Effect of Maintenance Antibiotic Therapy on the Cystic Fibrosis Sputum Microbiome

Maria Therese Nelson

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2020

Reading Committee:

Lucas R. Hoffman, Chair

Stephen Salipante

David Fredricks

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

Effect of Cycled Antibiotic Therapy on the Cystic Fibrosis Respiratory Microbiome

Maria Therese Nelson

Chair of the Supervisory Committee:
Dr. Lucas R. Hoffman, MD PhD
Professor, Department of Pediatrics
Adjunct Professor, Department of Microbiology

Cystic fibrosis (CF) is a genetic disease resulting in chronic, complex, polymicrobial respiratory infections, ultimately contributing to lung destruction and death. Despite extensive study, primarily using culture methods, the microbial determinants of clinical status and therapeutic response have yet to be completely elucidated. The most common antibiotic used to treat people with CF is inhaled tobramycin, administered as maintenance therapy for chronic *Pseudomonas aeruginosa* lung infections. While the effects of inhaled tobramycin on *P. aeruginosa* abundance and lung function diminish with continued therapy, this maintenance treatment improves long-term CF morbidity and mortality, underscoring how little is known about why antibiotics work in CF infections, their effects on complex CF sputum microbiomes, and how to improve these treatments.

Metagenomic sequencing is a promising approach for identifying and characterizing organisms and their functional characteristics in complex, polymicrobial infections, such as airway infections in people with CF. These analyses are often hampered, however, by overwhelming quantities of human DNA, yielding only a small proportion of microbial reads for analysis. Additionally, many abundant microbes in respiratory samples can produce large

quantities of extracellular bacterial DNA originating either from biofilms or dead cells. We first describe a method for simultaneously depleting DNA from intact human cells and extracellular DNA (human and bacterial) in sputum, using selective lysis of eukaryotic cells and endonuclease digestion. We show that this method increases microbial sequencing depth and, consequently, both the number of taxa detected and coverage of individual genes such as those involved in antibiotic resistance, underscoring the substantial impact of DNA from sources other than live bacteria in microbiological analyses of complex, chronic infection specimens.

We then collected sputum from 30 individuals with CF at standardized time points before, during and after a month-long course of maintenance inhaled tobramycin. We used culture, quantitative PCR and metagenomic sequencing to define the dynamic effects of inhaled tobramycin on sputum microbiomes, including abundance changes in both clinically targeted and untargeted bacteria, and in functional gene categories. CF sputum microbiota changed most markedly by one week of therapy and plateaued thereafter. This shift was largely driven by changes in non-dominant taxa. Genetically conferred functional capacities (metagenomes) of subjects' sputum communities changed little with antibiotic perturbation, despite taxonomic shifts, suggesting functional redundancy within the CF sputum microbiome. What changes did occur were primarily driven by non-dominant taxa. These results highlight the importance of monitoring the broad microbiological effects of antibiotics and other treatments as we work towards accurately defining and improving the clinical impact of these therapies.

Dedicated to my parents, the original Drs. Nelson

Acknowledgements

I would like to extend my deepest gratitude to all individuals that have not only made this work possible but have contributed to my intellectual growth up to this point.

First and foremost, I would like to thank all the individuals who participated in this study and all people living with cystic fibrosis and their families. Thirty of you generously shared your time, thoughts, symptoms and samples with us and tolerated a lot of nagging to ensure the strength of our sample set. To be living with CF in and around Seattle is likely to be involved in a number of studies, ostensibly on faith that it will benefit everyone living with CF. I am grateful for their participation and I hope this work may someday live up to that faith.

I would like to thank all the study coordinators who recruited participants and collected samples: Laura Nay, Sharon McNamara, Cheryl Majors and Ellen Willhelm. Without your kindness, dedication and persistence this study would not have been possible. Without your willingness to answer 11th hour study questions I would not have been able to submit my F30. By extension, I would like to thank the entire CF research community, both within the US and internationally: this community is a remarkable and educational example of what scientific collaboration can achieve. I have specifically benefitted from the mentorship of John LiPuma, DB Sanders, Pradeep Singh, Ajai Dandekar, Matt Parsek, Chris Goss and the entire Seattle Children's Hospital pulmonary department, both directly by critiquing by written work and oral presentations and indirectly by simply sharing their scientific and clinical expertise, all with kindness. In particular, I would like to thank all the amazing women in CF research, past and present: you are a constant inspiration and who I strive to be.

To all members of the Hoffman Lab, past and present: you have provided the best graduate school experience I could have hoped for. Thank you to Mimi Precit and Catherine Armbruster for providing an example to follow and helping me when I first started in the lab. Thanks to Lauren Gonsalves for working with me now and putting up with my disorganization; I hope I can be for you what Mimi and Catherine were to me. Thanks to Robyn Marsh for being a fantastic mentor, personal and professional, and a great friend. And for teaching me how to be deliberate and thoughtful in my science. Thanks to Dan

and Chris: the older brothers of the lab. You were my on-the-ground mentors and my graduate education is due to you two as much as to anyone else. Thank you for sharing your expertise with me, for being willing to answer questions and provide guidance even with your busy schedules. Thank you for respecting my opinion and scientific work. It was this belief that allowed be to grow over the last five years. And of course, thanks to Dr. Luke Hoffman. I have learned so much from you about mentorship, the importance of communicating your science to all audiences, and how to conduct patient centered research. I am eternally grateful for the opportunity to work with you and to learn not only science but how to be a scientist.

Finally, thank you to Sam Welker for doing the most even when you thought you did the least.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

“The child will soon die whose brow tastes salty when kissed” [1]

-Almanac of Children’s Games and Songs of Switzerland, 1857

Cystic Fibrosis: history and disease manifestation

Today, cystic fibrosis (CF) is the most common life-shortening recessive genetic disorder in individuals with European ancestry, occurring among approximately 1:3,000 live births in the United States [2]. CF disease results from mutations in the widely expressed cystic fibrosis transmembrane regulator (CFTR), which regulates ion transport across cellular membranes in multiple organ systems, with the respiratory and gastrointestinal tracts two of the most prominently affected. Although cystic fibrosis (CF) is now often thought of primarily as a respiratory disease, it was initially thought to be a form of celiac disease, underscoring the importance of the nutrition and gastrointestinal consequences of CF. There are numerous historical accounts of a childhood wasting disease characterized by fat malabsorption and intestinal obstruction as well as folk accounts of "cursed" children with salty skin [1]. Dr. Dorothy Andersen was the first to describe a separate condition she termed "cystic fibrosis of the pancreas" based on pathological evidence of cystic ducts in the pancreata of affected children in 1938 [3]. Andersen later described a number of cases of infants who died of bronchiolitis, but upon autopsy were found to have cystic pancreatic ducts, further defining cystic fibrosis of the pancreas as a unique clinical entity with accompanying respiratory complications [4]. Andersen believed the extra-gastrointestinal manifestations of CF to be due to vitamin A malabsorption secondary to exocrine pancreatic insufficiency. It wasn't until 1946 that the recessive, Mendelian nature of what is now known as CF was described [5]. Later, the clinical recognition of high sodium chloride in the sweat of infants with CF enabled the early diagnosis of "pancreatic-sufficient" children with the disease [6,7], cementing this characterization as a unique clinical entity, facilitating research into early stages of the disease and allowing clinicians to begin treatment in the vital early life period even in the absence of symptoms.

The 1940s also marked the advent of another landmark treatment which would revolutionize CF care: the antibiotic penicillin. Early reports of *Staphylococcus* in the lungs of children with cystic fibrosis of the pancreas sparked the development of aerosolized formulations of penicillin [8]. It was quickly recognized that early diagnosis and treatment was vital in the clinical care of infants with CF, and through the 1940s and 1950s aggressive maintenance antibiotics, as well as a number of nutritional supplements and respiratory physiotherapies, were introduced as part of CF clinical care and began to increase the survival of

affected children, some into their teens. Antibiotic therapy primarily comprised anti-*Staphylococcal* therapy, the taxon most commonly seen in the respiratory tracts of infants with CF at the time. Even as early as 1951, reports began to surface of an emerging pathogen in CF that arose after, seemingly in response to, such aggressive antibiotic therapy: *Pseudomonas aeruginosa* [9,10]. It wasn't until the 1970s, however, that *P. aeruginosa* came to be considered a serious pathogen in CF associated with worse clinical outcomes [11,12]. While *Staphylococcus* had been the primary pathogen identified in early CF research and clinical care, by the 1980s *P. aeruginosa* was recognized as the dominant CF pathogen, and aggressive antibiotic treatments often targeted at this taxon [13]. Morbidity and mortality outcomes for individuals with CF continued to improve over the next decades due to the development of a number of medical, mechanical and surgical interventions to address sequalae of CF, such as aggressive antibiotic therapy, improvements in respiratory pharmacotherapy and physiotherapy, nutritional support, and heart-lung transplantation for those with end stage disease. Within the last decade, novel drugs have been introduced that address the CFTR defect itself [14,15] which have resulted in sustained improvements in a number of clinical metrics, with presumed (and highly anticipated) benefit on mortality outcomes [16,17]. The first of these drugs to be introduced, ivacaftor, has been shown to reduce *P. aeruginosa* respiratory colonization rates [18] and is associated with improvements in inflammatory markers [19]. Despite these revolutionary advancements, respiratory infections still remain the leading cause of morbidity and mortality in people with CF [2,20,21]. Infections often persist even after CFTR function is partially restored [19], and *P. aeruginosa* infection has been shown to rapidly reestablish after lung transplantation [22]. Therefore, despite decades of awareness, research and clinical care, respiratory infections remain a pressing problem in CF.

Infections in cystic fibrosis: from culture to sequencing

In the lungs of individuals with CF, aberrant fluid secretion results in buildup of thick, intractable mucus that is prone to microbial colonization, resulting in chronic, progressive respiratory infections. These polymicrobial infections require frequent antibiotic treatments for both maintenance and periodic pulmonary exacerbations, which are associated with reduced quality of life and long-term clinical decline [23,24]. Antibiotic treatments are both common and effective for CF lung disease. While CF respiratory infections contribute considerably to reduced quality and length of life [23,24], neither the microbial determinants of CF respiratory disease

severity nor of response to antibiotic treatment are well understood. Current therapies in CF generally target culture-identifiable organisms, but CF lungs remain persistently infected with these "standard pathogens" throughout patients' lifetimes despite frequent antibiotic treatments. Traditionally, these infections have been diagnosed and treated using standard clinical culturing methods, which generally identify only the small number of classic CF pathogens that are effectively detected by these methods (organisms that grow quickly and aerobically on laboratory media, e.g.), such as *S. aureus* and *P. aeruginosa* [25,26]. Transient respiratory microbiological changes with antibiotic therapy identified by culture are often most pronounced among aerobic taxa, most notably in *P. aeruginosa* [27,28], a common target of antibiotics in CF. However, while one might predict that antibiotic-induced reductions in sputum abundances of "pathogenic" bacteria (such as *P. aeruginosa*) are responsible for clinical improvement, previous studies have not supported this hypothesis [29,30], usually identifying minimal and/or transient abundance changes during treatment or exacerbations that do not consistently correlate with changes in either symptoms, lung function or *in vitro* clinical susceptibility testing [31,32]. Interestingly, there is little evidence that *P. aeruginosa* density changes appreciably during exacerbation [33], and the degree to which antibiotics reduce *P. aeruginosa* levels following treatment has not been correlated with therapeutic success [34,35]. Thus, classic clinical culture has not fully explained the complexity of CF respiratory microbiology or its relation to clinical outcomes.

Recently, wider availability of next generation sequencing has facilitated more nuanced and comprehensive investigations of these infections. These culture-free methods have shown the microbiota (the identities and relative abundances of the full complement of bacterial taxa present) in CF respiratory samples to be more diverse than previously thought, often comprising species not detected by routine clinical culture [36–38]. These newer techniques also suggest CF sputum microbiota are relatively stable despite clinical change [39–41]. Furthermore, the sputum microbiota appeared to be resilient, returning to their baseline states after treatment ceased [39–44]. Longitudinal studies have also demonstrated inter-patient heterogeneity in CF respiratory microbiota to be higher than intra-patient heterogeneity and that there is substantial variation in the rate of microbiota change within patients [39,42].

While stability and resilience are core features of the CF respiratory microbiome, both cross-sectional and long-term longitudinal studies have identified a number of host and therapeutic factors that

are associated with differences and changes in microbiota, many of which vary over time not only within individuals, but also in CF patient populations, as care evolves [45]. Studies have demonstrated that, on average, CF respiratory microbial communities tend to be dominated by oral anaerobes early in life followed by gradual decreases in alpha diversity due to increasing abundances (relative and absolute) of "traditional" CF pathogens [46–49]. These changes are associated both with decreasing lung function and also increasing antibiotic exposure [50,51]. Treatment with novel CFTR modulators, which treat the underlying defect in CF, was recently shown to be followed by an initial decrease in absolute and relative abundances of *P. aeruginosa* and with an accompanying increase in relative abundances of *Streptococcus* and *Prevotella spp*. These changes resulted in increased overall alpha-diversity (taxonomic diversity in a single given samples), all of which accompanied a concurrent increase in lung function [52]. Interestingly, baseline alpha-diversity and worse disease state also appear to predict the degree of community perturbation at exacerbation [53,54] and decreased microbial diversity early in life has also been associated with a faster rate of lung function decline [55]. The onset of pulmonary exacerbation has specifically been associated with transient changes in relative abundance of a number of taxa, such as anaerobes [53] and *Gemella* [54]. However, these changes in sputum microbiota do not necessarily signify the mechanism of the clinical effects of antibiotics [56,57]: changes in absolute bacterial abundance are associated with neither the symptomatic onset of exacerbation nor response to treatment on average [58,59]. Conversely, the presence in CF sputum of diverse microbiota diversity and presence of microbes (predominantly anaerobes) also typically found in healthy lungs [59] has also been associated with better clinical status [60,61], but the mechanisms of these associations and their roles in therapeutic responses are unclear. For example, higher sputum microbial diversity may signify lower historical antibiotic burden due to milder disease; alternatively, higher diversity could indicate low absolute abundances of pathogens that only lead to worse disease when their abundances rise.

Thus, despite these advances in understanding the microbial correlates of CF lung disease, studies of the CF sputum microbial community using these newer methods have thus far failed to find consistent predictors of either therapeutic success or clinical status. While a number of studies of suggested associations between aspects of the CF respiratory microbiota and therapeutic outcomes or clinical states, the direct therapeutic or clinical factors contributing to these effects are unclear. Prior studies may

not have been able to adequately identify the microbial determinants of clinical status because they were limited by studying diverse treatment regimens, and by their retrospective approaches: study samples are typically from individuals on different antibiotics and collected at different time points, making controlling for confounding variables challenging, particularly when considering the relatively low population prevalence of CF (limiting sample size and power). Furthermore, microbial communities in CF sputum can differ dramatically between individuals with similar clinical characteristics. These observations, together with the diagnostic imprecision of routine clinical culture, make it difficult to infer which taxa are most responsible for clinical status or response to treatment. Thus, a deeper understanding of sputum microbial community constituency than that provided by current methods could help to determine mechanisms by which microorganisms persist, and how they may be more effectively treated. Finally, while most CF microbiome research has focused on the microbial taxa ("microbiota") in sputum, the full complements of CF sputum microbial genes ("metagenomes") have not been thoroughly defined.

Tobramycin use in CF

Tobramycin is an aminoglycoside, a class of drugs that act by inhibiting protein translation by preventing formation of a complete ribosomal complex [63]. All aminoglycosides are actively taken up by cells [64], a process requiring working electron transport chains that tend to be most active under aerobic conditions; therefore anaerobes are considered to be inherently relatively resistant to antibiotics in this class. However, there are reports *in vitro* of aminoglycoside activity against anaerobic taxa [65–67]. Aerobic taxa have also been shown to modulate uptake of aminoglycosides based on carbon source [68] in addition to enzymatic modification of the drug and efflux pumps as mechanisms of resistance [63].

Intravenous (IV) aminoglycoside therapy, primarily gentamicin, has been a component of CF therapy since the 1970s after the emergence of *P. aeruginosa* as a dominant CF respiratory pathogen [69,70]. IV tobramycin, an aminoglycoside-class antibiotic with fewer side effects [71], was subsequently demonstrated to be non-inferior to gentamicin *in vivo* [72] and superior *in vitro* [73] and subsequently became standard of care in the US. Inhaled formulations of tobramycin have been used as anti-*Pseudomonal* therapy in CF since the 1980s [74,75]. Today tobramycin is the most commonly prescribed

antibiotic for CF pulmonary infections in the US, with approximately 70% of patients prescribed this drug for chronic use [2]. While IV tobramycin is still used during pulmonary exacerbation, inhaled tobramycin is typically used as maintenance therapy by people with CF chronically infected with *P. aeruginosa*. The standard inhaled tobramycin formulation was designed, tested and approved in alternating one-month-on, one-month-off cycles [76,77], a regimen demonstrated to reduce long-term exacerbation rates and lung function decline [76,78,79]. While the short-term effects of tobramycin in patients new to this therapy include a significant reduction in *P. aeruginosa* burden in respiratory secretions and improved respiratory function, both of these effects are attenuated with each subsequent on-off cycle of drug treatment [76], and therapeutic response can differ between patients regardless of *in vitro* susceptibility of infecting *P. aeruginosa* to tobramycin [76,80–82]. Furthermore, some individuals fail to respond or worsen over the first treatment period, and even more do so during a follow-up maintenance cycle. This discrepancy between waning short-term effectiveness and robust long-term therapeutic benefit suggests that factors in addition to canonical anti-*Pseudomonal* action contribute to clinical response and resistance, such as off-target effects against other taxa. Although tobramycin has been shown to have *in vitro* activity against many common CF pathogens, including gram-negative organisms other than *P. aeruginosa* [63] and *Staphylococcus aureus*, its effect on the entire sputum microbial community has not previously been assessed *in vivo*. Whether clinical response to tobramycin therapy is mediated through modulation of standard CF pathogens, non-canonical pathogenic species or “healthy” respiratory bacteria is unclear. Aminoglycoside antibiotics have even been shown to directly increase levels of functional CFTR protein *in vitro* by enhancing read-through of stop codons, raising the possibility that tobramycin may even be improving clinical outcomes by direct action on defective human proteins [83,84].

The change in the genetic composition of the community (the “metagenome”), which reflects microbial community function in addition to membership, in response to antibiotic perturbation has not been studied in the context of a standard treatment regimen. Such efforts are complicated by complex nature of these respiratory samples: CF sputum is dominated by host DNA due to chronic inflammation, resulting in frequent turnover of immune and lung cells. Furthermore, the microbial communities in CF sputum undergo frequent bacterial cell turnover (due to antibiotics, immunity, and inter-species competition), and many members are known to be prolific biofilm producers that produce large quantities of extracellular

DNA [85]. Consequently, the majority of DNA in CF sputum is human and extracellular microbial DNA that, if not excluded, can result in a biased assessment of community structure [86–89]. Extracellular bacterial DNA alone has been shown previously to bias determination of sputum community structure, particularly after antibiotic therapy, which can result in increased bacterial cell death and lead to inflated estimates of resistance to therapy [89]. Furthermore, the overwhelming abundance of host DNA (95% on average) can make metagenomic sequencing of large sample sets to adequate depth cost-prohibitive and technically difficult. Thus, most previous studies investigating the sputum metagenome have suffered from very small sample sizes and single time points [90–92], making identifying functional correlates of clinical response difficult.

Summary and Objectives

We hypothesize that therapeutic response to tobramycin correlates with features of both the microbiota and metagenome (the community members and their functional capacity) in sputum at baseline and change in these features across therapy, and that defining these microbial factors will help to identify the mechanisms by which chronic infections in CF persist despite chronic antibiotic exposure. Furthermore, we hypothesize that a standardized treatment, collection, sample processing and data analysis protocol will help to control for clinical and microbiological heterogeneity between individuals and allow for identification of the effect of such a chronic therapy on the CF respiratory microbial community and that depletion of human and extracellular bacterial DNA will allow for sufficient microbial sequencing coverage to adequately profile the change in these microbial communities in response to such a perturbation. In chapter 2 I discuss the development and optimization of a method for metagenomic sequencing of complex clinical samples such as CF sputa. I then apply this refined method to a collection of CF sputa collected before, weekly during and after a standard, one-month course of daily tobramycin and detail the drugs effect on the taxonomic (chapter 3) and genetic functional constituency (chapter 4) of the community. Finally, in chapter 5 I discuss ramifications and future directions of this work.

