Table 7]). A different sample from the same individual (6968) was used in the metagenomic sequencing and the qPCR assay. Both results showed relative closeness to each other, with the metagenomic sequencing giving an average lung tissue sample reading of 2.8±0.9 bacteria/1000 human cells [Table 6] and the qPCR on the

single different sample from the same individual being 5.0 bacteria/1000 human cells [Table 11].

5.5.2 Terminal Restriction Fragment Length Polymorphism Analysis

As seen in the results section, not every initial PCR reaction was positive. The COPD (GOLD 4) group and cystic fibrosis group had the highest percent positive reactions (81.3±25.9 and 75±25.9 respectively) than the non-smokers [Figure 16]. The standard deviations were quite large but may be indicative of large spatial variation in the amount of bacteria present in each sampled individual. The negative controls were positive a similar number of times as both the non-smoking and smoking groups. When the relative peak intensity of the restriction fragments generated, by this analysis, was graphed against the fragment length, certain similarities and differences could be seen. First, the positive control that used *E.coli* had bands that matched very well to bands in the non-smoking, smoking, and COPD (GOLD 4) group. The positive control did not have peaks that matched well with the cystic fibrosis group. Second, the negative controls had two samples that were below the level of detection and two samples that had a single peak that matched a peak found only in the cystic fibrosis samples. Third, there were small differences that could be seen between the non-smoking group versus the COPD (GOLD 4) group and the smoking group versus the COPD (GOLD 4) group (most notable in the 87-120bp fragment range). The fact that the positive controls matched well with those in the lung samples, excluding the cystic fibrosis group, may mean that either background contamination from the positive interfered with these results or that a bacterium affiliated with *E.coli* was present in the lung samples. The most likely reasoning as to why the

negative controls had a peak identical to that found in the cystic fibrosis group would be that the negatives were contaminated by the cystic fibrosis samples. Alternatively, it could be proposed that the negative controls could be contaminating the cystic fibrosis samples.

When an NMS ordination graph was created [Figure 19] there were three distinct groups identified. The non-smokers and smokers made up one group, the COPD (GOLD 4) samples comprised another group, and the cystic fibrosis samples comprised the last group. Although not shown the negative non-template controls clustered with the cystic fibrosis group and the positive *E.coli* controls spanned both the group made up of the non-smokers and smokers as well as the group made up of the COPD (GOLD 4). This suggests that the difference seen between the non-smoking and smoking group versus the COPD (GOLD 4) group could be due to species that are similar to *E.coli*. With that being said further positive controls would need to be included to allow for a complete group analysis to be done using this method.

When both the positive and negative controls were subtracted from the results and a relative intensity re-calculated, the difference between the non-smoking and smoking group versus the COPD (GOLD 4) group remains and was still statistically significant (non-smokers versus COPD (GOLD 4), p-value = 3.54×10^{-3} and smokers versus COPD (GOLD 4), p-value = 3.52×10^{-3}) [Table 13]. This indicates that the peaks affiliated with *E.coli* within samples do not explain the variation that was observed. The MRPP analysis confirmed that the groups were significantly different from one another. This analysis used statistical terms used in ecology (T-value and A-value) to help define differences between groups. Regardless of the manipulations done, the analysis showed that both the

non-smokers and smokers were similar both between group (A-Value) and across group (T-Value). Further these analyses consistently showed that the COPD (GOLD 4) group differed from the non-smoking and smoking group and cystic fibrosis group. Also the individual samples within each of the different groups were similar to each other. This consistency in MRPP analysis suggests that the differences between the groups may not be an artifact of simple *E.coli* contamination of samples.

There are a few important limitations that need to be stressed with respect to this technique in assessing bacterial communities. First, the initial PCR will tend to amplify the more abundant species over the less abundant species or differentially amplify certain fragments over others [185, 186]. This can lead to an incorrect picture of community structure. Second, there can be partial digestion of products by the restriction enzyme which could overestimate the true community composition of the sample [185, 187]. This partial digestion cannot be entirely corrected for [187, 188]. This is important to remember since it would increase the community composition observed in the T-RFLP analysis. Along this line, pseudo-terminal restriction fragments can also be observed [188]. These additional restriction fragments of unexpected size are detected when using cloned copies of a specific bacterial 16S rRNA gene where only one labeled fragment of known size should be detected [188]. These unexpected fragments can originate from single stranded amplicons due to secondary structure formation of the template molecule. This allows for partially single stranded DNA to be cut by the restriction enzyme (as long as this secondary structure contains the GCGC cut site of the *HHaI* restriction enzyme that was used). Thus, it is possible that some of the peaks in the positive control were due to pseudo-T-RFLP and what should have been a single clear peak turned out to be

multiple peaks. Again, the presence of the pseudo-T-RFLP fragments may mean that there was an overestimation of the bacterial populations. It is important to keep in mind that though we have found bacterial 16S rRNA gene community composition differences between groups there are some drawbacks to this technique that could explain some of the possible variances between these groups. Therefore the ability to sequence these samples to identify the possible class and phyla of bacteria present is important.

5.5.3 GS-FLX 454 Targeted Sequencing

For the non-smoker, smoker, and COPD (GOLD 4) samples our attempt to use 454 sequencing resulted in a number of problems, as detailed in the results section. Although these problems also occur with the cystic fibrosis samples, their impact was not as great since adequate PCR products were generated. One of the problems that occurred for every sample was the amplification of a secondary product around the 100 bp range. These secondary products could be due to the large primers that had to be used for this experiment. Attempts to purify the PCR products resulted in variable success. Sometimes the Machary-Nagel PCR purification kit worked, however, it did not work on some samples and the primers and secondary product remained even after PCR purification. Further there was not enough amplicon DNA after one PCR reaction for sequencing. This meant pooling the best two to three reactions together. Even with pooling we were not able to obtain enough PCR amplicon products to allow for good sequencing for the majority of the samples. The underlying problem for both roadblocks was either due to the low amount of bacterial DNA in the samples to begin with or due to an inefficiency of the primers to amplify the DNA optimally. Further there may have

been problems with the Machary-Nagel PCR purification kit itself since it had a variable PCR purification reliability. This was a major reason why a switch to the Qiaquick Gel Extraction kit was done. The results from the sequenced bacteria in the cystic fibrosis samples do suggest that the T-RFLP analysis does agree relatively well with the targeted sequencing. This helps to provide support that the differences we observe in the groups other than that of the cystic fibrosis group by T-RFLP analysis may be real. Thus a major priority should be to try and work out a solution to sequencing very low starting bacterial DNA within a background of a large amount of human DNA that can be reproducible.