Chapter 2

Depletion of human and extracellular bacterial DNA for optimized microbiome profiles

The majority of the work presented in this chapter is published as:

Nelson MT, Pope CE, Marsh RL, Wolter DJ, Weiss EJ, Hager KR, Vo AT, Brittnacher MJ, Radey MC, Hayden HS, Eng A, Miller SI, Borenstein E, Hoffman LR. Human and Extracellular DNA Depletion for Metagenomic Analysis of Complex Clinical Infection Samples Yields Optimized Viable Microbiome Profiles. C*ell report*s. 2019 Feb 19;26(8):2227-40. doi: 10.1016/j.celrep.2019.01.091. PMID: 30784601

Introduction

The declining cost of high-throughput, next generation sequencing technology has permitted culture-free analysis of CF sputum, a respiratory specimen that variably samples secretions from the mouth to lower airways, most often by sequencing the bacterial 16S ribosomal RNA gene (16S amplicon sequencing). These culture-free methods have shown the microbiota (the full complement of bacterial taxa present) in CF respiratory samples to be more diverse than previously thought, often comprising species not detected by routine clinical culture [36–38]. However, studies of the CF sputum microbial community using these newer methods have thus far failed to find consistent predictors of either therapeutic success or clinical status. Furthermore, relatively little is known about the genetically conferred functional capacity of CF sputum microbial communities. Although bioinformatic pipelines exist to infer functional capacity of a community from 16S amplicon sequencing [93], these methods can only use what is available in annotated bacterial genomic databases and can miss differences in accessory genomes across strains. Sequence analysis of the "metagenome," the total complement of genes present in a community, can provide further insight into not only the taxonomic composition of the microbiota, but also its functional capacity directly from sequencing data [94]. Metagenomic analysis has been used in fecal samples [95] and, to a limited extent, in respiratory samples [90,96,97] and has the potential to identify functional traits that are required for persistence in chronic infections like those of the CF lung. Unlike some complex microbiota communities, such as those in fecal samples from the healthy GI tract and in soil, many complex clinical samples include large quantities of immune cells and comparatively low microbial loads. As a result, metagenomic sequencing of sputum and other respiratory samples can be hindered by the overabundant proportions of human DNA, relative to microbial DNA; for example, approximately 95% of metagenomic sequencing reads from CF sputum samples are annotated as human [96,97]. High ratios of human-to-microbial DNA are a barrier shared by many complex, human-associated microbial communities, such as those found in healthy oral [98], skin [99], vaginal [100], blood [101], middle ear and nasopharyngeal samples [102]. Human-to-microbial DNA ratios may be even higher in samples from inflamed and/or infected sites due to the influx of immune cells [103,104]. The cost of metagenomic analysis of samples with high human-to-microbial DNA ratios can thus become prohibitive when considering the depths required for thorough investigation of the microbiome in a large number of such

complex samples. For this reason, functional characterization of sputum and other samples containing relatively high ratios of human-to-microbial DNA has lagged behind studies investigating stool and other microbial communities that contain a high microbial abundance.

A related problem affecting metagenomic analyses of infected tissues is the presence of extracellular DNA derived from human cells and/or microbes. Extracellular DNA in complex infections can be produced as a result of the processes that occur during and in response to such infections, including cellular turnover accelerated by antibiotic therapy, inter-bacterial competition and host defense, as well as from the production of microbial biofilms (DNA is an abundant component of these extracellular matrices for certain microbes). Failing to address the impact of extracellular DNA on molecular microbial analyses has the potential for systematic bias when the intended focus is on viable microbial cells. For example, proportions of extracellular microbial DNA may vary when collected before and after antibiotic treatment [105] and may bias characterization of live bacteria persisting after antibiotic treatment if not accounted for during processing. Therefore, methods to more accurately characterize the functional capacity of infecting communities using metagenomic sequencing not only require human DNA depletion, but also enrichment for the cellular DNA fraction.

A number of studies have detailed techniques for depleting human DNA from microbial communities [98,106–110]. Few have evaluated the efficiency of these different methods for depleting human DNA from inflammatory and (relatively) low-abundance infections, their ability to remove extracellular DNA of either human or microbial origin, or the bias introduced by these methods to the phylogenetic composition of these samples. An extensive investigation and characterization of such methods is an important step to take prior to conducting metagenomic sequencing studies of CF sputum. In this thesis, this step was required in order to maximize microbial sequencing coverage and minimize the contribution of DNA from human cells or extracellular bacterial DNA as we planned studies to detect and interpret taxonomic and functional correlates of clinical state or response to treatments such as antibiotics (Chapters 3-4), which can increase the proportion of bacterial extracellular DNA, thus masking the genetic contributions of surviving cells. The aim of this study was to compare the results of different methods designed to deplete human DNA extracted from CF sputum, focusing in particular on (1) the

relative effect of each method on the calculated taxonomic constituency of the community, and (2) how accurately the results of each method reflect the viable microbial fraction. We quantified human DNA depletion via both next generation sequencing and qPCR, demonstrating that a combination of hypotonic lysis and nuclease digestion most effectively reduces human and extracellular microbial DNA in CF sputum. Using phylogenetic analysis of metagenomic and 16S amplicon sequencing data, we compared the effects of this and other methods on microbial sequence read depth and on detection of rare taxa, as well as the contribution of extracellular bacterial DNA to apparent community composition, in clinical sputum samples. We then compared the accuracy of these methods for measuring the constituency of mock communities containing cultured bacterial species common in CF sputum. Finally, we demonstrate that the increased read depth afforded by this method enhanced sensitivity for detecting genes of potential clinical importance. This chapter describes the optimization and characterization of this method, which I used subsequently prior to metagenomic sequencing of clinical samples in studies investigating the effects of a common antibiotic treatment on sputum microbiota (Chapter 3) and metagenomes (Chapter 4).

Results

Efficiency of human DNA depletion

We first compared four methods, each used in published studies [98,106–110] or available as commercial kits specific for human DNA depletion, for their ability to deplete human DNA from CF sputum samples (Fig. 1). “Antibody Depletion” utilizes immunoglobulins against methylated epitopes specific to eukaryotic DNA; “Cell Lysis” employs selective eukaryotic cell lysis with Trypsin-EDTA and Tween-20; “Molysis” involves chaotropic lysis of human cells followed by endonuclease digestion of extracellular DNA; and “Benzonase1” utilizes hypotonic lysis of human cells followed by endonuclease digestion of extracellular DNA. Each of these methods was used to supplement a standard phenyl:chloroform-based DNA extraction method, which was also tested without pre-processing (“Standard extraction”).

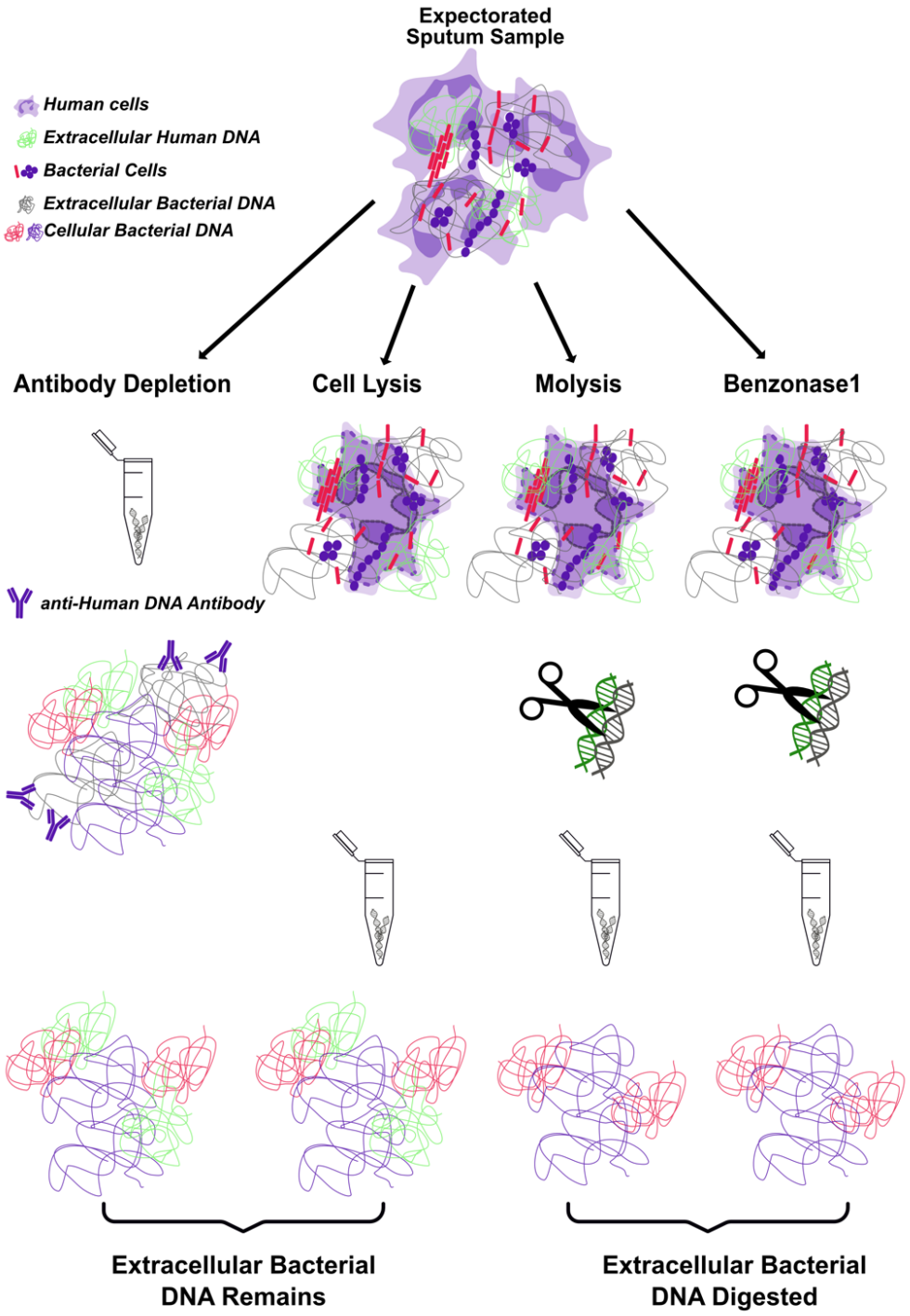

Figure 1. Diagrammatic representation of Human DNA depletion methods used in this study

Sputum samples from eight CF patients (a collection we refer to here as “Test Set 1”) were processed using each method for selective depletion of human DNA followed by Standard DNA extraction, or Standard extraction alone to assess the degree of selective human DNA reduction, we calculated the proportion of human to total (human and bacterial) DNA in terms of genome equivalents (GE) per µL identified in respective broad-range qPCR reactions (Fig. 2a), for all samples and extraction methods. In parallel, we performed metagenomic sequencing of all DNA extracts and then calculated the proportion of human to total sequence reads by mapping reads to the human genome (Fig. 2b). By both sequencing and qPCR measures, non-depleted samples contained 96% human DNA on average, consistent with previous reports [96,107]. In this study, the Benzonase1 method was the most efficient for selective human DNA removal from Test Set 1; on average, 68% of reads from Benzonase1 treated samples mapped to the human genome. This reduction in human reads resulted in a 5-fold increase in microbial reads on average and increased microbial sequencing depth (Fig 2c). Importantly, reduction in human DNA did not correspond with a decrease in the total bacterial load after any extraction method, indicating increased microbial sequencing depth could be achieved (Fig 2c), without significant loss of bacterial DNA (Figure 2d).

Effect of human DNA depletion on apparent microbial community composition

We next assessed the effect of the different depletion methods on apparent microbial community structure for Test Set 1. We hypothesized that selectively depleting human DNA in these samples would result in higher effective microbial sequence read depth and would improve detection of low-abundance taxa in these samples. MetaPhlAn2 [111] was used to define the taxonomic composition of Test Set 1

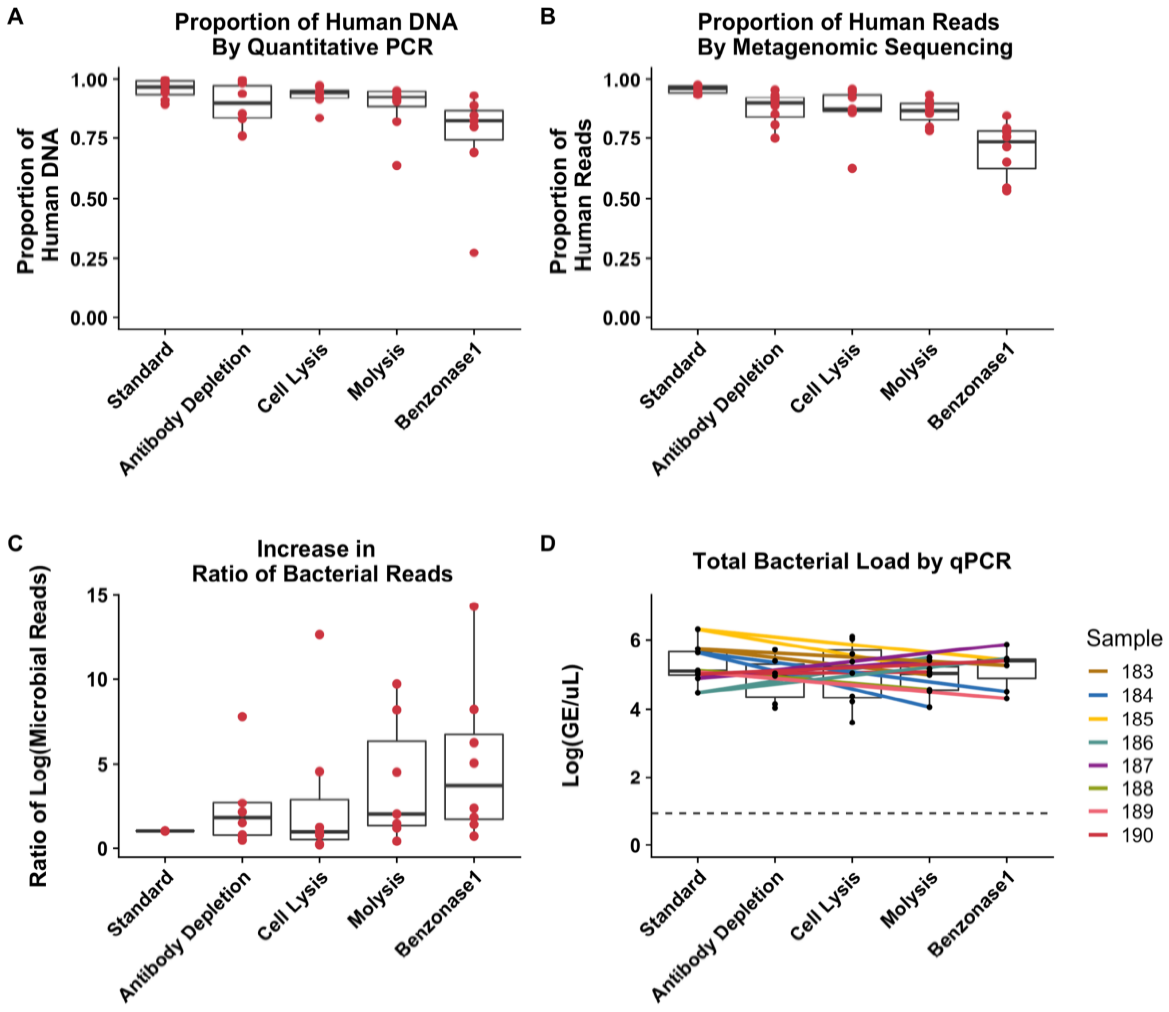

Figure 2. Comparison of four methods for selective depletion of human DNA from CF sputum samples

DNA was extracted from eight CF sputum samples with Standard extraction and four different methods for selectively depleting human DNA. A) Proportion of human DNA to total calculated DNA as determined by quantitative PCR of both human and bacterial DNA. B) Proportion of human to total reads, calculated by mapping all metagenomic sequencing reads to the human genome. C) Ratio of microbial metagenomic sequencing reads yielded by each extraction method compared to Standard extraction. D) Total bacterial load yielded by each extraction method, calculated by qPCR targeting the 16S rRNA gene. The colored lines connect data points from the same sample and the dotted line indicates the limit of detection. Boxes represent the interquartile region and black lines within boxes indicate the median value. Pairwise Wilcoxon signed rank tests were used to compare results from selective human DNA depletion methods to Standard extraction, with a Benjamini-Hochberg correction for multiple comparisons with all comparisons combined (A-D). No comparison reached the level of significance ($p<0.05$).

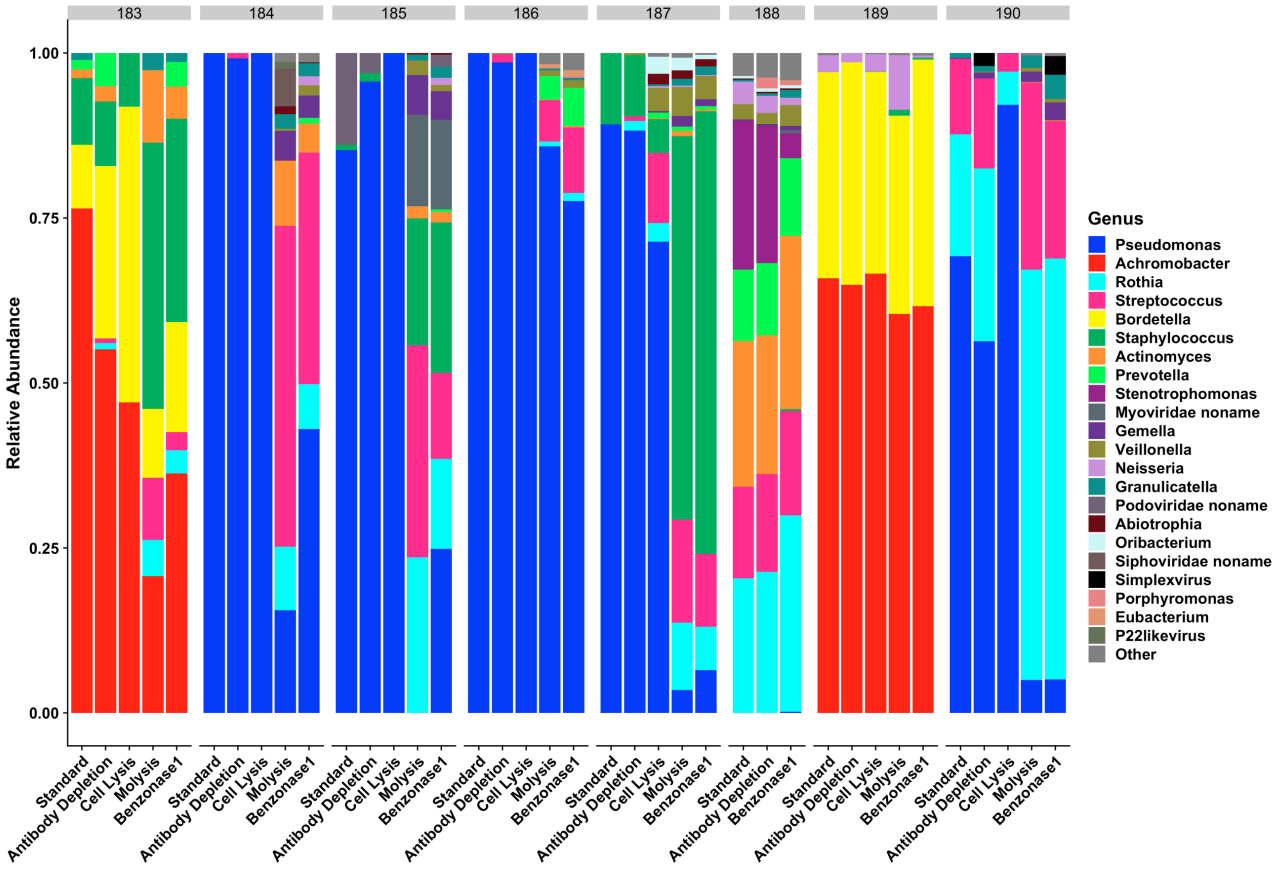

Figure 3. Phylogenetic profile of the 8 Test Set 1 CF sputum samples comparing 5 different processing and extraction methods

Phylogenetic composition determined via metagenomic sequencing followed by MetaPhlAn2. All taxa of relative abundance less than or equal to 1% in all samples are grouped into the “other” category (see Table 3 for details).