5.6 Positive "Negative" Controls

It was not possible under our conditions to obtain a consistent negative signal for our negative non-template controls. Cycle number could not be reduced since the level of signal from our non-smokers, smokers, and COPD (GOLD 4) groups were too close to that of the negative controls. Further both the negative controls and the samples, excluding the cystic fibrosis samples, were positive a similar number of times. This is best shown in [Figure 16 & 21]. The only sample group that had a significantly consistent higher positivity than the negative controls was the cystic fibrosis group. Thus not only were the non-smoking, smoking, and COPD (GOLD 4) groups close in total bacteria to the negative controls but also were very similar in the number of times samples would come up positive. However, the negative controls had peaks that were most similar to the cystic fibrosis samples and not the non-smokers, smokers, and COPD (GOLD 4) samples [Figure 17]. The differences seen in [Figure 17] suggest that the PCR products are from bacteria that are initially present within the samples and are not the

same as the bacteria that contaminated the negative controls. Further, the T-RFLP analysis showed the there was a difference between the non-smoking and smoking group versus the COPD (GOLD 4) group. These findings provide support for the Hilty, et al's conclusion that the total number of bacteria does not change but the bacterial community composition narrows or shifts in disease [158]. This would mean that the total number of bacteria is similar to background but may be different between the different groups (non-smokers, smokers, and COPD (GOLD 4)). However, without sequencing and subsequent identification of the bacteria present in the non-cystic fibrosis samples, the T-RFLP findings cannot be confirmed.

Chapter 6: Conclusion

We have been able to successfully produce a methodology that can be used to identify and quantify bacteria within lung tissue of patients. Further, we were able to confirm what Hilty, et al has reported in a previously published work using bronchial brushings and BAL instead of lung tissue samples. The data in this thesis extend the current literature on the lung microbiome by showing, within lung tissue, that a distinct bacterial signature can be detected in the most severe form of COPD which is significantly different than the non-smoking and smoking controls. However, even though total bacteria can be identified from lung tissue and a difference in community composition can be demonstrated the one major thought that prevails is whether or not such a low number of total bacteria can have a significant impact on disease progression and outcome.

References

1. Celli, B., W. MaNee, and C. Memebers, *Standards for the Diagnosis and Treatment of Patients with COPD: A Summary of the ATS/ERS Position Paper.* Eur Respir J., 2004. **23**: p. 932-946.

2. Pauels, R., et al., *Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease: NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop Summary.* Am J Respir Crit Care Med, 2001. **163**: p. 1256-1276.

3. Jr., M.H.W. and N.S. Seriff, *Chronic Obstructive Pulmonary Disease: An analysis of Clincial, Physiologic, and Roentgenologic Features.* The American Journal of Medicine, 1962. **35**(1): p. 20-30.

4. Lowell, F.C., et al., *Chronic Obstructive Pulmonary Emphysema: A Disease of Smokers.* Ann Intern Med, 1956. **45**(2): p. 268-274.

5. Fletcher, C. and R. Peto, *The Natural History of Chronic Airflow Obstruction.* British Medical.Journal, 1977. **1**: p. 1645-1648.

6. Kanner, R., et al., *Lower Respiratory Illnesses Promote FEV(1) decline in Current Smokers but not Ex-Smokers with Mild Chronic Obstructive Pulmonary Disease: Results From the Lung Health Study.* Am J Respir Crit Care Med, 2001. **164**(3): p. 358-364.

7. Becklake, M., *Chronic Airflow Limitation: It's Relationship to Work in Dusty Occupations.* Chest, 1985. **88**(4): p. 608-617.

8. Churg, A., et al., *Chronic Exposure to High Levels of Particulate Air Pollution and Small Airway Remodeling.* Environmental Health Perspectives, 2003. **111**(5): p. 714-718.

9. Churg, A. and J. Wright, *Airway Wall Remodeling Induced by Occupational Mineral Dusts and Air Pollutant Particles.* CHEST, 2002. **122**: p. 306-309.

10. Kuempel, E., et al., *Contributions of Dust Exposure and Cigarette Smoking to Emphysema Severity in COal Miners in the United States.* Am J Respir Crit Care Med, 2009. **180**: p. 257-264.

11. Mathers, C. and D. Loncar, *Projections of Global Mortality and Burden of Disease from 2002 to 2030.* PLOS Medicine, 2006. **3**(11): p. 2011-2030.

12. Teculescu, D., *[Clear Definitions of Chronic Bronchopulmonary Diseases: A Necessity for Practice, Teaching and Research in Pneumology* Rev Pneumol Clin, 1990. **46**(5): p. 194-199.

13. Thurlbeck, W. and G. Angus, *A Distribution Curve for Chronic Bronchitis.* Thorax, 1964. **19**: p. 436-442.

14. Lumsden, A., A. McLean, and D. Lamb, *Goblet and Clara cells of Human Distal Airways: Evidence for Smoking Induced Changes in their Numbers.* Thorax, 1984. **39**: p. 844-949.