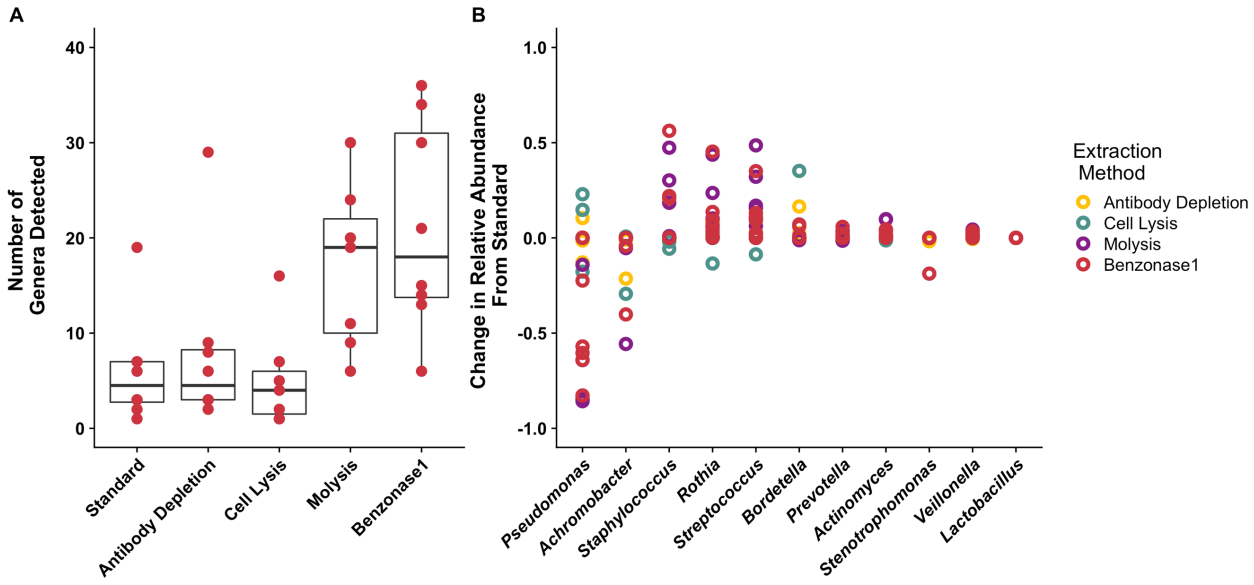

Figure 4. Effect of extraction method on metagenomic sequencing taxonomic profile and viable counts of the eight Test Set 1 CF sputum samples

A) Number of distinct genera detected from DNA prepared with each extraction method. Results from extractions including selective depletion of human and extracellular DNA were compared to standard extraction conditions. Box represents the interquartile region and black lines indicated the median value. Pairwise, Wilcoxon signed rank tests were performed comparing each extraction method to Standard with a Benjamini-Hochberg correction for multiple comparisons, with no comparison reaching the level of significance ($p<0.05$). B) Difference in relative abundance identified by each extraction method (compared with standard extraction) for the ten most abundant taxa. Each data point indicates an individual sample.

metagenomes (Fig. 3), and the effects of different depletion methods on the resulting calculated sputum microbial taxonomic profiles were compared. We identified more microbial taxa (Fig. 4a), as well as an apparent shift in the relative abundance of a number of taxa, in samples processed by either nuclease-based method (Molysis and Benzonase1, Fig. 4b, Fig. 3). Although this difference between each depletion method and Standard DNA extraction was not statistically significant, we did observe a large decrease in *P. aeruginosa* relative abundance, an important CF pathogen that is known to exude extracellular DNA [112,113], with both nuclease-based processing methods. Because of this pronounced effect of the nuclease-based processing methods on *P. aeruginosa* relative abundance (Fig. 3, Fig. 4b), and because of the importance of this pathogen in CF, we wondered whether our sample collection (which routinely includes freezing at -80°C) or processing methods could have impacted *P. aeruginosa* viability. We compared DNA extracted from four Test Set 1 samples both before and after freezing and showed no appreciable alteration in microbiota profiles (Fig. 5).

We next wondered whether aspects of our extractions methods themselves were artificially lowering the apparent relative abundance of *P. aeruginosa.* In the Benzonase1 method, the endonuclease is inactivated before downstream DNA extraction using EDTA and 0.15M NaCl prior to a final wash and resuspension of bacterial cells. We therefore tested whether the observed impact of nuclease-based extraction methods was due to premature lysis of bacterial cells during initial processing with EDTA, which has been demonstrated to kill *P. aeruginosa* growing planktonically [114] and in biofilms [115]. We tested the effect of hypotonic lysis followed by nuclease treatment (“Benzonase1”) on viable counts of six clinical *P. aeruginosa* isolates cultured from Test Set 1. We also modified the Benzonase1 protocol by adding EDTA later during processing (Fig. 6), after the final wash step, (referred to here as “Benzonase2”) and tested this procedural change on *P. aeruginosa* viability. These experiments demonstrated that, while Benzonase1 extraction resulted in an average 1.3 $\log_{10}$ reduction in *P. aeruginosa* CFU, the Benzonase2 method had no detectable impact on viable count (Fig. 7a), suggesting that earlier enzyme inactivation with EDTA may have artificially lowered the apparent *P. aeruginosa* relative abundance with Benzonase1. Interestingly, we did not observe a similar effect with other taxa commonly identified in CF sputum (Fig. 7b).

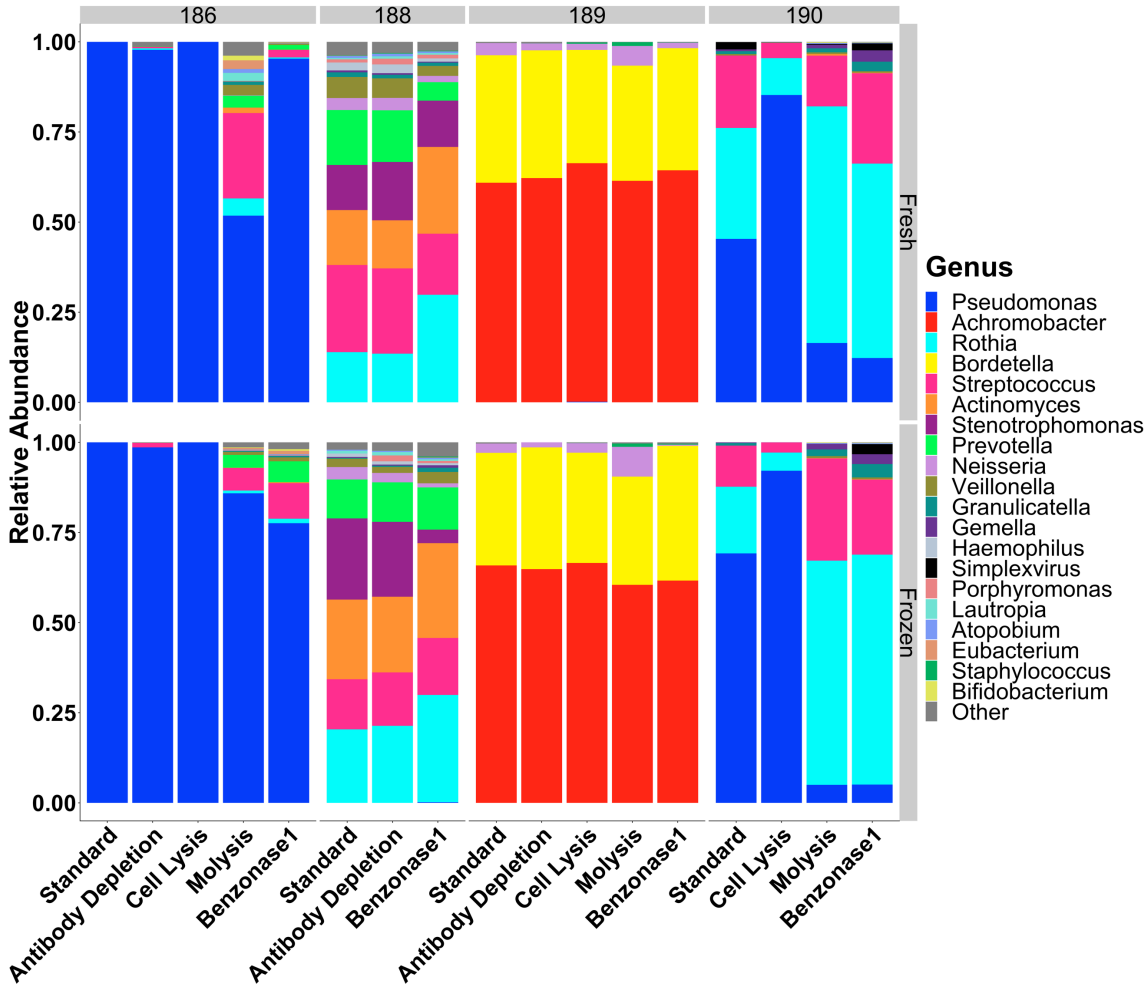

Figure 5. Effect of a single freeze-thaw cycle on phylogenetic composition

Bacterial phylogenetic composition determined by MetaPhlAn2 of four Test Set 1 CF sputum metagenomes, without DTT homogenization and extracted either before (top, “Fresh”) or after (bottom, “Frozen”) freezing at -80℃.

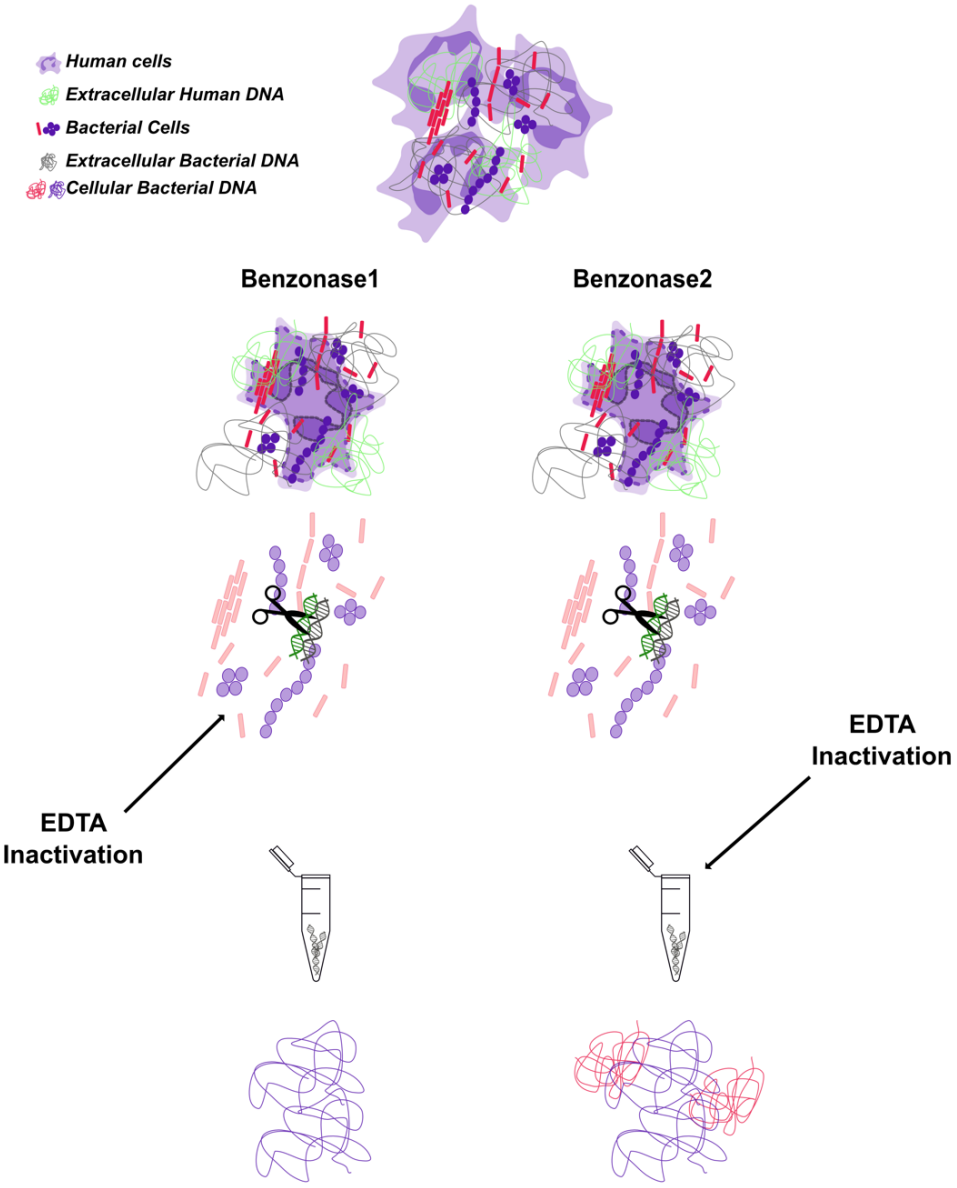

Figure 6. Diagrammatic representation of the distinction between the Benzonase1 and Benzonase2 methods

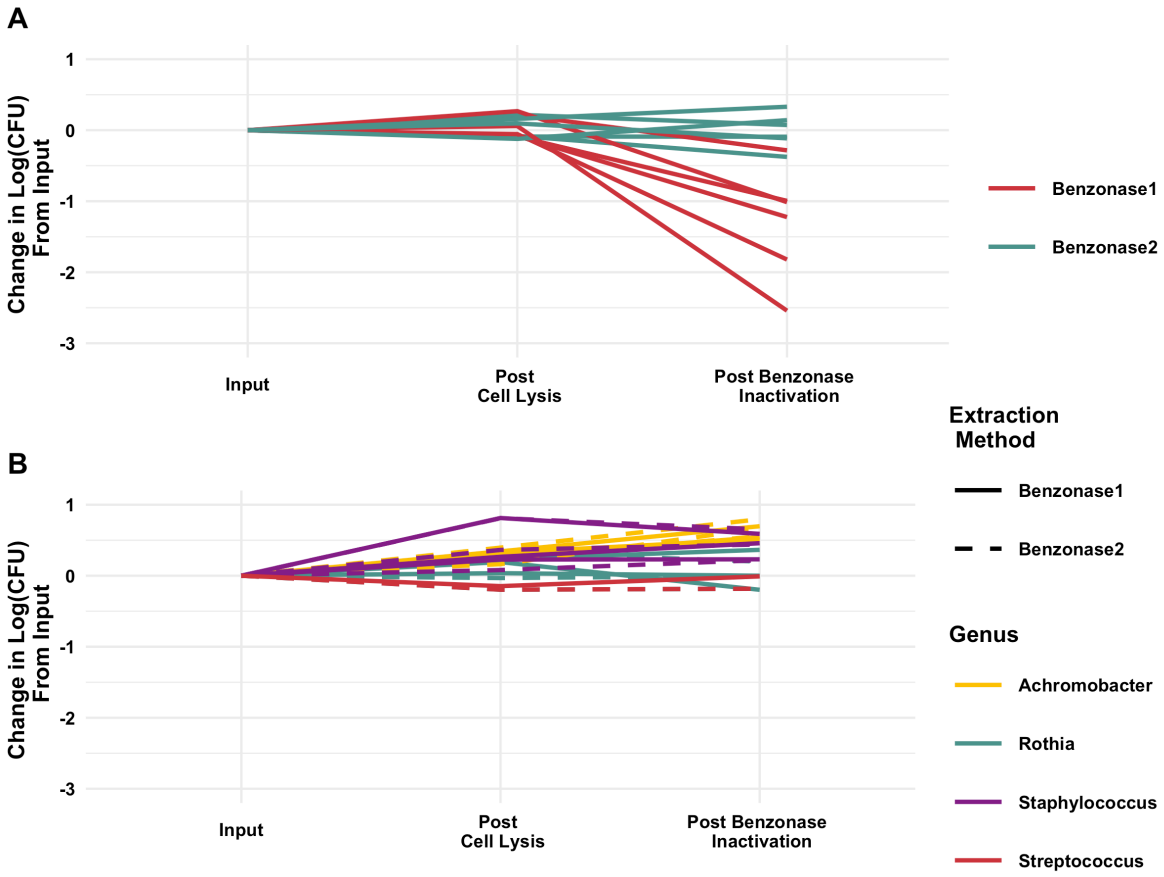

Figure 7. Effect of Benzonase treatment on viable counts

A) Effect of Benzonase treatment on viable *P. aeruginosa* counts. Cultures of six separate *P. aeruginosa* clinical isolates were subjected to both Benzonase1 and Benzonase2 processing methods, which differ only in the order in which EDTA is added. Viable counts were measured after the hypotonic, host-cell lysis step and after enzyme inactivation, as indicated. B) Effect of Benzonase treatment on multiple taxa. Cultures of eight separate clinical isolates from Test Set 1 (3 *Achromobacter,* 3 *Rothia,* 3 *Staphylococcus* and 1 *Streptococcus)* were subjected to both Benzonase1 and Benzonase2 processing methods, which differ in the order in which EDTA is added. Viable counts were measured after the hypotonic, host-cell lysis step and after enzyme inactivation, as indicated.

Effect of adapted Benzonase2 protocol on metagenomic and 16S amplicon sequencing

To further compare the effects of the Benzonase1 and Benzonase2 methods on taxonomic profiles from metagenomic sequencing, we constructed a smaller set of four sputum samples ("Test Set 2"), one from Test Set 1 (186) and a sample each from three additional CF patients. Each sample was treated with the mucolytic agent dithiothreitol (DTT) [116] and then mixed to fully homogenize, thus limiting bias from potential regional differences among the different sputum aliquots used for extraction. DNA was extracted with Standard extraction, Benzonase1 and Benzonase2 methods, and taxonomic profiles were defined by metagenomic sequencing followed by MetaPhlAn2 analysis. As before, we observed a decrease in *P. aeruginosa* relative abundance and increased detection of lower-abundance taxa with both Benzonase methods compared with Standard extraction (Fig. 8, top). This observed decrease in *P. aeruginosa* relative abundance may solely be due to an increase in reads from other microbes after Benzonase processing, perhaps as a result of increased microbial sequencing depth, rather than reduction of the absolute abundance of *Pseudomonas*. Because taxonomic coverage in 16S amplicon sequencing is less impacted by large quantities of human DNA due to amplification of a bacterial-specific gene, we addressed this possibility by analyzing each sample and extraction method using 16S amplicon sequencing (Fig. 8, bottom). It is important to note that although large quantities of human DNA can inhibit amplification of the 16S gene [117], we saw similar total 16S amplicon read numbers and proportions of amplicon sequencing reads annotated as human across different extraction methods when applied to Test Set 1, suggesting this effect was minimal (Table 1). 16S amplicon sequencing demonstrated decreases in *P. aeruginosa* relative abundance after Benzonase1 and Benzonase2 extractions that were similar to those demonstrated by metagenomic sequencing, although of a lower magnitude. On average, 97% of metagenomic sequencing reads from DNA prepared with the Standard extraction and 60% of reads in Benzonase-treated extracts mapped to the human genome (Fig. 9a,b). Microbial reads increased 15-fold following Benzonase1 treatment and 14-fold following Benzonase2, on average, compared with Standard extraction (Fig. 9c). As before, the total bacterial load was similar between extraction methods, indicating minimal effect of these processing methods on microbial DNA extraction efficiency (Fig. 9d). Our data indicate a further decrease in the proportion of human reads and increase in microbial reads when comparing Standard extraction to Benzonase treatment in Test Set 1 vs. Test Set 2 (68% vs. 60%, Fig. 2b-c vs. Fig. 9b-c), suggesting DTT-

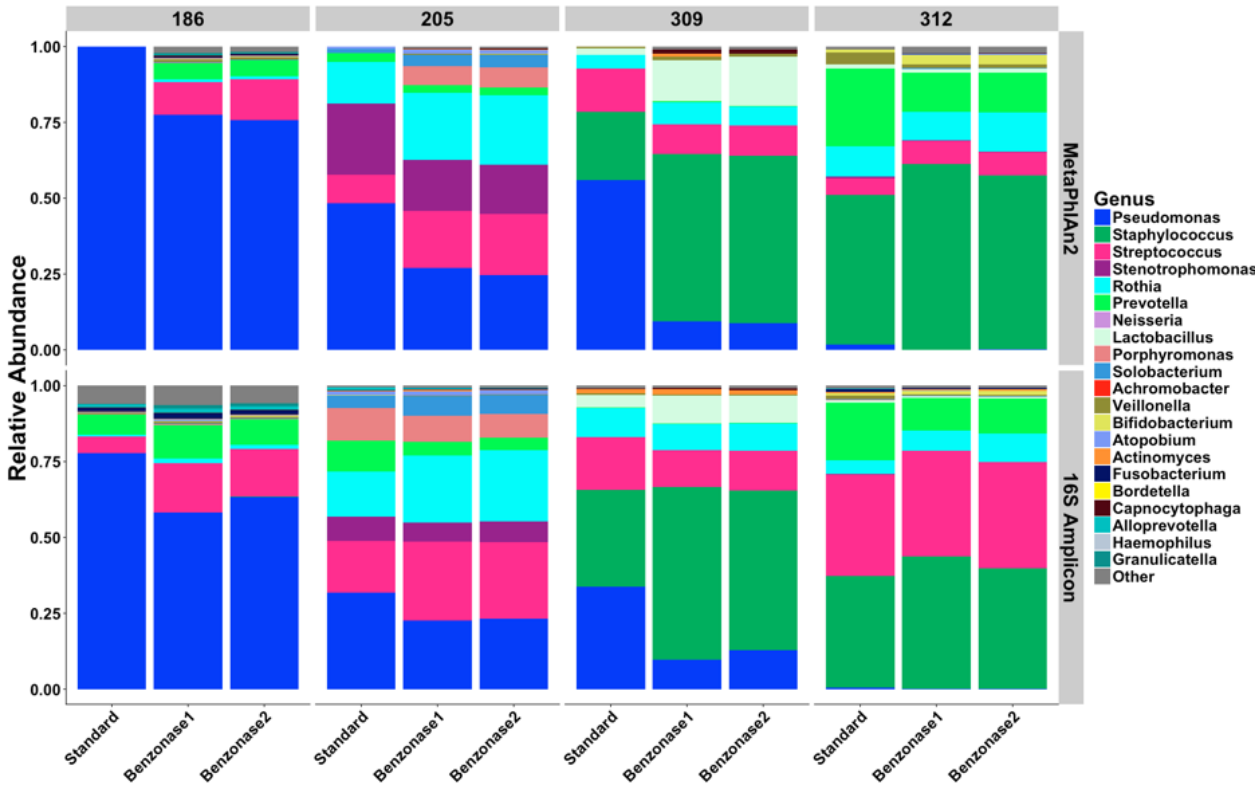

Figure 8. Sequence-based phylogenetic composition of DTT-treated Test Set 2 samples after Standard and Benzonase extraction

DNA extracted from each of four DTT-treated, homogenized CF sputum samples was analyzed using both metagenomic sequencing (MetaPhlAn2, top row) and 16S amplicon sequencing (16S Amplicon, bottom row).

Sample	Raw Sequences	Proportion Remaining Post QC	Proportion Remaining Post Merge	Proportion Remaining Post Chimera Removal	Proportion Remaining From Raw	Proportion Human Reads to Post Chimera
186 Standard	50445	0.913	0.983	0.906	0.814	0.001
186 Benzonase1	45736	0.922	0.988	0.953	0.868	0.006
186 Benzonase2	64622	0.924	0.993	0.967	0.887	0.004
205 Standard	72895	0.914	0.986	0.926	0.834	0.001
205 Benzonase1	47488	0.922	0.987	0.939	0.853	0.003
205 Benzonase2	49677	0.914	0.987	0.939	0.847	0.002
309 Standard	46277	0.931	0.995	0.981	0.908	0.002
309 Benzonase1	54987	0.918	0.983	0.905	0.816	0.003
309 Benzonase2	58418	0.915	0.985	0.915	0.824	0.004
312 Standard	71960	0.913	0.991	0.913	0.826	0.002
312 Benzonase1	112453	0.907	0.991	0.915	0.823	0.001
312 Benzonase2	130460	0.912	0.991	0.914	0.825	0.001
Mock Benzonase2	24605	0.922	0.988	0.847	0.771	0
Mock Benzonase1	23938	0.916	0.989	0.86	0.779	0
Mock Standard	22043	0.912	0.989	0.85	0.767	0
Mock Cell Lysis	25303	0.927	0.989	0.88	0.806	0
Mock PMA	25467	0.92	0.988	0.849	0.772	0
Mock Molyis	22570	0.911	0.991	0.861	0.777	0
Molysis Blank	217	0.899	0.933	1	0.839	0
NTC1	2	1	0.5	1	0.5	0
NTC2	31	0.226	1	1	0.226	0
Benzonase2 Blank	188	0.904	0.929	0.987	0.83	0
Benzonase1 Blank	53	0.868	0.848	0.949	0.698	0
PMA Blank	125	0.88	0.918	1	0.808	0
Standard Blank	81	0.802	0.738	0.979	0.58	0
Tween Blank	100	0.91	0.912	0.988	0.82	0

Table 1. Number of 16S amplicon sequences remaining after each filtering step in the DADA2 pipeline

Raw Sequence Number: Number remaining after sequencing in both forward and reverse sequencing files.

Proportion after QC: Proportion of raw reads remaining after QC filtering.

Proportion after merging: Proportion of post-QC reads remaining after merging forward and reverse reads.

Proportion after chimera removal: Proportion of post-merging reads remaining after chimera removal.

Proportion from raw: Proportion of raw reads remaining after chimera removal. This represents the proportion of reads remaining after all quality control steps.

Proportion Human reads: Proportion of post-chimera removal reads annotate as human. Human reads were determined using BLASTn. The RDP Bayesian Classifier was unable to annotate these reads during initial analysis.

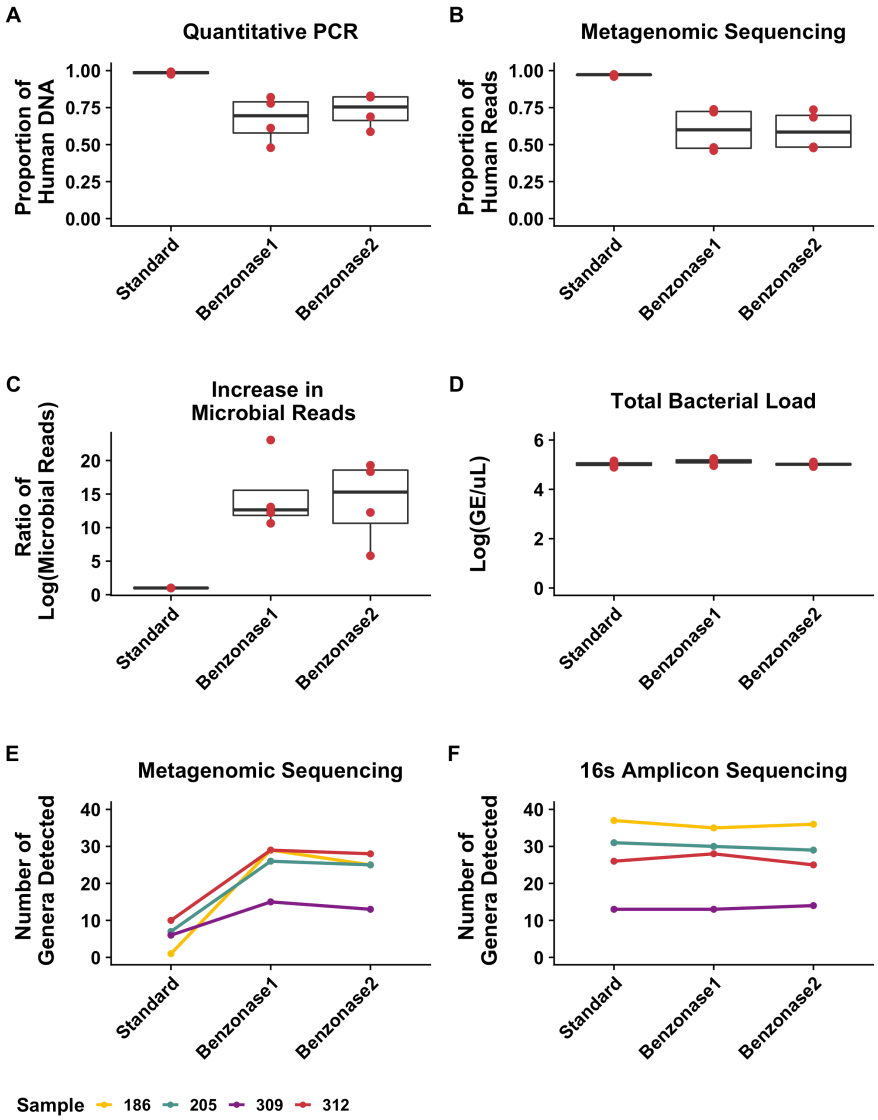

Figure 9. Effect of the refined Benzonase2 extraction method on selective human DNA depletion and microbial sequencing depth

Total DNA from the four Test Set 2 sputum samples was extracted using Standard, Benzonase1 and Benzonase2 extraction methods and analyzed to show: A) Proportion of human DNA relative to total DNA as determined by qPCR. B) Proportion of human to total reads calculated by mapping all metagenomic sequencing reads to a reference human genome. C) Ratio of microbial shotgun sequencing reads yielded by each extraction method compared to the Standard extraction. D) Total bacterial load (genome equivalents, GE) yielded from each extraction method as determined by qPCR targeting the 16S rRNA gene. Boxes represent the interquartile region and black lines indicate the median value. Number of genera detected in each extract using E) metagenomic sequencing or F) 16S amplicon sequencing. Each color represents a different sample. Results from each extraction method were compared to the standard extraction conditions using pairwise, two-sided t-tests with a Benjamini-Hochberg correction for multiple comparisons, identifying no significant differences.

based homogenization may increase efficiency of nuclease-mediated reduction in host DNA. Finally, the number of genera detected after both Benzonase treatments compared to Standard extraction alone was increased in metagenomic sequencing but not 16S amplicon sequencing (Fig. 9e-f). Notably, for three out of the four samples analyzed, a similar number of genera were detected by metagenomic sequencing and 16S amplicon sequencing following either Benzonase treatment, suggesting that nuclease-based methods provide similar microbial sequencing depths via both metagenomic sequencing and amplification of a bacterial-specific gene.

Nuclease-based extraction better reflects true, viable diversity in the sputum microbiota than other tested methods

Because Benzonase treatment depletes human reads by digesting extracellular DNA, it remained possible that the taxonomic differences found with Benzonase processing compared with Standard extraction (Fig. 3) were due primarily to degradation of extracellular microbial DNA rather than improved microbial sequence read depth. Extracellular DNA is also excluded from sequencing by another processing method, treatment with propidium monoazide (PMA), a chemical that cross-links extracellular DNA and selectively prevents its amplification [118]. PMA has been used in clinical samples to specifically focus on viable microbial cells [119,120] (similar to the intended use of Benzonase processing) and has been used in CF sputum 16S amplicon sequencing studies for this reason [113,121]. PMA has also been used to deplete human DNA from human-associated microbiota [108]. However, these prior studies focused on complex clinical samples, as in our experiments, and thus did not determine the relative effects on exclusion of microbial versus human DNA. Therefore, to more rigorously define and compare the effects of Benzonase and PMA processing on microbial community structure, we constructed an *in vitro* bacterial mock community containing cultured cells of taxa commonly present in CF respiratory samples. DNA was extracted from these communities using six methods: Standard extraction or Standard extraction with additional Cell Lysis, Molysis, Benzonase1, Benzonase2, or PMA treatment. We cultured the mock community to determine the viable cell counts of each community member and to confirm viable relative abundances. We were also able to detect extracellular DNA in the supernatant of these cultured taxa using qPCR (Fig. 10). We then compared phylogenetic composition identified by 16S amplicon

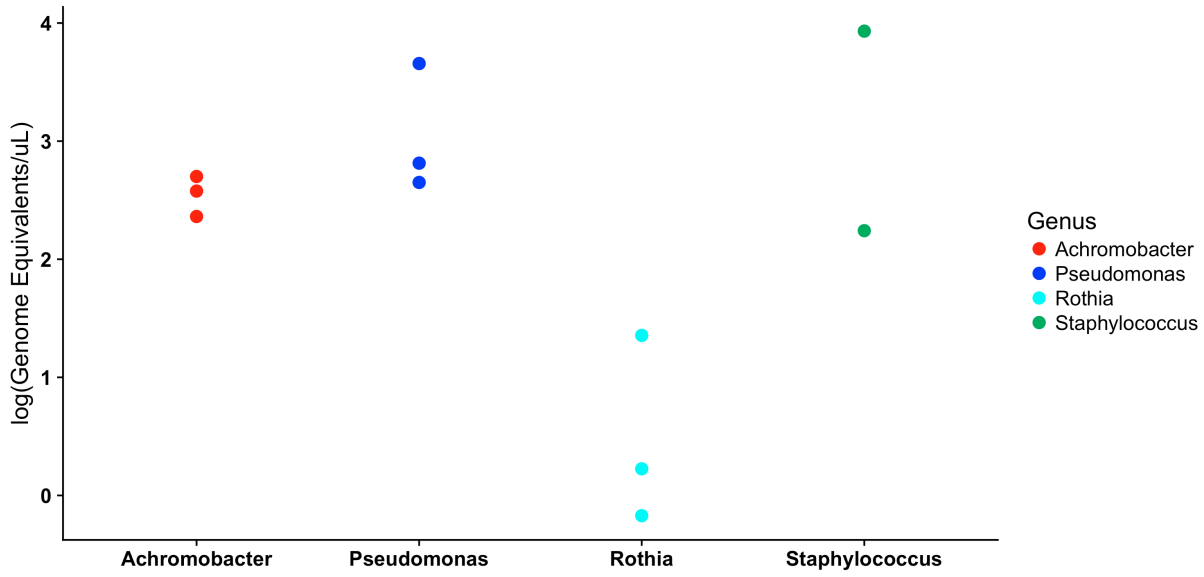

Figure 10. Detection of extracellular DNA in culture supernatants

Three isolates of *A. xylosoxidans,* three isolates of *P. aeruginosa,* two isolates of *Rothia dentocariosa*, one isolate of *R. aeria* and two isolates of *S. aureus* were cultured overnight. Quantitative PCR against the bacterial 16S gene was performed on culture supernatant to detect extracellular DNA from each taxon.

sequencing and metagenomic sequencing to culture-validated input community composition (Fig. 11). We predicted that the Benzonase2 taxonomic profile from 16S amplicon sequencing would most closely resemble PMA (reflecting depletion of extracellular bacterial DNA), and that the Benzonase2 taxonomic profile from metagenomic sequencing (which is less subject to the amplification bias of 16S amplicon sequencing [122]) would most closely resemble calculated input. As predicted, the taxonomic composition identified by metagenomic sequencing after Benzonase2 processing was most similar to culture-validated input (Fig. 11, left), indicating that DNA extracted after Benzonase2 processing most closely resembles the DNA of the viable bacterial community, whereas relatively low concordance was observed between cultured results and the PMA-treated DNA extracts. In contrast, community composition identified by 16S amplicon sequencing after Molysis and Benzonase2 processing was most similar to PMA-treated communities (rather than calculated input), consistent with the ability of PMA to prevent amplification of extracellular DNA (Fig. 11, right). Molysis (which differs from the Benzonase methods in both the nuclease and eukaryotic cell lysis method used) produced a community structure similar to that produced by Benzonase2 as identified both by metagenomic sequencing and 16S amplicon sequencing, further underscoring the effect of endonuclease digestion of extracellular DNA on apparent phylogenetic composition. Therefore, nuclease-based processing prior to DNA extraction performed at least as well (16S amplicon sequencing) or better (metagenomic sequencing) than PMA at identifying the viable bacterial constituency of polymicrobial mixtures.

Selective depletion of human DNA increases coverage of microbial genes in metagenomic sequencing

To assess the effect of selective human DNA depletion on sequence coverage of microbial genomes, we computationally constructed contigs from the metagenomic sequencing dataset from all four Test Set 2 sputum samples using Standard, Benzonase1 and Benzonase2 processing. We then mapped all microbial reads from the dataset to these contigs, quantifying the mean coverage (average sequencing depth at each base pair) of each DNA extract as a measure of microbial sequence read depth. For both Benzonase methods, mean coverage increased across all contigs compared to Standard extraction (Fig. 12a), despite similar raw read counts between extraction methods (Table 2), indicating higher microbial

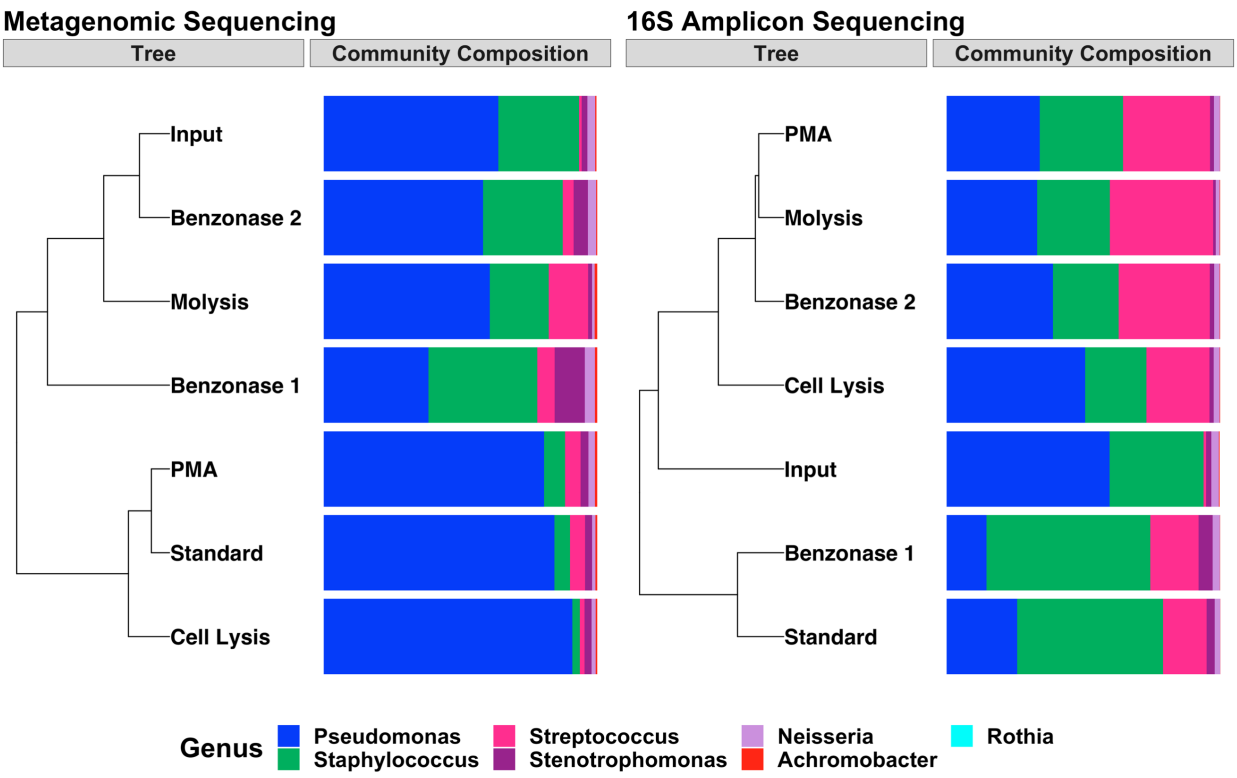

Figure 11. Effect of extraction method on sequencing-based taxonomic profile of a bacterial mock community

Data represents a single mock community extracted in parallel with each extraction method. Phylogenetic tree constructed using Bray-Curtis dissimilarity ("Tree") and the corresponding phylogenetic composition ("Community Composition") determined via A) metagenomic sequencing and B) 16S amplicon sequencing. Input refers to relative abundance of the mock community before extraction based on quantitative culture. Viable counts were corrected for 16S copy number in analysis of 16S amplicon sequence data.

Sample	Raw Sequences	Post Deduplication	Post QC Trim	Post Human Removal	Human Reads	Proportion Human
186 Standard	1.36E+07	1.36E+07	1.14E+07	3.12E+05	1.11E+07	0.973
186 Benzonase1	1.88E+07	1.85E+07	1.56E+07	4.08E+06	1.15E+07	0.738
186 Benzonase2	1.79E+07	1.75E+07	1.45E+07	3.83E+06	1.07E+07	0.736
205 Standard	1.99E+07	1.98E+07	1.65E+07	6.65E+05	1.59E+07	0.96
205 Benzonase1	3.54E+07	3.47E+07	2.84E+07	1.53E+07	1.30E+07	0.459
205 Benzonase2	2.91E+07	2.86E+07	2.37E+07	1.22E+07	1.15E+07	0.485
309 Standard	2.65E+07	2.63E+07	2.33E+07	6.00E+05	2.27E+07	0.974
309 Benzonase1	2.70E+07	2.65E+07	2.28E+07	6.39E+06	1.64E+07	0.719
309 Benzonase2	1.34E+07	1.33E+07	1.11E+07	3.49E+06	7.56E+06	0.685
312 Standard	5.99E+07	5.95E+07	5.48E+07	1.48E+06	5.33E+07	0.973
312 Benzonase1	3.95E+07	3.88E+07	3.48E+07	1.81E+07	1.67E+07	0.481
312 Benzonase2	6.26E+07	6.10E+07	5.47E+07	2.86E+07	2.61E+07	0.477
Mock Standard	4.87E+07	4.68E+07	3.61E+07	3.61E+07	1.96E+02	0
Mock Tween	2.81E+07	2.74E+07	2.03E+07	2.03E+07	1.36E+02	0
Mock Molysis	7.49E+07	7.06E+07	5.80E+07	5.80E+07	3.09E+02	0
Mock Benzonase2	2.91E+07	2.85E+07	2.35E+07	2.35E+07	1.83E+02	0
Mock Benzonase1	3.72E+07	3.63E+07	2.81E+07	2.81E+07	2.88E+02	0
Mock PMA	3.56E+07	3.45E+07	2.78E+07	2.78E+07	2.26E+02	0

Table 2. Number of metagenomic sequences remaining after each filtering step by sample for Test Set 2 and Mock Communities

Raw Sequence Number: Total number of metagenomic sequencing reads.

Number after deduplication: Total number of reads remaining after removing duplicate reads.

Number after QC filter: Total number of reads remaining after quality control filtering.

Number after human DNA removal: Total number of reads remaining after removal all reads mapping to the human genome.

Human reads: Total number of reads mapping to the human genome.