15. Yoshida, T. and R. Tuder, *Pathobiology of Cigarette Smoke-Induced Chronic Obstructive Pulmonary Disease.* Physiol Rev, 2007. **87**: p. 1047-1082.

16. Reid, L., *Measurement of the bronchial mucous gland layer: Diagnostic yardstick in chronic bronchitis.* Thorax, 1960. **15**: p. 132.

17. Reid, L., *Chronic Bronchitis and Emphysema.* Advances in Internal Medicine: Editors; William Dock, I. Snapper, 1964. **12**: p. 256-294.

18. Lamb, D. and L. Reid, *Goblet Cell Increase in Rat Bronchiolar Epithelium After Exposure to Cigarette and Cigar Smoke.* British Medical.Journal, 1969. **i**: p. 33-35.

19. Mullen, J., et al., *Reassessment of Inflammation of Airways in Chronic Bronchitis.* British Medical.Journal, 1985. **291**: p. 1235-1239.

20. Niewoehner, D., *The Role of Chronic Bronchitis in the Pathogenesis of Chronic Obstructive Pulmonary Disease.* Semin Respir Infect, 1988. **3**(1): p. 14-26.

21. Sethi, S., et al., *New Strains of Bacteria and Exacerbations of Chronic Obstructive Pulmonary Disease.* N Engl J Med, 2002. **347**(7): p. 465-471.

22. Gandevia, B. and D. Cowling, *Bacteriological Studies in Chronic Bronchitis.* Australasian Ann. Med, 1961. **10**: p. 275.

23. Balter, M., et al., *Canadian Guidelines for the Management of Acute Exacerbations of Chronic Bronchitis: Executive Summary.* Can Respir J, 2003. **10**(5): p. 248-258.

24. Seneff, M., et al., *Hospital and 1-Year Survival of Patients Admitted to Intensive Care Units with Acute Exacerbations of Chronic Obstructive Pulmonary Disease.* JAMA, 1995. **274**: p. 1852-1857.

25. Snider, G., et al., *The Definition of Emphysema: Report of a National-Heart-Lung-And-Blood-Institute, Division of Lung-Diseases Workshop.* Am Rev Respir Dis, 1985. **132**: p. 182-185.

26. Ryder, R., M. Dunnill, and J. Anderson, *A Quantitative Study of Bronchial Mucus Gland Volume, Emphysema, and Smoking in a Necropsy Population.* J Pathol, 1971. **104**(1): p. 59-71.

27. Anderson, J., M. Dunnill, and R. Ryder, *Dependence of the Incidence of Emphysema on Smoking History, Age, and Sex.* Thorax, 1972. **27**: p. 547-551.

28. Wright, J., et al., *Ranking the Severity of Emphysema on Whole Lung Slices. Concordance of Upper Lobe, Lower Lobe, and Entire Lung Ranks.* Am Rev Respir Dis, 1986. **133**(5): p. 930-931.

29. Rosenblum, L., et al., *Computed Tomography of the Lung.* Radiology, 1978. **129**(2): p. 521-524.

30. Sanders, C., P. Nath, and W. Bailey, *Detection of Emphysema with Computed Tomography: Correlation with Pulmonary Function Tests and Chest Radiography.* Invest Radiol, 1988. **23**: p. 262-266.

31. Petty, T., G. Silvers, and R. Stanford, *Mild Emphysema is Associated with Reduced Elastic Recoil and Increased Lung Size but not with Air-Flow Limitiation.* Am Rev Respir Dis, 1987. **136**: p. 867-871.

32. Reid, L., *Emphysema: Classification and Clinical Significance.* British Journal of Disease of the Chest, 1966. **60**(2): p. 57-65.

33. Calverley, P. and N. Pride, *Chronic Obstructive Pulmonary Disease*. Defining Chronic Obstructive Pulmonary Disease, ed. G. Snider1995: Chapman & Hall Medical.

34. Gough, J., *Pathological Changes in the Lungs Associated with Cor Pulmonale.* Bull. N.Y. Acad. Med, 1965. **41**(9): p. 927-941.

35. Thurlbeck, W., *The Incidence of Pulmonary Emphysema with Observations on the Relative Incidence and Spatial Distribution of Various Types of Emphysema.* Am Rev Respir Dis, 1963. **87**: p. 207-215.

36. Heard, B., *Further Observations on the Pathology of Pulmonary Emphysema in Chronic Bronchitis.* Thorax, 1958. **14**: p. 58-90.

37. Jr, A.A. and A. Foraker, *Centrilobular Emphysema and Panlobular Emphysema: Two Different Diseases.* Thorax, 1973. **28**: p. 547-550.

38. Jr, J.P.S., G.A. Falk, and G.W. Siskind, *Serum Immunoglobulins and Alpha-1-Antitrypsin: Variation with Clinical Type of Chronic Obstructive Pulmonary Disease.* Chest, 1971. **59**(5): p. 17-19.

39. Falk, G. and W. Briscoe, *Alpha-1-Antitrypsin Deficiency in Chronic Obstructive Pulomary Disease.* Ann Intern Med, 1970. **72**(3): p. 427-429.

40. Gross, P., et al., *Experimental Emphysema: It's Production with Papain in Normal and Silicotic Rats.* Arch. Environ. Health, 1965. **11**: p. 50-58.

41. Janogg, A., et al., *Experimental Emphysema Induced with Purified Human Neutrophil Elastase: Tissue Localization of the Instilled Protease.* Am Rev Respir Dis, 1977. **116**: p. 461.

42. MacNee, W., et al., *Cigarette Smoking Delays Polymorphonuclear Leukocyte Transit Through the Pulmonary Circulation.* Am Rev Respir Dis, 1987. **135**: p. 146.

43. Hiemstra, P., S.v. Wetering, and J. Stolk, *Neutrophil Serine Proteinases and Defensins in Chronic Obstructive Pulmonary Disease: Effects on Pulmonary Epithelium.* Eur Respir J., 1998. **12**: p. 1200-1208.