Proportion human reads: Proportion of post-QC filter reads mapping to the human genome.

Concentration (ng/uL): Concentration of DNA in extracted DNA from each sample. This represented input into metagenomic sequencing libraries.

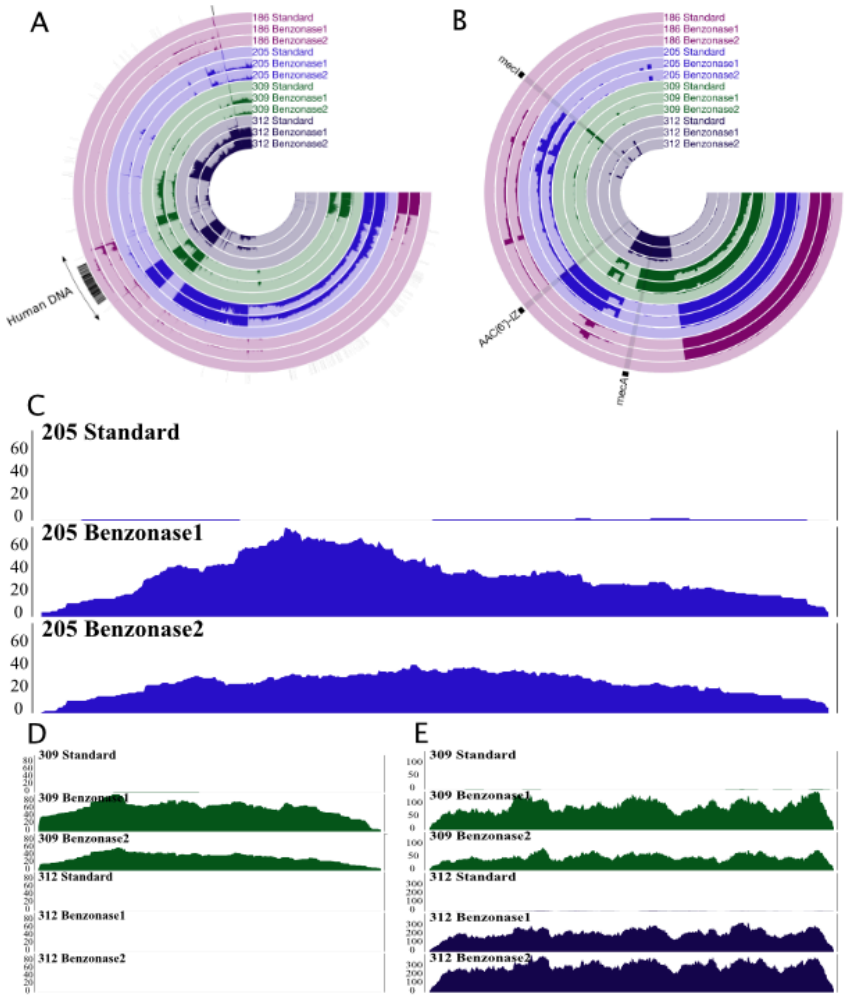

Figure 12. Increase in microbial sequence coverage after human DNA depletion of the four Test Set 2 sputum samples

A) Contigs were assembled using the same four sputum samples and processing methods as in Figures 4 and 5 and all reads were subsequently mapped back onto these contigs. Each radial line represents a single contig ordered by Euclidean distance based on sequence content. Height of each bar of darker shading represents the average coverage across that contig (average of the sequence coverage of each nucleotide across a given contig) in a given sample from 0-10x. Radial black lines in outermost ring indicate contigs annotated as human. B) Average coverage of 145 identified antibiotic resistance genes in each sample as a proportion of 10x coverage. Each radial line represents an individual antibiotic resistance gene. Three genes described in the text are indicated. Height of bars represents mean coverage for each gene from 0-10x. C-E) Coverage map of three antibiotic resistance genes for select samples and extraction methods. The y-axis represents coverage at each base pair and the x-axis represents the length of each ORF. C) AAC(6')-IZ, D) *mecI* and E) *mecA*. Note the absence of *mecI* detection in sample 312, but the presence of *mecA* in both 309 and 312.

sequencing coverage was achieved by Benzonase processing. Subsampling metagenomic sequencing reads followed by MetaPhlan2 analysis suggested this increased microbial sequencing depth was adequate to profile community structure (Fig. 13a). For both Benzonase extractions, the graph of species richness versus number of sequences reached a plateau for three out of four samples at 1-5 million reads, with sample 186 beginning to plateau at 7 million reads. These data concur with the results presented in Figure 9e-f, which demonstrated that Benzonase extraction results in taxonomic richness similar to 16S amplicon sequencing. It is interesting to note that although the rarefaction curves for Standard extraction plateau for all samples, fewer species were detected with Standard than with Benzonase extraction, further revealing the improved microbial detection afforded by Benzonase-based extraction (Fig. 12a). Furthermore, Figure 6 indicates that the microbiota identified by metagenomic sequencing only resembles that of 16S amplicon sequencing when the latter was performed after nuclease-based extraction, again suggesting the species richness calculated after Benzonase-based extraction more closely resembles the “true” CF sputum microbiota, at least as determined by the sequencing method used most often in this field (16S amplicon sequencing).

We next assessed the effects of Benzonase processing on detection and characterization of individual microbial genes in the sputum metagenome data. We focused our analysis on antibiotic resistance genes due to their potential clinical importance. All reads from all samples were mapped against the Comprehensive Antibiotic Resistance Database (CARD), an annotated database of antibiotic resistance genes and their associated proteins and phenotypes [123]. Of the 2,239 genes in this collection, 145 were at least one of these four samples (at least one read mapping to a given gene in at least one sample). Across all four samples, an average of 67% of the antibiotic resistance genes detected from extracts prepared with either Benzonase method were not detected from DNA prepared using the Standard extraction. No resistance genes were detected solely in Standard extraction samples. For Benzonase2 extracted samples, the mean coverage for detected resistance genes was 11.0x, 17.8x, 6.5x and 35.4x in samples 186, 205, 309 and 312 respectively. In contrast, the highest mean coverage across all Standard extractions was 0.4x (Fig. 12b). This increased coverage was seemingly independent of the degree to which a given taxon was detected by metagenomic sequencing or 16S amplicon sequencing. For example, sequencing coverage for the chromosomal aminoglycoside acetyltransferase from

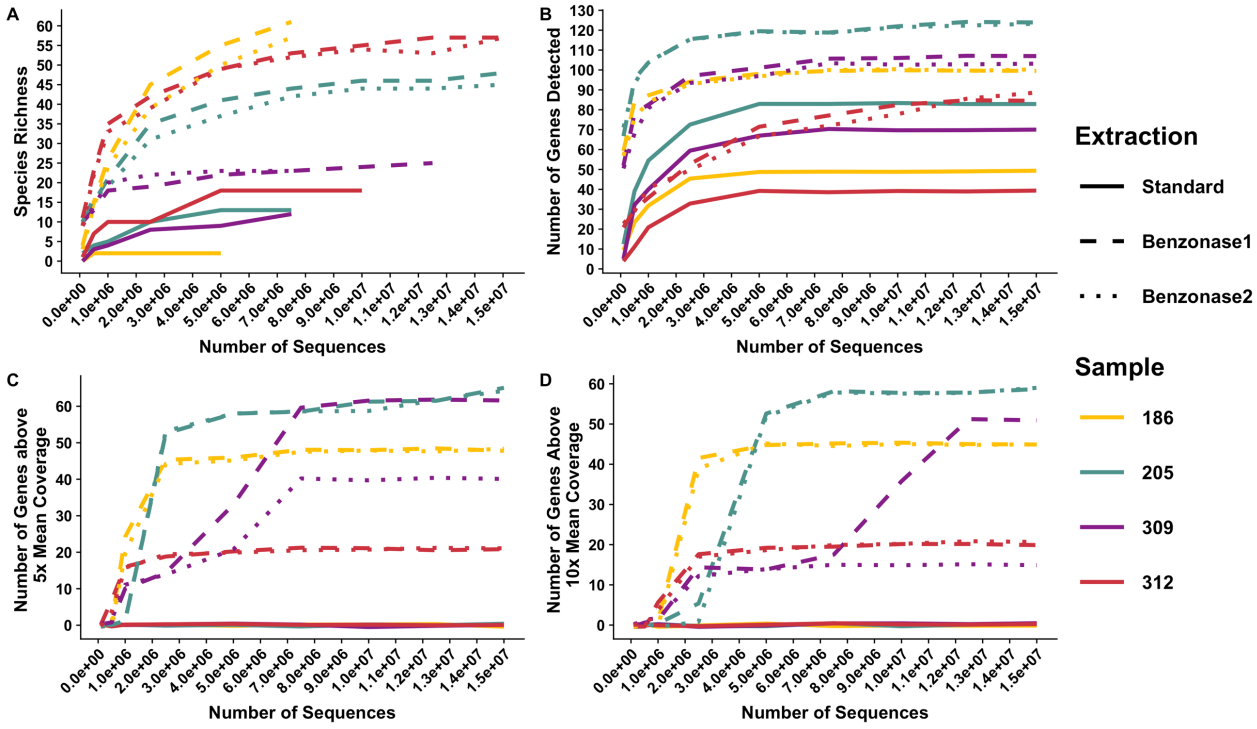

Figure 13. Determining adequacy of sequencing depth for detecting taxa and antibiotic resistance genes

All metagenomic sequencing files were subsampled to 0.1, 0.5, 1, 2.5, 5, 7.5, 10, 12.5 and 15 million reads. A) Taxonomic composition was determined independently for each subsampled sequencing file using MetaPhlAn2. Richness was calculated as the total number of detected species. B-D) Antibiotic resistance genes were detected as described in methods for Figure 6b. B) Number of genes with at least one read mapping to that gene, C) number of genes with at least 5x average coverage across the gene and D) number of genes with at least 10x average coverage across the gene.

Stenotrophomonas, AAC(6')-IZ, was <1x after Standard extraction but 30x after Benzonase2 (Fig. 12c), despite this genus being detected at a similar relative abundance after Standard or Benzonase extractions (Fig. 8). Metagenomic sequencing also detected differences in species-specific antibiotic resistance gene profiles among different samples with that species. For example, both *mecA*, a gene that confers MRSA status to *S. aureus*, and *mecI*, encoding an inhibitor of *mecA*, were detected in sample 312, whereas only *mecA* was detected in sample 309, despite a similar relative abundance of *S. aureus* in both samples (Fig. 8). Standard extraction afforded <1x average coverage for both of these genes (Fig. 12d-3). It is particularly useful to note that that we employed the lowest limit of detection possible (one read mapping to a given gene), and it is likely that higher stringency criteria required to confidently detect antibiotic resistance genes using metagenomic sequencing would have further limited the sensitivity of Standard extraction. Finally, subsampling metagenomic sequencing reads followed by mapping these rarefied reads against the same antibiotic resistance gene database indicated that the increased microbial sequencing depth provided by Benzonase extraction was sufficient to characterize the inferred antibiotic resistance profile of our communities (Fig. 13b-d).

Discussion

Although metagenomic sequencing of clinical samples has the potential to provide insight into both the taxonomic and function features of these microbial communities, current methods for analyzing these polymicrobial infections are insufficient to handle samples with overwhelming quantities of human DNA relative to microbial DNA. Metagenomic sequencing of sputum from people with chronic airway infections due to CF results in a majority of human reads that must be computationally identified and removed prior to downstream analyses, a process that commonly results in inadequate microbial sequence read depth. Bacteria in the CF lung can undergo frequent cellular turnover due to interspecies competition, antibiotic therapy and host defense, resulting in large quantities of extracellular bacterial DNA that can persist in infected tissues and secretions. Copious amounts of extracellular DNA can lead to inaccurate estimates of viable bacterial cell abundance and can undermine the ability of metagenomic sequencing to analyze the functional capacity of a microbial community by masking functions or taxa in a community, especially those persisting after antibiotic therapy.

Here, we describe an optimized method for DNA extraction from CF sputum, a prototypical complex clinical sample, that enriches for microbial DNA from viable cells by (1) selective lysis of human cells with hypotonic treatment, and (2) endonuclease digestion of both human and microbial extracellular DNA. Our analysis demonstrated this approach has superior ability to deplete human DNA from CF sputum compared with four other published or commercially available methods, resulting in a 14-fold increase in microbial reads without affecting total bacterial load measured by qPCR.

This method also shifted the community structure compared with the Standard extraction method, most notably by reducing the relative abundance of the canonical CF pathogen *P. aeruginosa*. A similar decrease was observed for a related Gram-negative taxon, *Achromobacter*, also an opportunist and biofilm former common in CF infections [124,125], in one Test Set 1 sample (Fig. 3). There are several potential reasons for these observations. For example, it is possible that heterogeneity in microbial distribution within sputum samples led repeatedly to these differences occurring by chance; however, we consider this to be unlikely, as we observed similar differences between Standard and Benzonase extraction for all samples with *Pseudomonas*. While it was also possible that treatment with Benzonase selectively lyses *Pseudomonas* and *Achromobacter* cells, we identified no change in viable counts of either taxon after Benzonase2 treatment (Fig. 5). Instead, analysis of mock communities suggests that reduction of extracellular bacterial DNA by nuclease treatment explained the relative abundance shifts observed for sputum samples. Thus, nuclease treatment likely enables more optimized characterization of viable bacterial communities than Standard DNA extraction methods. Furthermore, the increased sequence read depth afforded by digestion of all extracellular DNA enables higher microbial gene coverage and, thus, may allow improved detection of taxa present at low relative abundance, allowing more in-depth study of the contribution of these less-well-studied taxa to disease status and response to therapy.

There are a number of lines of evidence supporting this interpretation. For example, we were able to culture taxa from 5/8 Test Set 1 samples that were only detected by sequencing after nuclease processing (both Molysis and Benzonase1, results not shown), confirming their presence in the samples. Although this does not provide conclusive evidence that those taxa detected after Benzonase2 extraction

are solely those that were alive in CF sputum, it does indicate that Standard extraction methods would fail to detect potentially important, viable species in CF respiratory samples, and that Benzonase2 extraction provides a more focused reflection of the viable microbial community in these samples. In addition, several taxa detected by metagenomic sequencing only after nuclease processing were detected via 16S amplicon sequencing after all extraction methods, suggesting Benzonase treatment achieved metagenomic coverage to determine microbial community composition reflecting that of a method less biased by sequence read depth. Together, these results indicate that the increased microbial sequence read depth provided by nuclease processing improves detection of lower abundance taxa by metagenomic sequencing. Nevertheless, the relative abundances of *Pseudomonas* and *Achromobacter* yielded by the two nuclease methods (Benzonase and Molysis) were similar despite differences in human DNA depletion, indicating that depletion of human DNA alone was not responsible for this effect. *Pseudomonas* is known to extrude extracellular DNA into its environment, particularly when forming biofilms [85], further supporting the idea that the taxonomic differences observed after Benzonase2 extraction were due to depletion of extracellular bacterial DNA.

Using mock bacterial communities, we further demonstrated that endonuclease treatment results in a more accurate representation of viable community diversity compared with Standard DNA extraction. When analyzed by metagenomic sequencing, mock community profiles generated from DNA prepared with Benzonase2 processing most closely resembled the input community structure, again suggesting extracellular DNA may influence sequencing-based taxonomic profiling, even with *in vitro* cultured bacteria. The larger reduction in *Pseudomonas* relative abundance that we observed for some samples after Benzonase1 processing, which differs from the Benzonase2 protocol only in timing of EDTA treatment, reflected the loss of viable counts seen with this extraction method, providing further support for the close relationship between results from sequencing and culture. Results after extraction with the Molysis method, another nuclease-based processing method that does not use EDTA to inhibit enzymatic digestion of extracellular DNA, also closely resembled the input mock community, further underscoring the effect of extracellular DNA depletion on phylogenetic composition defined by sequencing.

When mock communities were analyzed via 16S amplicon sequencing, results from both Benzonase2 and Molysis processing, both of which involve nuclease treatment, closely resembled those from PMA extraction (a method used commonly to prevent amplification of extracellular DNA), further underlining the negative impact extracellular DNA can have on sequence-based characterization of viable community structure. PMA treatment creates double strand breaks and DNA aggregates only in DNA it can access (i.e. extracellular) [126,127], limiting amplification of longer targets (e.g. amplicon sequencing). Because metagenomic sequencing principally utilizes smaller fragments and involves little amplification, it is less likely to be impacted by this depletion method, a prediction supported by the differences we found between results from PMA and nuclease methods by metagenomic versus 16S sequencing (Fig. 5). Two previous studies used PMA to exclude extracellular DNA prior to metagenomic sequencing [109,128], but both used an additional whole-genome amplification step before sequencing. Our results indicate that PMA is not ideal with limited-amplification sequencing methods currently in use (such as in this study). Marotz et al. successfully used PMA to deplete human DNA from saliva for metagenomic sequencing [108], although they did not specifically explore its effects on extracellular bacterial DNA. It is unclear why PMA was not as useful in reflecting viable input than were nuclease methods in our mock communities by metagenomic sequencing. It is possible that PMA efficacy is limited by the physical properties of either extracellular DNA or sample chemistry. For example, in the limited number of sputum samples tested, we did not see an effect of PMA on the proportion of human DNA (data not shown). The complexity and viscosity of these CF sputum samples compared to saliva may explain this effect.

We found that Benzonase processing substantially increased metagenomic sequencing coverage of microbial genes in CF sputum compared with Standard extraction. While people with CF are frequently treated with antibiotics, the microbial determinants of response (or lack of response) to these treatments are poorly understood. Metagenomic sequencing offers a promising approach to this problem; for example, these methods can detect, classify, and quantify longitudinal changes in antibiotic resistance genes in infectious bacterial communities during treatment. However, the sequencing read depth provided by Standard extraction is insufficient for confident detection or classification of many of these genes and gene variants, which will limit the utility of metagenomic sequencing to infer functionality.

These limitations can be overcome through the increased microbial coverage yielded by Benzonase2 processing. Furthermore, metagenomic sequencing with Benzonase2 processing identified differences in antibiotic resistance gene profiles among samples with similar relative abundances of a given taxon; by contrast, functional inference from 16S amplicon sequencing of the same samples would not indicate any differences, highlighting the potential of metagenomic sequencing and this extraction method for infectious disease studies. Comparison of sequencing results from mock communities revealed the potential bias introduced by extracellular DNA in defining viable community structures in infections, which are continuously perturbed and remodeled by host immune activity, antibiotics and other treatments, nutrient limitation, interspecies interactions, natural cell turnover, and during biofilm production. This creates the potential for the extracellular DNA load to be systematically larger among subjects in treatment arms of antibiotic therapy trials, which would bias the comparison of metagenomes between sample groups if extracellular DNA was not excluded. We found that the nuclease-based processing technique Benzonase2 resulted in the most accurate representation of viable community structure among the tested methods. While Molysis also provided an accurate picture of the viable bacterial community, this method was not as effective in enriching for bacterial reads in our samples.

Many of the methods used in this comparative study have been used individually in CF respiratory microbiome studies. For example, PMA treatment has been used prior to 16S amplicon sequencing [88,89] and metagenomic sequencing [108]. Regarding nuclease treatment, Lim et al. used DNAseI to reduce human DNA in sputum samples prior to metagenomic sequencing [91] and demonstrated a reduction in human genome equivalents but did not define the impact of this processing on microbial reads or coverage. Furthermore, 25 mM EDTA was used to inactivate the enzyme in these studies; here we show that EDTA concentrations as low as 5 mM can lyse *P. aeruginosa*, with unclear effects on community profiles. Leo et al. used the Molysis method to treat CF bronchoalveolar lavage fluid (BALF) [129], a method that resulted in 72% human DNA after depletion. By comparison, we found this method to be less efficient at reducing human DNA in CF sputum, perhaps due to the relatively high viscosity and complexity of sputum compared to BALF. In addition, we are aware of a commercially available kit for depletion of human DNA that uses detergent-mediated lysis of human cells followed by Benzonase digestion of extracellular DNA, but that differs from the methods described here in the use of

proteinase K to inactivate the endonuclease. Although not tested here, we believe this method may not be optimal for analyzing CF sputa or, perhaps, other clinical samples, which often require enzymatic digestion steps after extracellular DNA removal to efficiently extract DNA from common taxa, such as *S. aureus* [130].

Regardless, this method represents an important step towards extending metagenomic sequencing, a powerful and promising technology, for analyzing complex clinical samples such as CF respiratory samples. While this method affords both enhanced sequencing coverage adequate for functional metagenomics, and a sharper focus on viable bacterial cells, it is not yet known whether microbiome analyses focused on viable bacterial cells will correlate more closely with clinical outcomes than do results from other microbiological approaches. In chapters 3-4, we apply this refined extraction method to a large sample set of CF sputa collected before, during and after antibiotic treatment, with matched clinical data to begin tackling this question.

Methods

Table 3. Oligonucleotides used in this work

Name/Target	Name/Target	SOURCE
Human beta-Globin Forward	GGGCAACGTGCTGGTCTG	Handschur et al., 2009
Human beta-Globin Reverse	AGGCAGCCTGCACTGGT	Handschur et al., 2009
Universal 16S Forward	TCCTACGGGAGGCAGCAGT	Nadkarni et al., 2002
Universal 16S Reverse	GGACTACCAGGGTATCTAATCCT GTT	Nadkarni et al., 2002

Table 4. Software used in this work

Name	Source	Version
Bio-Rad CFX Manager 3.1	Biorad	1845000
SeqUniq	https://github.com/standage/sequniq	Version 0.1
KneadData	https://bitbucket.org/biobakery/kneaddata/wiki/Home	Version 0.6.1
Trimmomatic	Bolger et al., 2014	Version 0.33
BMTagger	https://bioconda.github.io/recipes/bmtagger/README.html; Human Microbiome Project	Version 3.101
MetaPhlAn2	Thompson et al., 2017[131]; Truong et al., 2015[111]	Version 2.2.0
DADA2	Callahan et al. 2016 [132]	Version 1.6.0
RDP Bayesian classifier	Wang et al., 2007 [133]	Implemented through DADA2 (Version 1.6.0)
R	R Core Team, 2017	Version 3.4.2
ggplot2	Wickham, 2009 [134]	Version 2.2.1
Vegan	Oksanen et al., 2017 [135]	Version 2.4.4
Ape	Paradis et al., 2004 [136]	Version 5.0
ggtree	Yu et al., 2016 [137]	Version 1.10.0
Megahit	Li et al., 2015	Version 1.1.2
Bowtie2	Langmead and Salzberg, 2012	Version 2.2.6
Prodigal	Hyatt et al., 2010	Version 2.6.3
Centrifuge	Kim et al., 2016	Version 1.0.3-beta
Anvi'o	Eren et al., 2015	Version 4
Inkscape	www.inkscape.org	Version 0.92.3

Human sputum samples used in this work

This study was approved by the Seattle Children's Hospital (SCH) Institutional Review Board. Sputum samples were collected from children diagnosed with CF who presented to Seattle Children's Hospital as part of regular clinical care, were willing to provide samples and provide informed consent, and were able to expectorate at least 1mL of total sputum (so that we would have enough sputa to split each sample into multiple aliquots). Participants were selected to reflect a range of prior sputum culture results (both in terms of dominant culturable species and abundances), and they were on a mix of inhaled, oral and/or IV antibiotics and presented with a range of clinical statuses (stable, exacerbation, in treatment). Finally, samples 183 and 189 were from the same individual, collected on different days. Test Set 1 samples were homogenized by passing through a 1mL syringe approximately 10 times, aliquoted and either processed immediately or frozen at -80°C prior to extraction. Test Set 2 samples were diluted 1:1 with 10% Sputolysin (EMD Millepore, 56-000-010ML), aliquoted evenly and frozen prior to extraction.

Mock Community Construction

Seven individual taxa isolated from CF sputa (*Pseudomonas aeruginosa*, *Staphylococcus aureus*, *Neisseria sp. unclassified*, *Achromobacter xylosoxidans*, *Streptococcus salivarius*, *Stenotrophomonas maltophila* and *Rothia mucilaginosa*), were grown in liquid culture in to mid-log phase in Brain Heart Infusion broth (*Neisseria*, *R. mucilaginosa*, *S. salivarius*) or Tryptic soy broth (*P. aeruginosa*, *S. aureus*, *A. xylosoxidans*, *S. maltophila*) and taxonomy confirmed by Sanger sequencing of the 16S rRNA gene. Specified volumes of each culture were mixed by gently vortexing and inverting and were then diluted to generate a mock community. Serial dilutions of each taxa were simultaneously grown on 5% Sheep's Blood Agar for viable counts to retrospectively determine viable community composition, resulting in the following calculated relative abundances: 63.8 % *Pseudomonas*, 29.5 % *Staphylococcus*, 2.89% *Neisseria*, 0.398% *Achromobacter*, 1.02% *Streptococcus*, 2.04% *Stenotrophomonas* and 0.370% *Rothia*. For comparison with 16S amplicon sequencing, we adjusted relative abundances by 16S rRNA copy number, resulting in 60.0% *Pseudomonas*, 34.3% *Staphylococcus*, 2.89% *Neisseria*, 0.398%

Achromobacter, 1.02% *Streptococcus*, 2.04% *Stenotrophomonas* and 0.370% *Rothia*. DNA extraction was performed immediately after mock community construction as detailed below.

To quantify extracellular DNA in *in vitro* culture supernatants, three separate isolates of *A. xylosoxidans*, three of *P. aeruginosa*, two of *R. dentocariosa*, two of *S. aureus* and one of *R. aeria* were cultured as indicated above and 1mL of culture was removed. Bacterial cells were pelleted by spinning at 10,000g for 3 minutes, supernatant was collected and syringe filtered using a 0.22μm PES filter. DNA was extracted using Standard extraction as detailed below, beginning at the Proteinase K step. Total bacterial load was determined via 16S qPCR as detailed below.

DNA Extraction

For all samples and depletion methods, Standard DNA extraction was performed as follows: approximately 200 mg of sputum or 200 μL of mock community was suspended in 1 mL of PBS and centrifuged at 13,000 g for 3 min. The pellet was suspended in 400 μL TE. A mixture of 1 mm and 0.1 mm silica:zirconia beads and a single tungsten-carbide bead was added to TE solution and followed by bead-beating for one minute in a BioSpec MiniBeadBeater. The resulting solution was boiled for 5 min at 95°C. Lysozyme (Sigma L6876, 3 mg/mL final) and lysostaphin (Ambi LSPN, 0.14 mg/mL final) were added, and the sample was incubated for one hour at 37°C. Proteinase K (Invitrogen 25530049, 1.4 mg/mL final) and SDS (1.8% final) were added, and the resulting solution incubated at 56°C for 30 min before cooling to room temperature. The solution was removed to a separate tube and 5 M NaCl was added (2 M final) before adding phenol:chloroform:isoamylalcohol (25:24:1) at a 1:1 volume. The solution was then incubated for 20 min at room temperature, centrifuged at 13,000 g for 20 min and the top aqueous layer was collected. 0.133 volume equivalent of 7.5 M ammonium acetate was added to the aqueous layer, and the resulting solution diluted 1:1 with cold 100% ethanol to precipitate DNA. DNA product was cleaned with a spin column.

For the “Antibody Depletion” method, 1 μg of total extracted DNA was processed with the NEBNext Microbiome DNA Enrichment Kit (NEB E2612S) according to the manufacturer’s instructions. For the “Cell Lysis” method [138], prior to Standard DNA extraction sputum was suspended in 1 mL PBS,

vortexed and centrifuged at 13,000 g for 2 min. The resulting pellet was suspended in 1 mL TrypZean (Sigma T3449) and 0.05% Tween-20 (Sigma P9416) and incubated at 37°C for 60 min. The solution was then vortexed to mix and centrifuged at 5,000 g for 2 min to pellet eukaryotic cells. The supernatant was then removed and centrifuged at 13,000 g for 10 min to pellet any prokaryotic cells and suspended in PBS before proceeding with Standard DNA extraction. For the “Molysis” method, prior to Standard DNA extraction sputum was initially processed with Molzym’s Molysis Human DNA removal kit (D-300-050) according to manufacturer’s instructions, pausing before the “Buglysis” step. Cells were washed once with PBS and suspended in TE before proceeding with Standard DNA extraction. For the “Benzonase1” method [138], sputum was suspended in 7 mL dH_2O and incubated at room temperature for one hour with gentle agitation. 10x strength Benzonase buffer (200 mM Tris-HCl, 10 mM $MgCl_2$) to a final 1X and 250U Benzonase (Sigma E-1014) was added and the sample was incubated at 37°C for two hours with gentle agitation. The Benzonase reaction was quenched by adding EDTA (5 mM final) and NaCl (150 mM final). The resulting solution was spun at 8000 g for 10 minutes and the pellet washed once with PBS, then suspended in 400µL TE before proceeding with Standard DNA extraction. The “Benzonase2” method differed from Benzonase1 in the moving of one step: after the two-hour nuclease incubation and before EDTA inhibition, the bacteria were pelleted by centrifugation at 8000 g for 10 min and washed once in PBS, then suspended in 400µL TE, at which point EDTA (5 mM final) was added to inactivate the endonuclease before proceeding directly to Standard extraction. For the “PMA” method, 300uL PBS was added to 200uL of the bacterial mock community, 1 µL propidium monoazide (20mM in water) was added and the solution incubated at room temperature for 5 minutes. The solution was then incubated under 160 LED white light for 15 minutes with gentle agitation. The bacteria were pelleted by centrifugation at 13,000g for 10min, washed once in PBS and then suspended in 400 µL TE before proceeding with Standard extraction. Reagent blanks consisting of PBS alone were processed for each extraction method and sequenced via 16S amplicon sequencing.

Quantitative PCR

Total bacterial load was determined using quantitative PCR with PowerUp SYBR Green Master mix (Applied Biosystems A25742) and previously published primers and reaction conditions [139]. Total

human DNA load was determined using primers targeting the beta-globin gene, as previously described [140]. Data were analyzed with the Bio-Rad CFX Manager 3.1 software, using software-defined Cq thresholds. Proportion of human DNA was determined by calculating quantity of human and bacterial DNA from genome equivalents (GE), using a genome size of $6.5x10^9$ bp for human and $5x10^6$ bp for the average microbial genome. GEs for human cells and bacterial cells were multiplied by their respective genome sizes to calculate number of base pairs per microliter. “Total DNA” per microliter was calculated by adding the total bacterial and human base pairs. The proportion of human DNA was calculated by dividing total human base pairs per microliter by total base pairs per microliter. Based on the following commonly-used definition of limit of detection (LOD) for qPCR [141,142], “the lowest copy number associated with the serial dilution that gave a positive PCR response on 95% of occasions,” the LOD of our qPCR assay is approximately 8.24 Genome Equivalents (GE) per uL.

Effect of Benzonase processing on bacterial viability

Six separate isolates of *P. aeruginosa*, three separate isolates of *S. aureus*, two separate isolates of *Rothia dentocariosa* and one isolate each of *Streptococcus salivarius*, *Rothia aeria* and *Achromobacter xylosoxidans* cultured from Test Set 1 samples were grown in liquid culture to mid-log phase. Cells were washed in PBS and plated for viable counts on Brain Heart Infusion agar or Tryptic Soy Agar. Cells were then subjected to Benzonase1 and Benzonase2 extraction as described above, with aliquots plated for viable counts at input, after hypotonic lysis and after enzyme inactivation.

Phylogenetic composition from metagenomic shotgun sequencing

Next generation sequencing libraries were prepared for all samples using the Nextera DNA Sample Prep Kit (Illumina FC-121-1031) following manufacturer’s instructions. Three of the Test Set 1 samples (183-185) were sequenced on the Illumina MiSeq platform, producing an average of 2.55×10^6 reads per sample. The remaining samples (186-190, 205, 309 and 312) were sequenced on the Illumina HiSeq platform, producing an average of 1.55×10^7 reads per sample in our first set of 8 (Test Set 1) sputum samples and 3.48×10^7 reads in Test Set 2. Sequencing data from all samples were de-duplicated using SeqUniq (version 0.1) https://github.com/standage/sequniq) and quality filtered using KneadData (version

0.6.1) and Trimmomatic (version 0.33) [143]. Human reads were identified and removed with BMTagger (version 3.101) and community phylogenetic composition was determined using MetaPhlAn2 [111,131] (version 2.2.0) to produce a MetaPhlAn2 taxa table. All commands were executed with default settings, with the exception of KneadData, which was used with the "–run-bmtagger" flag. For our mock communities, the *Achromobacter* genus represents reads originally annotated either as *Achromobacter* or *Bordetella*, the latter of which we believe represents misidentified *Achromobacter* reads. We have found MetaPhlAn2 consistently splits this taxon into *Achromobacter* and *Bordetella* even when this later taxon was not added to mock communities and was not detected in 16S amplicon sequencing of the same samples.

Phylogenetic composition from 16S amplicon sequencing

The V4 region of the 16S rRNA gene was amplified using primers from the Earth Microbiome Project [131] and barcodes adapted from Kozich et al [144]. (detailed at https://github.com/SchlossLab/MiSeq_WetLab_SOP/blob/master/MiSeq_WetLab_SOP_v4.md). 16S amplicons were made under the following conditions: 94°C for 3min, 30 cycles of the following sequence: [94°C for 45s, 50°C for 60s, 72°C for 90s], and then 72°C for 10min. Libraries were constructed by pooling equimolar amounts of each sample or of each blank at the volume of the least concentrated sample. Libraries were sequenced on the Illumina MiSeq platform producing paired 300 bp reads.

16S amplicon sequencing data were analyzed using the denoising program DADA2 [133] (version 1.6.0) and the complete code is listed in Supplementary File 1. Briefly, we computationally trimmed 10bp off the beginning of the both the forward and reverse reads and truncated the forward read to 200bp and the reverse read to 100bp. We used our entire data set to define an error rate at each base pair and then denoised all sequences. Forward and reverse reads were merged and any pair without perfect overlap was removed. Finally, chimeric sequences were removed. This program produces a list of “Amplicon Sequence Variants” (ASV) analogous to OTUs generated with a 97% clustering method. Each ASV was annotated with the RDP Bayesian classifier [133] against the SILVA database [145] to produce

a 16S amplicon taxa table. ASVs identified as *Pseudomonas*, *Staphylococcus* and *Achromobacter* were analyzed with BLASTn to determine species identity.

Relative Abundance Analysis

After quality filtering, sputum samples produced an average of 56,347 reads per sample, mock communities produced an average of 18,687 reads, extraction blanks produced an average of 101 reads and amplicon blanks produced an average of 8 reads. Most reads in the blanks were taxa in common among samples with similar sequencing barcodes. As the absolute abundance of these reads was 2-3 logs lower in blanks than in neighboring samples during sequencing, we found it unlikely they contributed to the taxonomic profiles of our samples, but that these reads were more likely due to errors in barcode reading [146,147]. The remaining taxa were those noted by Salter et al. to be common reagent contaminants [148] and were found at <1% relative abundance in samples. We did not observe any increase in those taxa identified in extraction blanks with any extraction method compared to standard, suggesting that background contamination was not the source of the increase in taxonomic richness we observed with Benzonase extraction. As the number of reads from blanks were substantially outnumbered by those detected in sputum samples, we did not analyze the reads from these controls further (raw taxonomic tables, including taxa detected in controls, from 16S amplicon sequencing are in Table 5). We did not remove any reads from samples based on detection in controls.

Analysis of community composition was performed in R [149] (version 3.4.2) and visualized using ggplot2 [134] (version 2.2.1). Taxa tables from 16S amplicon sequencing and MetaPhlAn2 output were merged. For sputum samples, all taxa below 1% relative abundance in all samples or above 1% relative abundance solely in extraction blanks, were pooled into the “Other” category (these genera are listed in Table 2). For mock communities, we excluded two detected taxa which were not added to our mock community (*Pusillimonas* and *Caulobacter*), which both comprised <0.03% of the community.

Phylogenetic trees were constructed using the Vegan [135] and Ape packages [136] (versions 2.4.4 and 5.0 respectively) using the bray-curtis dissimilarity metric and the "Unweighted Pair Group Method with

Test Set 1		Test Set 2	
Acinetobacter	*Lautropia*	*Abiotrophia*	*Moraxella*
Actinobacillus	*Leptothrichia*	*Acinetobacter*	*Myoviridae noname*
Alloprevotella	*Lymphocryptovirus*	*Actinobacillus*	*Naumovozyma*
Alloscardovia	*Megasphaera*	*Alloscardoia*	*Oribacterium*
Atopobium	*Mycobacterium*	*Bacteroidetes noname*	*P22likevirus*
Bacteroidetes noname	*Naumovozyma*	*Brevibacterium*	*Parascardovia*
Bifidobacterium	*Parascardovia*	*Burkholderia*	*Parvimonas*
Brevibacterium	*Parvimonus*	*Camplyobacter*	*Peptoniphilus*
Burkholderia	*Peptoniphilus*	*Candida*	*Peptostreptococcus*
Campylobacter	*Peptostreptococcus*	*Cantonella*	*Podoviridae noname*
Candida	*Propionibacterium*	*Caulobacter*	*Propionibacterium*
Capnocytophaga	*Pusillimonas*	*Clostridiales Family XIII*	*Pusillimonas*
Catonella	*Rhodanobacter*	*Incertae Sedis noname*	*Rhodanobacter*
Caulobacter	*Reiemerella*	*Dialister*	*Riemerella*
Clostridiales Family XIII	*Scardovia*	*Escherichia*	*Scardovia*
Incertae Sedis noname	*Selemonas*	*Eubacterium*	*Selenomonas*
Corynebacterium	*Shuttleworthia*	*Gemella*	*Shuttleworthia*
Dialister	*Solobacterium*	*Kingella*	*Simplexvirus*
Ekienella	*Sphingobium*	*Kocuria*	*Siphoviridae noname*
Escherichia	*Sphingopyxis*	*Lachnoanaerobaculum*	*Sphingobium*
Fusobacterium	*Stomatobaculum*	*Lachnospiraceae*	*Sphigopyxis*
Haemophilus	*Tannerella*	*noname*	*Stomatobaculum*
Kingella	*Treponema*	*Lautropia*	*Tannerella*
Kocuria		*Lymphocryptovirus*	*Treponema*
Lachnoanaerobaculum		*Megasphaera*	*Corynebacterium**
Lachnospiraceae noname		*Mitsuokella*	*Mycobacterium**
Lactobacillus		*Eikenella*	

Table 5. Low relative abundance genera pooled into "other" category

Genera in Test Set 1 or Test Set 2 detected at <1% relative abundance across all samples or above 5% relative abundance solely in extraction blanks (*). Genera in test set 1 represent taxa detected with metagenomic sequencing. Genera in Test Set 2 represent taxa detected with either metagenomic sequencing or 16S amplicon sequencing.

Arithmetic Mean" agglomeration method for hierarchical clustering. Tree was visualized using the ggtree package [137] (version 1.10.0).

Microbial read depth analysis

Contigs from metagenomic sequencing data were assembled with Megahit (Version 1.1.2) [150], producing a total of 162,842. All reads were then mapped back to assembled contigs with Bowtie2 (Version 2.2.6) [151]. Taxonomy was determined by predicting open reading frames with Prodigal (Version 2.6.3) [152]. Contigs of human origin were identified with Centrifuge [153] based on these ORFs. The “nucleotide protein homolog model” collection from the Comprehensive Antibiotic Resistance Database [123] was used to detect antibiotic resistance genes. This model comprises 2,239 unique genes for which presence is sufficient to confer resistance (i.e. not including ubiquitous genes that only confer resistance when mutated such as DNA gyrase). All reads were mapped to this database with Bowtie2 as above. The complete code for this analysis is listed in Supplementary File 2. Mapping results were processed and visualized with Anvi’o (Version 4) [154] and figures were finalized with Inkscape (www.inkscape.org). For subsampled analysis, metagenomic sequencing reads after quality filtering and human read removal were subsampled to indicated depths and analysis was repeated as detailed above.

Statistical Analysis

All statistical analyses were performed in R. For all comparisons between standard extraction and processing methods for human DNA removal, p-values were calculated using a pairwise, Wilcoxon signed rank tests with a Benjamini-Hochberg correction for multiple comparisons. P-values smaller than 0.05 were considered significant.

Data availability

Metagenomic sequencing data is available through NCBI Bioproject (PRJNA516442)

Chapter 3

Maintenance Tobramycin Primarily Affects Untargeted Bacteria in the Cystic Fibrosis Sputum Microbiome

The majority of the work presented in this chapter under review as:

Nelson MT, Walter DJ, Eng A, Weiss EJ, Vo AT, Brittnacher MJ, Hayden HS, Ravishankar S, Bautista G, Ratjen A, Blackledge M, McNamara S, Nay L, Majors C, Miller SI, Borenstein E, Simon R, LiPuma JJ , Hoffman LR. Maintenance Tobramycin Primarily Affects Untargeted Bacteria in the CF Sputum Microbiome. *Thorax. In revisions, Dec. 2019.*

Introduction

Inhaled tobramycin is the antibiotic prescribed most often for CF lung disease [155] and has been demonstrated to decrease sputum *Pseudomonas aeruginosa* abundances and improve long term respiratory outcomes for individuals with CF on average when used in 28-day, alternate month cycles [76,78,79]. However, studies utilizing classic clinical culture have shown the short-term effects of tobramycin observed in treatment-naïve individuals—reduction in *P. aeruginosa* counts and improved respiratory function—are attenuated with successive drug cycles [76] and vary between individuals regardless of the *in vitro* susceptibilities of infecting *P. aeruginosa* to tobramycin [76,80–82].