44. Janoff, A., et al., *Experimental Emphysema Induced with Purified Human Neutrophil Elastase: Tissue Localization of the Instilled Protease.* Am Rev Respir Dis, 1977. **116**: p. 461.

45. Sethi, S., *Bacterial Infection and the Pathogenesis of COPD.* CHEST, 2000. **117**: p. 286-291.

46. Takubo, Y., et al., *Alpha(1)-Antitrypsin Determines the Pattern of Emphysema and Function in Tobacco Smoke-Exposed Mice Parallels with Human Disease.* Am J Respir Crit Care Med, 2002. **166**: p. 1596-1603.

47. Atkinson, J. and R. Senior, *Matrix Metalloproteinase-9 in lung remodeling.* Am J Respir Cell Mol Biol, 2003. **28**: p. 12024.

48. Beasley, M.B., *Smoking-Related Small Airway Disease-A Review and Update.* Adv Anat Pathol, 2010. **17**(4): p. 270-276.

49. Bhatia, R., *Diseases of Small Airways of Lung.* Journal, Indian Academy of Clincial Medicine, 2001. **2**(3): p. 222-224.

50. Hogg, J.C., et al., *The Nature of Small-Airway Obstruction in Chronic Obstructive Pulmonary Disease.* N Engl J Med, 2004. **350**(26): p. 2645-2653.

51. Wright, J., et al., *The Detection of Small Airways Disease.* Am Rev Respir Dis, 1984. **129**(6): p. 989-994.

52. Cosio, M., et al., *The Relations between Structural Changes in Small Airways and Pulmonary-Function Tests.* N Engl J Med, 1978. **298**: p. 1277-1281.

53. Niewoehner, D.E., J. Kleinerman, and D.B. Rice, *Pathologic Changes in the Peripheral Airways of Young Cigarette Smokers.* N Engl J Med, 1974. **291**: p. 755-758.

54. Cosio, M., K. Hale, and D. Niewoehner, *Morphologic and Morphometric Effects of Prolonged Cigarette Smoking on the Small Airways* Am Rev Respir Dis, 1980. **122**(2): p. 265-21.

55. Hogg, J., P. Macklem, and W. Thurlbeck, *Site and Nature of Airway Obstruction in Chronic Obstructive Lung Disease.* N Engl J Med, 1968. **278**: p. 1355-1360.

56. Hogg, J., *Pathophysiology of Airflow Limitation in Chronic Obstructive Pulmonary Disease.* Lancet, 2004. **364**: p. 709-721.

57. Orie, N. and J. Sluiter, *The Host Factor in Bronchitis.* In: Bronchitis proceedings of the international symposium on bronchitis, 1960.

58. Barnes, P., *Against the Dutch Hypothesis: Asthma and Chronic Obstructive Pulmonary Disease are Distinct Diseases.* Am J Respir Crit Care Med, 2006. **174**: p. 240-243.

59. Hahn, D., *Beyond the Dutch Hypothesis.* Am J Respir Crit Care Med, 2006. **174**: p. 1056.

60. Silva, G., et al., *Asthma as a Risk Factor for COPD in a Longitudinal Study.* CHEST, 2004. **126**: p. 59-65.

61. Barnes, P., S. Shapiro, and R. Pauwels, *Chronic Obstructive Pulmonary Disease: Molecular and Cellular Mechanisms.* Eur Respir J., 2003. **22**: p. 672-688.

62. Thomson, N., R. Chaudhuri, and E. Livingston, *Asthma and Cigaretter Smoking.* Eur Respir J., 2004. **24**: p. 822-833.

63. James, A., et al., *Decline in Lung Function in the Busselton Health Study: The Effects of Asthma and Cigarette Smoking.* Am J Respir Crit Care Med, 2005. **171**: p. 109-114.

64. Stein, R., et al., *Influence of Parental Smoking on Respiratiry Symptoms during the First Decade of Life.* American Journal of Epidemiology, 1999. **149**(11): p. 1030-1037.

65. Marco, R.d., et al., *Risk Factors for Chronic Obstructive Pulmonary Disease in a European Cohort of Young Adults.* Am J Respir Crit Care Med, 2010: p. 1-44.

66. Postma, D., et al., *Asthma and COPD: Common Genes, Common Environments?* Am J Respir Crit Care Med, 2011: p. 34.

67. Simpson, A., et al., *Polymorphisms in A Disintegrin and Metalloprotease 33 (ADAM33) Predict Impaired Early-Life Lung Function.* Am J Respir Crit Care Med, 2005. **172**: p. 55-60.

68. Diemen, C.c., et al., *Novel Strategy to identify genetic risk factors for COPD severity: A Genetic Isolate.* Eur Respir J., 2009. **35**(4): p. 768-775.

69. Anthonisen, N., *The British Hypothesis Revisited.* Eur Respir J., 2004. **23**: p. 657-658.

70. Harris-Stuart, C., M. Pownall, and C. Scothorne, *The Factor of Infection in Chronic Bronchitis.* Q J Med, 1953. **22**: p. 121-132.

71. Kanner, R., N. Anthonisen, and J.C.f.t.L.H.S.R. Group, *Lower Respiratory Illnesses Promote FEV1 Decline in Current Smokers but not Ex-Smokers with Mild Chronic Obstructive Pulmonary Disease.* Am J Respir Crit Care Med, 2001. **164**: p. 358-364.

72. Vestbo, J., et al., *Association of Chronic Mucus Hypersecretion with FEV1 Decline and Chronic Obstructive Pulmonary Disease Morbidity.* Am J Respir Crit Care Med, 1996. **153**: p. 1530-1535.

73. Donaldson, G., et al., *Relationship Between Exacerbation Frequency and Lung Function Decline in Chronic Obstructive Pulmonary Disease.* Thorax, 2002. **57**: p. 847-852.