Tobramycin is an aminoglycoside, a class of bactericidal antibiotics that act at bacterial ribosomes to inhibit protein synthesis and promote the production of toxic defective proteins[156]. Aminoglycoside-class antibiotics have been shown to have *in vitro* action against other common CF pathogens, including gram-negative organisms [157] and *Staphylococcus aureus* [158], in addition to *P. aeruginosa*. Together with the discrepancies between microbiological effects and clinical outcomes, this suggests that factors beyond *Pseudomonas*, perhaps involving other taxa or specific microbial genes, contribute to clinical and microbiological responses to this therapy. As elaborated in Chapter 1, metagenomic analysis provides an opportunity to study several poorly-understood features of CF lung disease, including the mechanisms underlying the resilience of CF respiratory infections, during potent antibiotic treatments [76,80–82].

We hypothesized that changes in CF sputum microbiomes (both microbiota and metagenomes) with inhaled tobramycin therapy would involve taxa beyond the intended target, *P. aeruginosa*, and that defining these microbiome dynamics would identify candidate mechanisms by which chronic infections in CF persist despite antibiotic treatment. Many previous studies have been limited by unstandardized study design and heterogeneous subject characteristics, making inferences from microbial analysis and clinical outcomes difficult. Thus, to test our hypotheses, we analyzed sputa collected from people with CF surrounding a standardized, routine maintenance course of inhaled tobramycin using culture, quantitative PCR (qPCR) and metagenomic sequencing using a method to specifically focus on the viable taxonomic constituency, as detailed in Chapter 2, to mitigate the effect of DNA from bacterial cells killed by antibiotic.

We compared microbiological changes during treatment with changes in pulmonary function measures and subjective symptom scores. Our goal was to specifically investigate the taxonomic and functional changes with a single course of maintenance inhaled tobramycin and to the determinants of sputum microbial persistence during antibiotic treatment in CF in order to direct the development of more focused and effective therapies for these and many other recalcitrant infections. This chapter will focus on taxonomic changes, while chapter 4 will elaborate on functional genomic differences with this maintenance therapy.

Results

Lung function and symptom scores changed minimally with inhaled tobramycin

We collected spontaneously-expectorated sputum samples from 30 individuals with CF (Table 1) before ("baseline"), weekly during ("weeks 1-4**")** and 1 month after ("Follow-up") a standard 28-day course of tobramycin inhaled powder (TIP), comprising 157 samples. Subjects had a mean age of 35 years (range 9-75) and most (24/30) subjects reported routine use of these treatments. Subjects received no antibiotics other than maintenance azithromycin for ≥1 month prior to or during the study, and subjects were withdrawn if they required other antibiotics during the study period to limit confounding clinical variables. Lung function measurements (forced expiratory volume in 1 second % predicted, $ppFEV_1$) and symptom scores (Chronic Respiratory Infection Symptom Score, CRISS) were collected (Fig. 1). Consistent with previous work [76], most participants' lung function measurements demonstrated little change with maintenance inhaled tobramycin treatment. We found a significant negative association between baseline $ppFEV_1$ and CRISS score, as well as between changes in these two metrics, with treatment (Fig. S1), indicating that improvement in lung function was associated with improvement in symptoms.

Changes in clinical metrics did not correlate with changes in sputum microbiota with inhaled tobramycin

Of the 30 study participants, 26 provided both baseline and week 4 sputum samples and performed spirometry. We decided *a priori* which microbial and clinical metrics to compare before analyzing the data: we compared changes in $ppFEV_1$ and CRISS scores with both baseline sputum microbiological measures

Category	Characteristic	Number of subjects (%)*
Study site	Seattle Children's Hospital	2 (7)
	University of Washington	17 (56)
	University of Michigan	11 (37)
Mean age in years (range)		35 (9-75)
Gender		
	Female	16 (53)
	Male	14 (47)
Genotype		
	Homozygous F508del	12 (40)
	Heterozygous F508del	16 (53)
	Other	2 (7)
Clinical characteristics		
	Mean Baseline ppFEV$_1$ (range)	69.1 (35.8 – 107.8)
	Mean Baseline CRISS Score (range)	34 (0 – 52)
Medical History		
	Impaired glucose tolerance[#]	9 (30)
	Liver dysfunction[$]	2 (7)
	Allergic bronchopulmonary aspergillosis	1 (3)
	Gastroesophageal reflux	3 (10)
	Asthma	2 (7)
Culture History (one year prior to baseline)		
	Pseudomonas aeruginosa	30 (100)
	Staphylococcus aureus, MSSA	15 (50)
	Staphylococcus aureus, MRSA	2 (7)
	Burkholderia cepacia complex	1 (3)
	Haemophilus influenzae	1 (3)
	Stenotrophomonas maltophila	5 (17)
	Achromobacter xylosoxidans	1 (3)
	Nontuberculous Mycobacterium	2 (7)
	Candida spp.	4 (13)
	Aspergillus spp.	10 (33)
Other medications		
	Azithromycin	19 (64)
	Inhaled DNAse	23 (77)
	Hypertonic saline	13 (43)
	Bronchodilators	28 (93)
	CFTR modulators	8 (27)

Table 1. Characteristics of individuals in this study

Includes history of abnormal liver function tests

* Unless otherwise noted in the "Characteristic" column

Includes any history of impaired glucose tolerance, including CF-related diabetes.

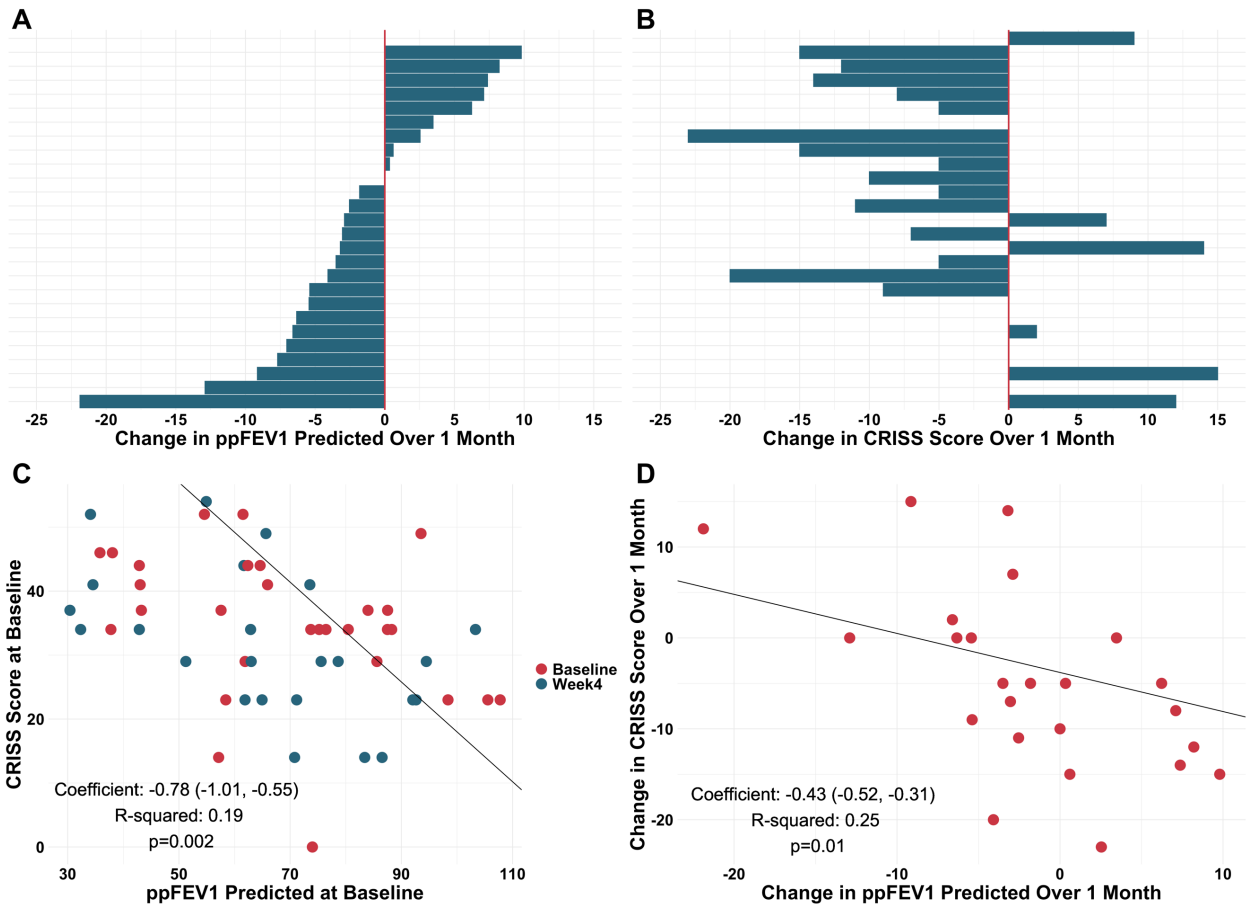

Figure 1. Clinic metrics and therapeutic response to one month of maintenance tobramycin therapy

ppFEV1 and symptom scores (CRISS) were collected from all participants at baseline and after one month of inhaled tobramycin. Change in ppFEV1 (A) and CRISS score (B) after one month of therapy is presented for each subject. Each horizontally aligned bar represents the same individual in both A and B. Positive numbers in (A) indicate improvements in lung function, while negative numbers in (B) represent improvements in symptoms. (C) Scatter plot comparing ppFEV1 and CRISS score for each individual at baseline and at Week 4. (D) Scatter plot of changes in CRISS score and ppFEV1. R2, coefficient and p values result from linear regression analysis (line), controlling for week on therapy in 1C. As a comparison, a common estimate of repeatability of standard spirometry is ~6 ppFEV1[159], and pulmonary exacerbations frequently involve a reduction of ≥10 ppFEV1 on average [160], while a change in CRISS score of 16 has been found to indicate a CF respiratory exacerbation.

and changes in those measures during therapy, including absolute cultured abundances of *S. aureus* and *P. aeruginosa*, qPCR-defined total bacterial load (TBL) and qPCR-defined absolute abundances of specific bacteria, sequencing-defined relative abundances of individual taxa, diversity measures, and normalized abundances of functional gene categories (further described in Chapter 4). As expected for this clinically stable population during maintenance therapy and a high dimensional dataset, we identified no significant associations between microbiological and clinical measures. We therefore focused on comprehensively defining how sputum microbiology changes with a cycle of maintenance inhaled tobramycin.

Primary sputum microbiological changes occurred at one week of inhaled tobramycin, with marked inter-individual variability

Culturable sputum abundances of lactose-fermenting, gram-negative bacteria (predominantly *P. aeruginosa*) and of *S. aureus* decreased an average of 1.7 and 1.1 logs, respectively, at one week of therapy. Viable counts of either plateaued or dropped relatively little after week 1, increasing thereafter during treatment, suggesting adaptation to antibiotic pressure that limited antibacterial efficacy (Fig. 3a-d). To investigate whether change in viable counts is a better measure of tobramycin therapeutic success, we also compared drop in viable count after one week in both *P. aeruginosa* and *S. aureus* to relative abundances of individual taxa, diversity measures, and normalized abundances of functional gene categories (further elaborated in Chapter 4) as described above and identified no significant associations between microbiological measures and change in viable counts of canonical CF pathogens.

To define total sputum microbiota dynamics, focusing on viable cells, we extracted DNA from all samples using Benzonase2, a method that depletes human and extracellular bacterial DNA [161] prior to sequencing described in detail in Chapter 2. In contrast to culture abundance, the mean absolute bacterial load, defined by broad-range qPCR, changed ≤0.5 logs with therapy (Fig. 2e). The different magnitudes of changes in TBL (Fig. 2e-f) and in classic pathogen culture counts (Fig. 2a-d) could in part be attributable to increased abundances of viable but unculturable cells, a state that can be induced by

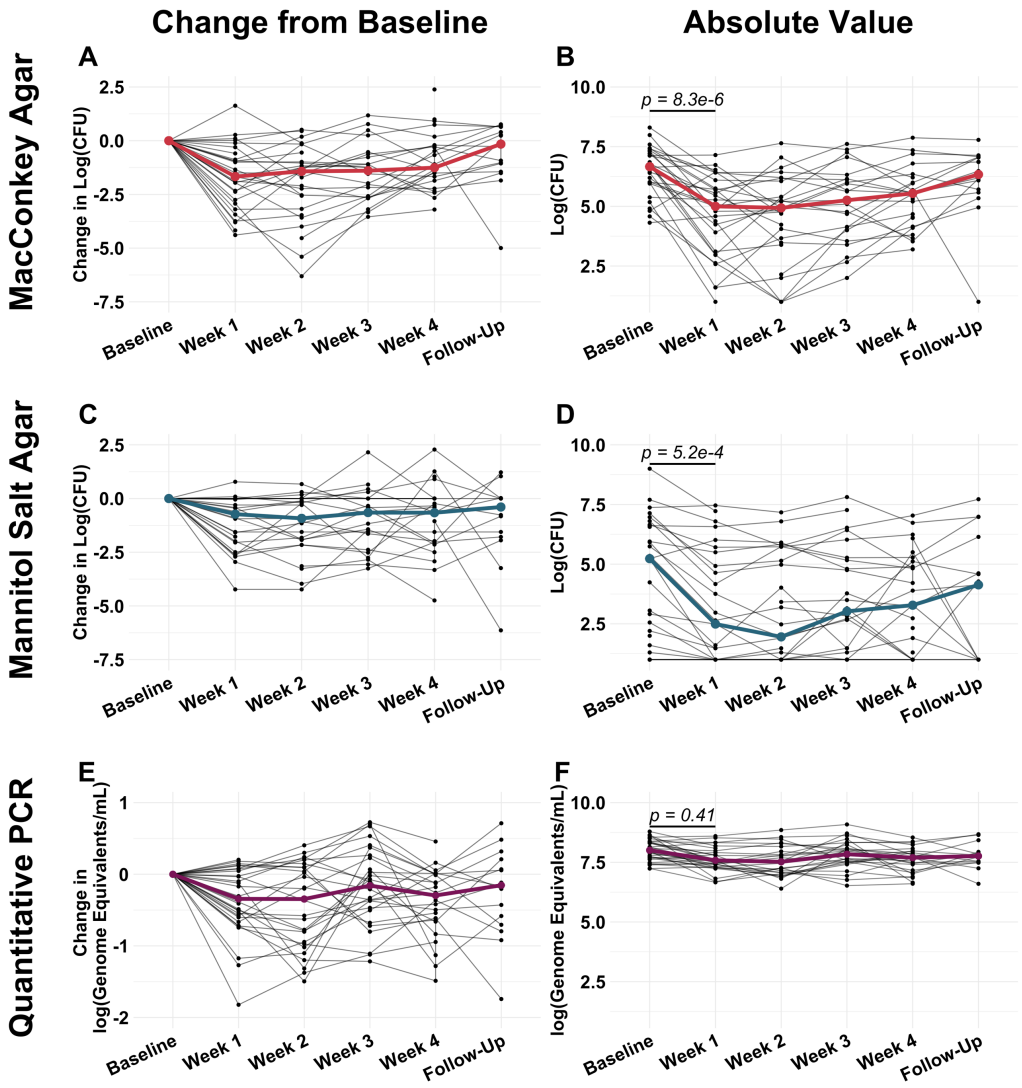

Figure 2. Inter-subject variability of sputum microbiological responses to a cycle of maintenance inhaled tobramycin was greater than intra-subject variability

A and B: Change in culturable colony counts on MacConkey agar (sequencing-based analyses demonstrated these to be predominantly P. aeruginosa with minor contributions of Serratia marcescens, Stenotrophomonas maltophila, Achromobacter xylosoxidans and an unidentified yeast taxon) from baseline (A) and absolute viable counts (B) by week on therapy. C and D similarly show change in viable counts on mannitol salt agar (S. aureus) from baseline (C) and absolute viable counts by week on therapy (D). Change in TBL from baseline (E) and absolute total bacterial load (TBL, F) by week on therapy. Black lines represent individual subjects and colored lines indicate medians. Baseline samples were collected prior to starting therapy, Weeks 1-4 represent weekly samples collected during therapy and Follow-Up samples were collected one month after cessation of therapy. All values are presented after log transformation.

antibiotic exposure [162,163], that could offset drops in DNA-based absolute abundances of canonical CF pathogens. Alternatively, DNA from dead but intact cells could have impacted PCR-based findings. To further investigate these questions at the species level we defined the taxonomic constituency of all samples using MetaPhlAn2 [111] (Fig. 3). Because metagenomic sequencing on a large scale has rarely been used previously to define CF sputum microbiota, we also sequenced a portion of samples with the more commonly-used 16S amplicon sequencing, yielding nearly identical results, albeit at the genus level compared with the species-level resolution of metagenomic analysis (Fig. 4). We then used quadruplex qPCR to measure absolute abundances of four taxa exhibiting changes during therapy: *P. aeruginosa, S. aureus, Prevotella* spp. and *Streptococcus* spp. While *P. aeruginosa* and *S. aureus* dynamics defined by qPCR were qualitatively similar to those observed from culture, the magnitudes of their qPCR-defined changes were smaller than those defined by culture (Fig. 2a-b vs. Fig. 5a-b). Notably, these differences in qPCR-defined and culturable abundances of *P. aeruginosa* and *S. aureus* have been observed before and could suggest important contributions of viable, nonculturable cells and/or residual extracellular DNA to CF sputum microbiota, as suggested previously [85,88,161]. This quadraplex qPCR was probe-based and all primer-probe combinations were present in the same reaction well, a necessity to conserve DNA sample, which may have also blunted potential changes in these taxa due to primer and/or probe competition.

Principal coordinate analysis (Fig. 6a) using the Aitchison dissimilarity metric, which is optimized for sparse, compositional data such as microbiota [164,165], identified clear distinctions among sputum bacterial microbiota between samples collected while off (Baseline and Follow-Up samples), compared with on (Weeks 1-4) tobramycin. PERMANOVA indicated species-level sputum microbiota differ significantly by week on therapy (Fig. 6a), while homogeneity of variance analysis demonstrated that these differences (distances between centroids, Fig. 6a) were not explained by differences in microbiota variances by week (distances between individual datapoints and their respective centroids). Therefore, the majority of sputum taxonomic changes with maintenance inhaled tobramycin occurred at one week of therapy. Interestingly, these treatment-emergent changes in sputum microbiota were most marked among taxa other than classic CF pathogens such as *P. aeruginosa* and *S. aureus* (Fig. 6b-c).

Figure 3. Taxonomic profile of all samples in this study

All samples were subjected to metagenomic shotgun sequencing followed and taxonomy was determined using MetaPhlAn2. Reads from the same genus (top) or species (bottom) were summed before plotting profiles for all samples, separated by individual. Each color represents a different taxon. Two subjects (202 and 209) provided an extra sample at specific weeks, which are designated “A” and “B” here.

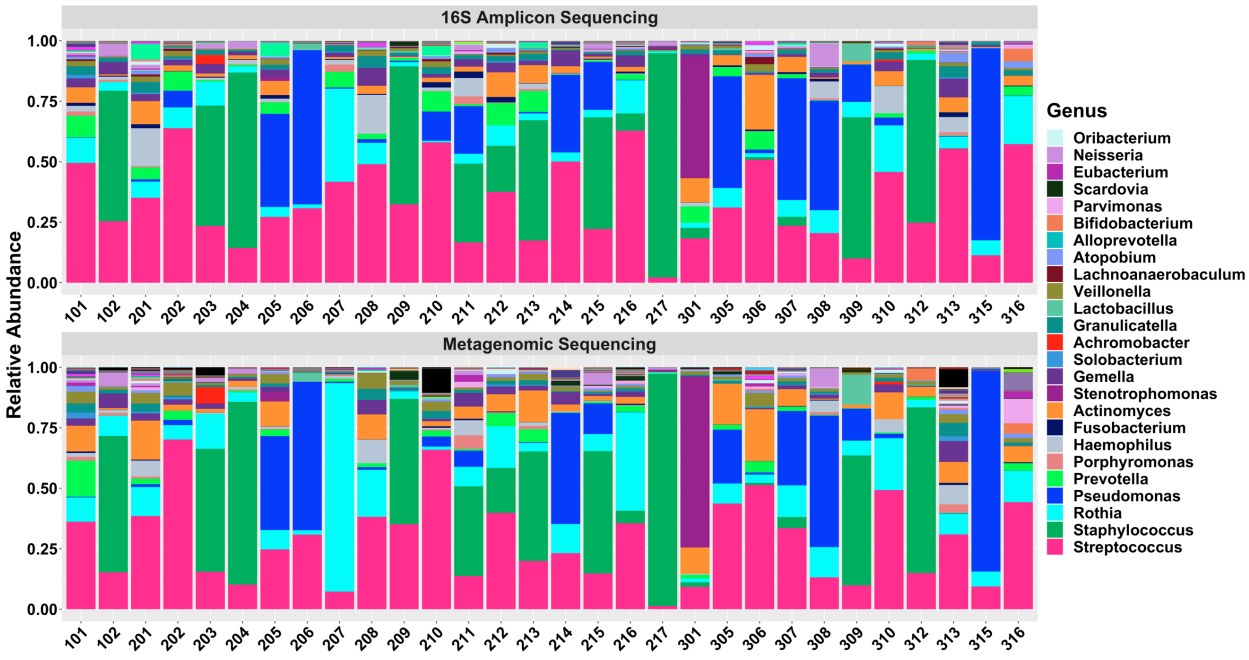

Figure 4. Comparison of taxonomic profiles analyzed via metagenomic shotgun sequencing vs. 16S amplicon sequencing

Taxonomic profiles of all 30 baseline samples at the genus level, determined using metagenomic shotgun sequencing (top) or 16S amplicon sequencing (bottom). Vertically aligned bars represent the same sample. Note that 16S amplicon sequencing does not reliably identify taxa at the species level, so only genera are displayed.