74. Wilkinson, T., et al., *Airway Bacterial Load and FEV1 Decline in Patients with Chronic Obstructive Pulmonary Disease.* Am J Respir Crit Care Med, 2003. **167**: p. 1090-1095.

75. Hurst, J., et al., *Susceptibility to Exacerbation in Chronic Obstructive Pulmonary Disease.* N Engl J Med, 2010. **363**: p. 1128-1138.

76. Hill, A., et al., *Association between Airway Bacterial Load and Markers of Airway Inflammation in Patients with Stable Chronic Bronchitis.* Am J Med., 1999. **109**: p. 288-295.

77. Sikkel, M.B., et al., *Respiratory Syncytial Virus Persistence in Chronic Obstructive Pulmonary Disease.* Pediatr Infect Dis J, 2008. **27**: p. 63-70.

78. McDowell, E., et al., *Abnormal Cilia in Human Bronchial Epithelium.* Arch Pathol Lab Med, 1976. **100**: p. 429-436.

79. Iravani, J. and G. Melville, *Long-Term Effect of Cigarette Smoke on Mucociliary Function in Animals.* Respiration, 1974. **31**: p. 358-366.

80. Read, R., R. Wilson, and A. Rutman, *Interaction of Nontypeable Haemophilus influenzae with Human Respiratory Mucosa In Vitro.* J Infect Dis, 1991. **163**: p. 549-558.

81. Vestbo, J., E. Prescott, and P. Lange, *Association of Chronic Mucus Hypersecretion with FEV1 Decline and Chronic Obstructive Pulmonary Disease Morbidity: Copenhagen City Heart Study Group.* Am J Respir Crit Care Med, 1996. **153**: p. 1530-1535.

82. Vestbo, J. and P. Lange, *Can GOLD Stage 0 Provide Information of Prognostice Value in Chronic Obstructive Pulmonary Disease?* Am J Respir Crit Care Med, 2002. **166**: p. 329-332.

83. Stefano, A.D., et al., *Airflow Limitation in Chronic Bronchitis is Associated with T-Lymphocytes and Macrophage Infiltration of the Bronchial Mucosa.* Am J Respir Crit Care Med, 1996. **153**: p. 629-632.

84. Bosken, C.H., et al., *Charcterization of the Inflammatory Reaction in the Peripheral Airways of Cigarette Smokers Using Immunohistochemistry.* Am Rev Respir Dis, 1992. **145**: p. 911-917.

85. Herr, C., et al., *Suppression of Pulmonary Innate Host Defence in Smokers.* Thorax, 2008. **64**: p. 144-149.

86. Stefano, A.D., A. Capelli, and M. Lusuardi, *Severity of Airflow Limitation is Associated with Severity of Airway Inflammation in Smokers.* Am J Respir Crit Care Med, 1998. **158**: p. 1277-1285.

87. Mantovani, A. and A. Sica, *Macrophage Polarization Comes of Age.* Immunity, 2005. **23**: p. 344-346.

88. Gordon, S., *Alternative Activation of Macrophages.* Nat. Rev. Immunol, 2003. **3**: p. 23-35.

89. Shapiro, S., *The Macrophage ins Chronic Obstructive Pulmonary Disease.* Am J Respir Crit Care Med, 1999. **160**: p. 29-32.

90. Turato, G., R. Zuin, and M. Saetta, *Pathogenesis and Pathology of COPD.* Respiration, 2001. **68**: p. 117-128.

91. Brightling, C., W. Monteiro, and R. Ward, *Sputum Eosinophilia and Short-Term Response to Prednisolone in Chronic Obstructive Pulmonary Disease: A Randomised Controlled Trial.* Lancet, 2000. **356**: p. 1480-1485.

92. Liu, H., et al., *Neutrophil Elastase and Elastase-rich Cystic Fibrosis Sputum Degranulate Human Eosinophils.* Am J physiol 1999. **276**: p. 28-34.

93. Junior, D., et al., *Immune System-Part II Basis of the Immunological Response Mediated by T and B Lymphocytes.* Bras J Rheumatol, 2010. **50**(5): p. 552-580.

94. Brucklacher-Waldert, V., et al., *Phenotypical Characterization of Human Th17 Cells Unambiguously Identified by Surface IL-17A Expression.* J Immunol, 2009. **183**: p. 5494-5501.

95. Boer, O.d., et al., *Immunhistochemcial Analysis of Regulatory T Cell Markers FOXP3 and GITR on CD4+CD25+ T Cells in Normal Skin and Inflammatory Dermatoses.* J Histochem Cytochem, 2007. **55**(9): p. 891-898.

96. Fowlkes, B., et al., *A Novel Population of T-Cell Receptor αβ-Bearing Thymocytes Which Predominantly Expresses a Single Vβ Gene Family.* Nature, 1987. **329**: p. 251-254.

97. Mizoguchi, A. and A. Bhan, *A Case for Regulatory B Cells.* J Immunol, 2006. **176**(2): p. 705-710.

98. Mizoguchi, A., et al., *Auppressive role of B Cells in Chronic Colitis of T Cell Receptor Alpha Mutant Mice.* J Exp Med, 1997. **186**(10): p. 1749-1756.

99. Brandsma, C.-A., et al., *Increased Levels of (class switched) Memory B cells in Peripheral Blood of Current Smokers.* Respiratory Research, 2009. **10**: p. 1-11.

100. Nierop, K.v. and C.d. Groot, *Human follicular dendritic cells: function, origin and development.* Seimars in Immunology, 2002. **14**: p. 251-256.

101. Kelsen, S.G., et al., *Lymphoid Follicle Cells in Chronic Obstructive Pulmonary Disease Overexpress the Chemokine Receptor CXCR3.* Am J Respir Crit Care Med, 2008. **179**: p. 799-805.

102. Kim, V., T.J. Rogers, and G.J. Criner, *New Concepts in the Pathobiology of Chronic Obstructive Pulmonary Disease.* Proc Am Thorac Soc, 2008. **5**: p. 478-485.

103. Weaver, C., et al., *IL-17 Family Cytokines and the Expanding Diversity of Effector T-Cells Lineages.* Annu Rev Immunol, 2007. **25**: p. 821-852.

104. Chen, Y., et al., *Stimulation of Airway Mucin Gene Expression by Interlukin (IL)-17 Through IL-6 Paracrin/Autocrine Loop.* J Biol Chem, 2003. **278**: p. 17036-17043.

105. Steinman, L., *A Brief History of Th17, the First Major Revision in the Th1/Th2 Hypothesis of T Cell-Mediated Tissue Damage.* Nature Medicine, 2007. **13** p. 129-145.

106. Sethi, S., et al., *Inflammatory Profile of New Bacterial Strain Exacerbations of Chronic Obstructive Pulmonary Disease.* Am J Respir Crit Care Med, 2007. **177**: p. 491-497.

107. Sethi, S., P. Mallia, and S.L. Johnston, *New Paradigms in the Pathogenesis of Chronic Obstructive Pulmonary Disease II.* Proc Am Thorac Soc, 2009. **6**: p. 532-534.

108. Matsuse, T., et al., *Latent Adenoviral Infection in the Pathogenesis of Chronic Ariways Obstruction.* Am Rev Respir Dis, 1992. **146**: p. 177-184.

109. Hogg, J.C., et al., *In situ Hybridization Studies of Adenoviral Infections of the Lung and Their Relationship to Follicular Bronchiectasis.* Am Rev Respir Dis, 1988. **139**: p. 1531-1535.

110. Wood, A., et al., *Smoke Exposure as a determinant of auto-antibody titre in AATD and COPD.* Eur Respir J., 2010.

111. Greene, C., et al., *Anti-proline-glycine-proline or Antielastin Autoantibodies are not evident in Chronic Inflammatory Lung Disease.* Am J Respir Crit Care Med, 2010. **181**(1): p. 31-35.

112. Rodriguez-Roisin, R., *Toward a Consensu Definition for COPD Exacerbations.* CHEST, 2000. **117**: p. 398-401.

113. Mannino, D., et al., *Chronic Obstructive Pulmonary Disease Surveillance-United States, 1971-2000.* Surveill Summ, 2002. **51**: p. 1-16.

114. Miravitlles, M., et al., *Effect of Exacerbations on Quality of Life in Patients with Chronic Obstructive Pulmonary Disease: A 2 Year Follow-up Study.* Thorax, 2004. **62**: p. 387-395.

115. Anthonisen, N., et al., *Antibiotic Therapy in Exacerbations of Chronic Obstructive Pulmonary Disease.* Ann Intern Med, 1987. **106**: p. 196-204.

116. Curran, T., et al., *Evaluation of Real-Time PCR for the Detection and Quantification of Bacteria in Chronic Obstructive Pulmonary Disease.* FEMS Immunol Med Microbiol, 2007. **1**: p. 112-118.

117. Serres, G.D., et al., *Importance of Viral and Bacterial Infections in Chronic Obstructive Pulmonary Disease Exacerbations.* Journal of Clinical Virology, 2009. **46**: p. 129-133.

118. Anderson, H., et al., *Air Pollution and Daily Admissions for Chronic Obstructive Pulmonary Disease in 6 European Cities: Results from the APHEA Project.* Eur Respir J., 1997. **10**: p. 1064-1071.

119. Nordenhall, C., et al., *Airway Inflammation Following Exposure to Diesel Exhaust: A Study of Time Kinetics Using Induced Sputum.* Eur Respir J., 2000. **15**: p. 1046-1051.

120. Rudell, B., et al., *Bronchoalveolar Inflammation After Exposure to Diesel Exhaust: Comparison Between Unfiltered and Particle Trap Filtered Exhaust.* Occup Environ Med, 1999. **56**: p. 527-534.

121. Arbex, M., et al., *Urban Air Pollution and Chronic Obstructive Pulmonary Disease-Related Emergency Department Visits.* J Epidemiol Community Health, 2009. **63**: p. 777-783.

122. Mohan, A., et al., *Prevalence of Viral Infection Detected by PCR and RT-PCR in Patients with Acute Exacerbation of COPD: A systematic Review.* Respirology, 2010. **15**: p. 536-542.

123. Kherad, O., et al., *Upper-Respiratory Viral Infection, Biomarkers, and COPD Exacerbations.* CHEST, 2010. **138**(4): p. 896-904.

124. Jr, C.C., et al., *Viral Pathogens in Acute Exacerbations of Chronic Obstructive Pulmonary Disease.* Intern Emerg Med, 2008. **3**: p. 355-359.

125. Ko, F., M. Ip, and P. Chan, *Viral etiology of Acute Exacerbations of COPD in Hong Kong.* CHEST, 2007. **132**: p. 900-908.

126. Zakharkina, T., et al., *Detection of Microorganisms in Exhaled Breath Condensate During Acute Exacerbations of COPD.* Respirology, 2011.

127. Morimoto, K., et al., *Adenovirus E1A Regulates Lung Epithelial ICAM-1 Expression by Interacting with Transcriptional Regulators at its Promoter.* Am J Physiol Lung Cell Mol Physiol, 2008. **296**: p. 361-371.

128. Verma, P., *Laboratory Diagnosis of Anaerobic Pleuropulmonary Infections.* Seminars in Respiratory Infections, 1986. **15**(2): p. 114-118.

129. Finegold, S. and J. Bartlett, *Anaerobic Pleuropulmonary Infections.* Clinical Considerations, 1975. **42**(1): p. 101-111.

130. Sethi, S., *Coinfection in Exacerbations of COPD.* CHEST, 2006. **129**(2): p. 223-224.

131. Ishige, I., et al., *Propionibacterium acnes is the Most Common Bacterium commenal in Peripheral Lung Tissue and Mediastinal Lymph Nodes from Subjects Without Sarcoidosis.* Sarcoidosis Vasculitis and Diffuse Lung Diseases, 2005. **22**: p. 33-42.

132. Martinez-Solano, L., et al., *Chronic Pseudominas aeruginosa Infection in Chronic Obstructive Pulmonary Disease.* Clinical Infectious Disease, 2008. **47**: p. 1526-1531.

133. Ko, F., et al., *A One-Year Prospective Study of Infectious Etiology in Patients Hospitalized with Acute Exacerbations of COPD and Concomitant Pneumonia.* Respiratory Medicine, 2008. **102**: p. 1109-1116.

134. Miravitlles, M., *Do we Need New Antibiotics for Treating Exacerbations of COPD?* Therapeutic Advances in Respiratory Disease, 2007. **1**(1): p. 61-76.

135. Jorgensen, J., et al., *Antimicrobial Resistance among Respiratory Isolates of Haemophilus influenzae, Moraxella catarrhalis, and Streptococcus pneumoniae in the United States.* Antimicrob Agents Chemother., 1990. **34**(11): p. 2075-2080.

136. Doern, G., *Trends in Antimicrobial Susceptibility of Bacterial Pathogens of the Respiratory Tract.* Am J Med., 1995. **99**(6B): p. 3-7.

137. Saint, S., et al., *Antibiotics in Chronic Obstructive Pulmonary Disease Exacerbations. A Meta-Analysis.* JAMA, 1995. **274**(14): p. 1131-1132.

138. Puhan, M., et al., *Exacerbations of Chronic Obstructive Pulmonary Disease: When are Antibiotics Indicated? A Systematice Review.* Respir Res., 2007. **8**(1): p. 1-11.

139. Nouira, S., et al., *Once Daily Oral Ofloxacin in Chronic Obstructive Pulmonary Disease Exacerbation Requiring Mechanical Ventilation: A Randomised Placebo-Controlled Trial.* Lancet, 2001. **358**(9298): p. 2020-2025.

140. Banerjee, D., O. Khair, and D. Honeybourne, *Impact of Sputum Bacteria on Airway Inflammation and Health Status in Clincial Stable COPD.* Eur Respir J., 2004. **23**: p. 685-691.

141. Turnbaugh, P., et al., *The Human Microbiome Project.* Nature, 2007. **449**: p. 804-810.

142. Dinsdale, E., et al., *Functional Metagenomic Profiling of Nine Biomes.* Nature, 2008. **452**: p. 629-632.

143. Pennisi, E., *Body's Hardworking Microbes Get Some Overdue Respect.* Science, 2010. **330**: p. 1619.

144. Palmer, C., et al., *Development of the Human Infant Intestinal Microbiota.* PLOS Biology, 2007. **5**(7): p. 1556-1573.

145. Ley, R., et al., *Microbial Ecology: Human Gut Microbes Associated with Obesity.* Nature, 2006. **444**: p. 1022-1023.

146. Lay, R., et al., *Microbial Ecology: Human Gut Microbes Associated with Obesity.* Nature, 2006. **444**: p. 1022-1023.

147. Benson, A., et al., *Individuality in Gut Microbiota Composition a Complex Polygenic Trait Shaped by Multiple Environmental and Host Genetic Factors.* PNAS, 2010: p. 1-6.

148. Hubbell, S., *Neutral Theory and the Evolution of Ecological Equivalence.* Ecology, 2006. **87**: p. 1387-1398.

149. Antonopoulos, D., et al., *Reproducible Community Dynamics of the Gastrointestinal Microbiota Following Antibiotic Perturbation.* Infection and Immunity, 2009. **77**(6): p. 2367-2375.

150. Turnbaugh, P., et al., *An Obesity-Associated Gut microbiome with Increased Capacity for Energy Harvest.* Nature, 2006. **444**: p. 1027-1031.

151. Costello, E., et al., *Bacterial Community Variation in Human Body Habitats Across Space and Time.* Science, 2009. **326**: p. 1694-1697.

152. Hummelen, R., et al., *Deep Sequencing of the Vaginal Microbiota of Women with HIV.* PLOS One, 2010. **5**(8): p. 1-9.

153. Ling, Z., et al., *Molecular Analysis of the Diversity of Vaginal Microbiota Associated with Bacterial Vaginosis.* BMC Genomics, 2010. **11**: p. 1-16.

154. Avila, M., D. Ojcius, and O. Yilmanz, *The Oral Microbiota: Living with a Permanent Guest.* DNA and Cell Biology, 2009. **28**: p. 405-411.

155. Frank, D., et al., *The Human Nasal Microbiota and Staphylococcus aureus Carriage.* PLOS One, 2010. **5**(5): p. 1-15.

156. Chilvers, M. and C. O'Callaghan, *Local Mucociliary Defence Mechanisms.* Paediatric Respiratory Reviews, 2000. **1**: p. 27-34.

157. Cox, M., et al., *Airway Microbiota and Pathogen Abundance in Age-Stratified Cystic Fibrosis Patients.* PLOS One, 2010. **5**(6): p. 1-10.

158. Hilty, M., et al., *Disordered Microbial Communities in Asthmatic Airways.* PLOS One, 2010. **5**(1): p. 1-9.

159. Huang, Y., et al., *A Persistent and Diverse Airway Microbiota Present during Chronic Obstructive Pulmonary Disease Exacerbations.* OMICS A Journal of Integrative Biology, 2010. **14**(1): p. 9-16.

160. Grice, E., et al., *A Diversity Profile of the Human Skin Microbiota.* Genome Research, 2008: p. 1-8.

161. Lynch, S., et al., *Diverse Bacterial Communities in the Airways of Patients with Sub-Optimally Controlled Asthma,* in *American Thoracic Society*2010: New Orleans.

162. Harris, J., et al., *Variability of the Microbiome in Bronchoalveolar Lavage versus Induced Sputum of Asthmatics* in *American Thoracic Society*2010: New Orleans.

163. Erb-Downward, J., et al., *Analysis of the Microbiome of the Normal and COPD Lung,* in *American Thoracic Society*2010: New Orleans.

164. Erb-Downward, J., et al., *Analysis of the Lung Microbiome in the "Healthy" Smoker and in COPD.* PLOS One, 2011. **6**(2): p. 1-12.

165. Brodie, E., et al., *Urban Aersols Harbor Diverse and Dynamic Bacterial Populations.* PNAS, 2007. **104**(1): p. 299-304.

166. Tringe, S., et al., *The Airborne Metagenome in an Indoor Urban Environment.* PLOS One, 2008. **3**(4): p. 1-10.

167. Doring, G., I. Parameswaran, and T. Murphy, *Differential Adaptation of Microbial Pathogens to Airways of PAtients with Cystic Fibrosis and Chronic Obstructive Pulmonary Disease.* FEMS Microbial Rev, 2010: p. 1-23.

168. Pang, H. and H. Winkler, *Copy Number of the 16S rRNA Gene in Rickettsia prowazekii.* Journal of Bacteriology, 1993. **175**(12): p. 3893-3896.

169. Neefs, J.-M., et al., *Compiliation of Small Ribosomal Subunit RNA Sequences.* Nucleic Acids Research, 1990. **18**: p. 2237-2317.

170. Fedurco, M., et al., *BTA, a Novel Reagent for DNA Attachment on Glass and Efficient Generation of Solid-Phase Amplified DNA Colonies.* Nucleic Acids Research, 2006. **34**(3): p. 1-13.

171. Tucker, T., M. Marra, and J. Friedman, *Massively Parallel Sequencing: The Next Big Thing in Genetic Medicine.* The American Journal of Human Genetics, 2009. **85**: p. 142-154.

172. Shendure, J. and H. Ji, *Next-Generation DNA Sequencing.* Nature Biotechnology, 2008. **26**(10): p. 1135-1145.

173. Bentley, D., et al., *Accurate Whole Human Genome Sequencing Using Reversible Terminator Chemistry.* Nature, 2008. **456**(7218): p. 53-59.

174. Turcatti, G., et al., *A New Class of Cleavable Fluorescent Nucleotides: Synthesis and Optimization as Reversible Terminators for DNA Sequencing by Synthesis.* Nucleic Acids Research, 2008. **36**(4): p. 1-13.

175. Zaikova, E., et al., *Microbial Community Dynamics in s Seasonally Anoxic Fjord: Saanich Inlet, British Columbia.* Environmental Microbiology, 2010. **12**(1): p. 172-191.

176. Sacchi, C., et al., *Sequencing of 16S rRNA Gene: A Rapid Tool for Identification of Bacillus anthracis.* Emerging Infectious Diseases, 2002. **8**(10): p. 1117-1123.

177. McNees, A., et al., *Specific and Quantitative Detection of Human Polyomaviruses BKV, JCV, and SV40 by Real Time PCR.* Journal of Clinical Virology, 2005. **34**: p. 52-62.

178. Liu, W.-T., et al., *Characterization of Microbial Diversity by Determining Terminal Restriction Fragment Length Polymorphisms of Genes Encoding 16S rRNA.* Applied and Environmental Microbiology, 1997. **63**(11): p. 4516-4522.

179. Schloss, P., et al., *Introducing Mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Mircobial Communities.* Appl Environ Microbiol, 2009. **75**(23): p. 7537-7541.

180. Liu, Z., et al., *Accurate Taxonomy Assignments From 16S rRNA Sequences Produced by Highly Parallel Pyrosequencers.* Nucleic Acids Research, 2008. **36**: p. e120.

181. Faultz, J., *Effects of Shelterwood Management on Flower-Visiting Insects and Their Floral Resources*, in *Entomology*2005, Montana State: Bozeman. p. 161.

182. Flint, H., et al., *Interactions and Competition within the Microbial Community of the Human Colon: Links Between Diet and Health.* Environmental Microbiology, 2007. **9**(5): p. 1101-1111.

183. Calderon, K., et al., *Analysis of Microbial Communities developed on the Fouling Layers of a Membrane-Couples Anaerobic Bioreactor Applied to Wastewater Treatment.* Bioresource Technology, 2011.

184. Bettazzi, E., et al., *Nitrite Inhibition and Intermediates Effects on Anammox Bacteria: A Batch-Scale Experimental Study.* Process Biochemistry, 2010. **45**: p. 573-580.

185. Clement, B., et al., *Terminal Restriction Fragment PAtterms (TRFPs), a Rapid, PCR-BAsed Method for the Comparison of Complex Bacterial Communities.* Journal of Microbiological Methods, 1998. **31**: p. 135-142.

186. Wintzingerode, F., U. Gobel, and E. Stackebrandt, *Determination of Microbial Diversity in Environmental Samples: Pitfalls of PCR-Based rRNA Analysis.* FEMS Microbial Rev, 1997. **21**: p. 213-229.

187. Osborn, A., E. Moore, and K. Timmis, *An Evaluation of Terminal-Restriction Fragment Length Polymorphism (T-RFLP) analysis for the Study of Microbial Community Structure and Dynamics.* Environmental Microbiology, 2000. **2**(1): p. 39-50.

188. Egert, M. and M. Friedrich, *Formation oof Pseudo-Terminal Restriction Fragments, a PCR-Related Bias Affecting Terminal Restriction Fragment Length Polymorphism Analysis of Microbial Community Structure.* Applied and Environmental Microbiology, 2003. **69**(5): p. 2555-2562.