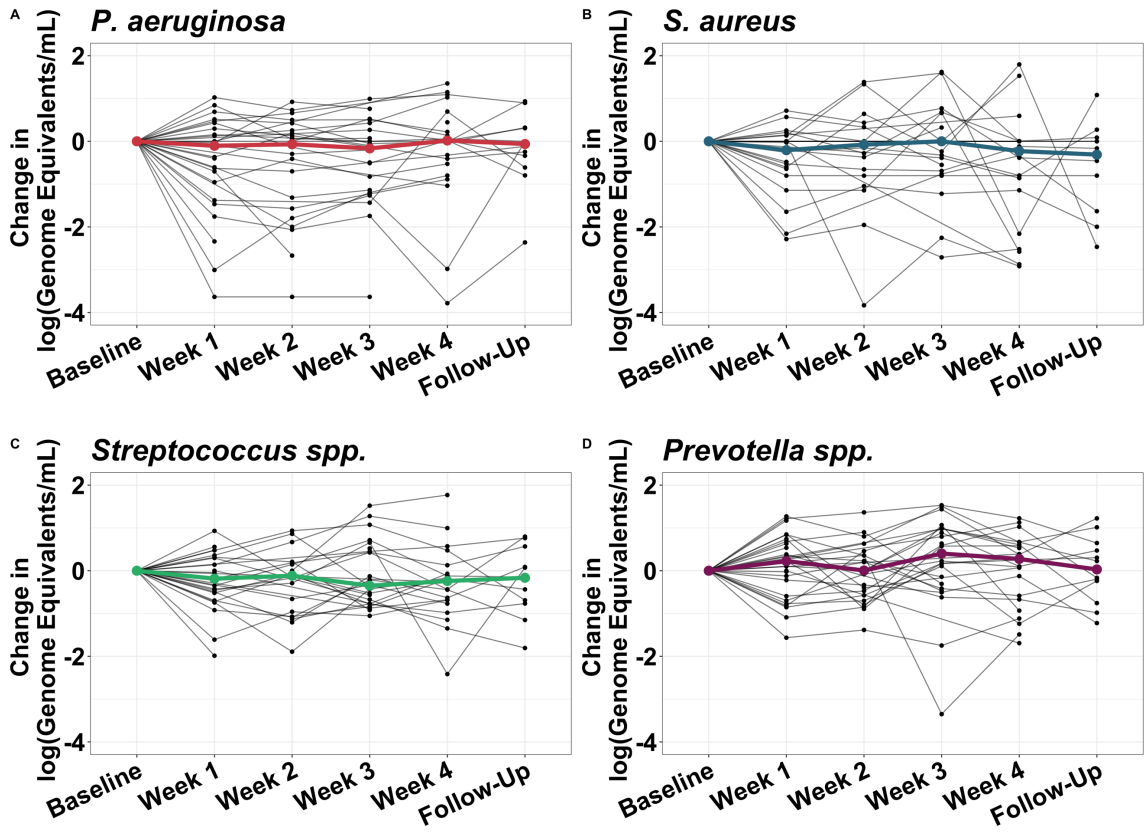

Figure 5. Change in absolute viable load of select taxa by quantitative PCR

Change in absolute load of *P. aeruginosa* (A), *S. aureus* (B), *Streptococcus* spp. (C) and *Prevotella* spp. from baseline by week on therapy. Each black line represents a single individual and the colored lines represent the median at each time point. Loads were calculated using a quadruplex quantitative PCR assay and were performed after removal of extracellular DNA from dead cells.

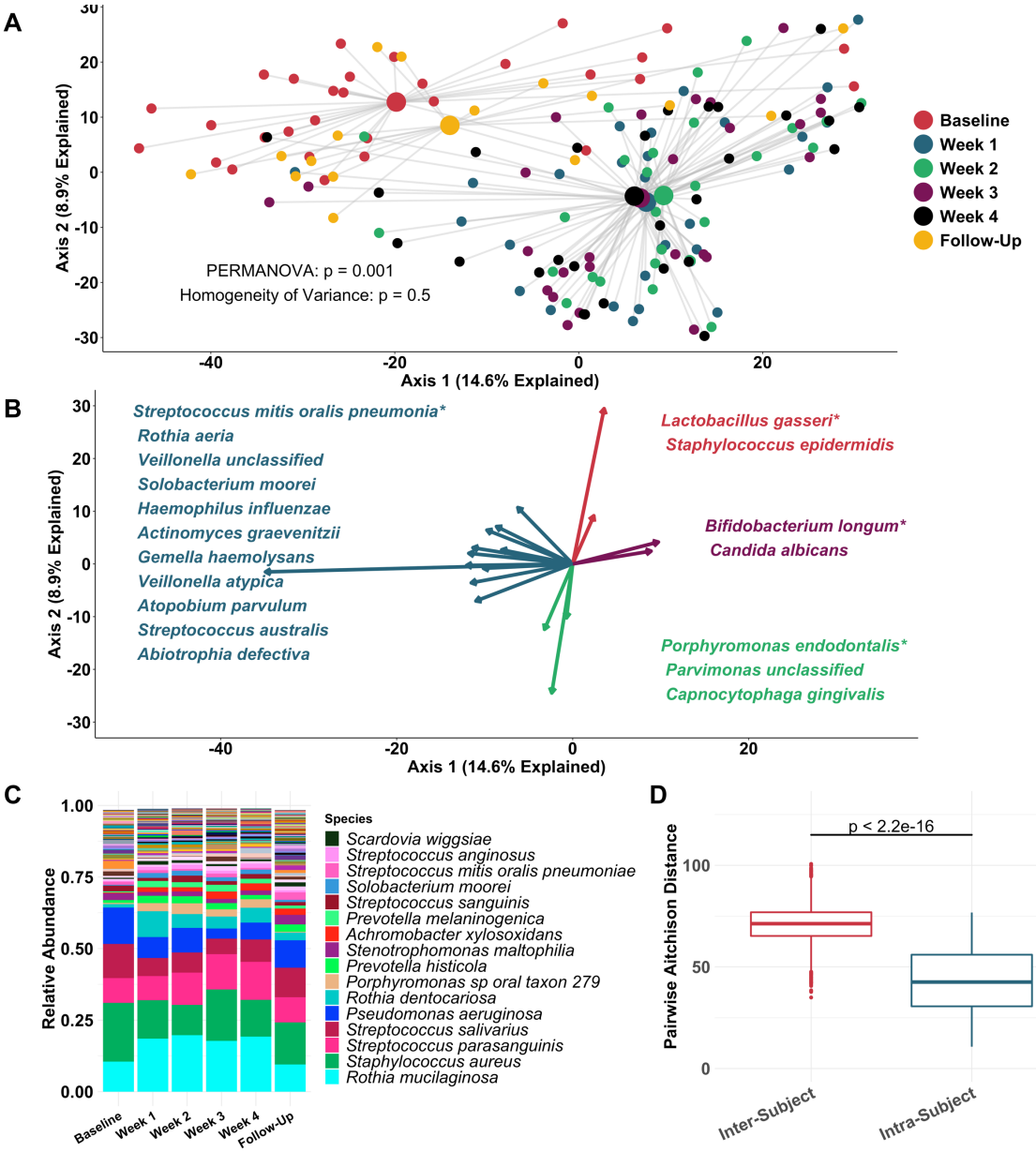
A
Axis 2 (8.9% Explained)
Axis 1 (14.6% Explained)
PERMANOVA: p = 0.001
Homogeneity of Variance: p = 0.5
Baseline
Week 1
Week 2
Week 3
Week 4
Follow-Up
B
Streptococcus mitis oralis pneumonia*
Rothia aeria
Veillonella unclassified
Solobacterium moorei
Haemophilus influenzae
Actinomyces graevenitzii
Gemella haemolysans
Veillonella atypica
Atopobium parvulum
Streptococcus australis
Abiotrophia defectiva
Lactobacillus gasseri*
Staphylococcus epidermidis
Bifidobacterium longum*
Candida albicans
Porphyromonas endodontalis*
Parvimonas unclassified
Capnocytophaga gingivalis
C
Relative Abundance
Species
Scardovia wiggsiae
Streptococcus anginosus
Streptococcus mitis oralis pneumoniae
Solobacterium moorei
Streptococcus sanguinis
Prevotella melaninogenica
Achromobacter xylosoxidans
Stenotrophomonas maltophilia
Prevotella histicola
Porphyromonas sp oral taxon 279
Rothia dentocariosa
Pseudomonas aeruginosa
Streptococcus salivarius
Streptococcus parasanguinis
Staphylococcus aureus
Rothia mucilaginosa
D
Pairwise Aitchison Distance
p < 2.2e-16
Inter-Subject
Intra-Subject

Figure 6. Sputum microbiota shifts after one week of inhaled tobramycin

Taxonomic profiles of all samples were calculated via metagenomic shotgun sequencing followed by metagenomic phylogenetic analysis. (A) Principal components analysis using the Aitchison dissimilarity metric – pairwise Euclidean distance between samples after a centered log-transformation of relative abundance data, which is optimized for sparse, compositional data such as microbiota [164,165] – of all samples (small dots) colored and grouped by week on therapy. Large dots represent the centroid of each group, with lines connecting individual sample dots to their respective centroids. PERMANOVA tested for difference between centroids, Homogeneity of Variance assessed whether dispersion in data within each timepoint (distance of each datapoint from the respective centroid) differed among groups. (B) Biplot demonstrating the 18 taxa most responsible for the taxonomic difference between samples in (A). Length of vectors indicates the extent to which taxa contribute to inter-sample dissimilarity; starred taxa are those with longest vectors in each color grouping. (C) Average taxonomic profiles of all samples by week on therapy at the species level. Only the top 14 most abundant species names are shown for ease of display. (D) Comparison of intra-subject vs. inter-subject microbiota dissimilarity at the species level by Aitchison dissimilarity, which takes into account both abundances and presence of individual taxa. Baseline samples were collected prior to starting therapy, Weeks 1-4 represent weekly samples collected while individuals were on therapy and Follow-Up samples were collected one month after cessation of therapy. Wilcoxon signed rank test was used to assess difference between groups. Taxonomic profiles of all samples individually are presented in Figures S4-5. Boxes represent interquartile region and middle represents the median.

Inspection of individual, sequencing-based taxonomic data at the genus and species levels (Fig. 3) demonstrated marked inter-subject variability in these treatment-emergent changes. To quantify this observation, we defined the diversity of sputum microbiota both within and between subjects, demonstrating much higher diversity between than within subjects (Fig. 6d): sputum taxonomic shifts with antibiotic perturbation were of a smaller magnitude than differences between any two samples from different subjects. We also calculated change in Shannon and Simpson diversity indices, evenness and richness for all samples by week on therapy (Fig. 7). Despite substantial inter-subject variability in these measures, we observed a statistically significant decrease in average species richness after one week of therapy, returning to baseline levels upon follow-up. We identified no association between microbial read depth and Shannon or Simpson index, and a paradoxical negative association between species richness and microbial read depth, indicating that sequencing depth did not impact these results (Fig. 8). We identified a number of microbial metrics, via both culture and molecular methods, that indicate that tobramycin exerts its major effect on the microbiota by just one week of therapy, that these changes are largely maintained during the treatment period, and that the microbiota rebound after antibiotic pressure is removed.

Sputum taxonomic shifts are driven by non-dominant, facultative and obligate anaerobes

To more rigorously identify which taxa changed most in sputum abundance with maintenance tobramycin treatment, we compared the sequencing-defined microbiota at baseline with those at week 1, when most changes occurred (Fig. 9a). Although *P. aeruginosa* was the intended target of inhaled tobramycin therapy [76], we found the most significant sputum relative and calculated absolute abundance changes occurred among non-dominant, low abundance taxa (Fig. 9b-c). In fact, no classic CF pathogen abundance changed significantly with tobramycin by molecular analysis (Fig. 9b-c, Fig. S9), although the calculated abundance changes of both *P. aeruginosa* and *S. aureus* at week 1 were similar to those we identified using traditional culture (Fig. 2b,d vs. Fig. 9c). To determine how specific taxa related to clinical response, we repeated the above analysis after partitioning subjects into dichotomous “responder” vs. “non-responder” categories (wherein “responders” were subjects in the upper tertile of $ppFEV_1$ change equivalent to >0 percentage point increase or lower tertile of CRISS change equivalent to < -8 point

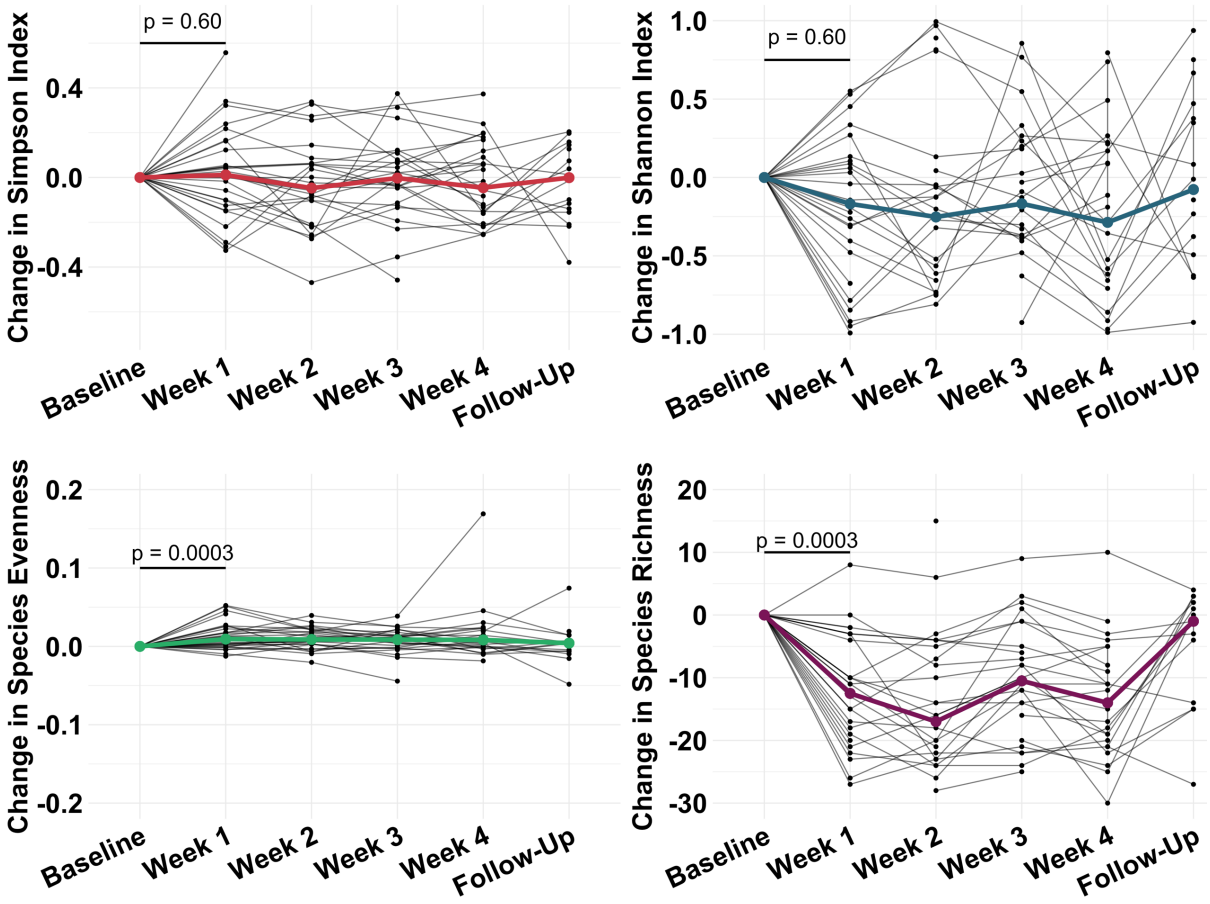

Figure 7. Diversity metrics of all samples by week on therapy

Change in Simpson index (A), Shannon index (B), evenness (C) and Richness (D) by week on therapy. Each black line represents a single individual and colored lines represents the median value. All values were calculated at the species level. Wilcoxon signed rank test was used to assess difference between groups.

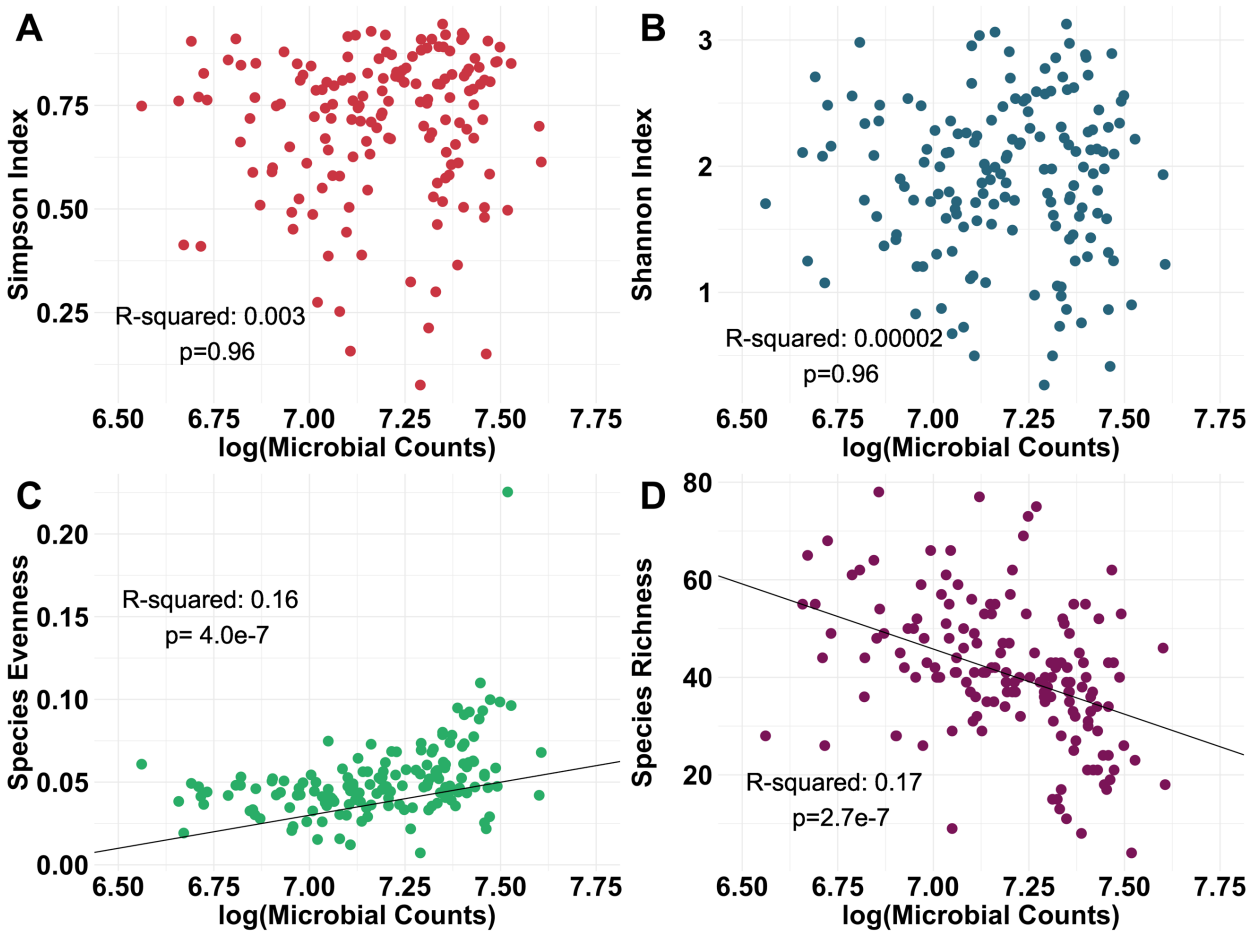

Figure 8. The effect of microbial sequencing depth on community diversity.

Dot plots comparing log microbial reads and Simpson index (A), Shannon index (B), evenness (C) and richness (D). Note that microbial reads, rather than total reads, are presented as percent human DNA differed between samples. R^2 and p values result from linear regression analysis (line). A Benjamini-Hochberg correction performed to account for multiple comparisons.

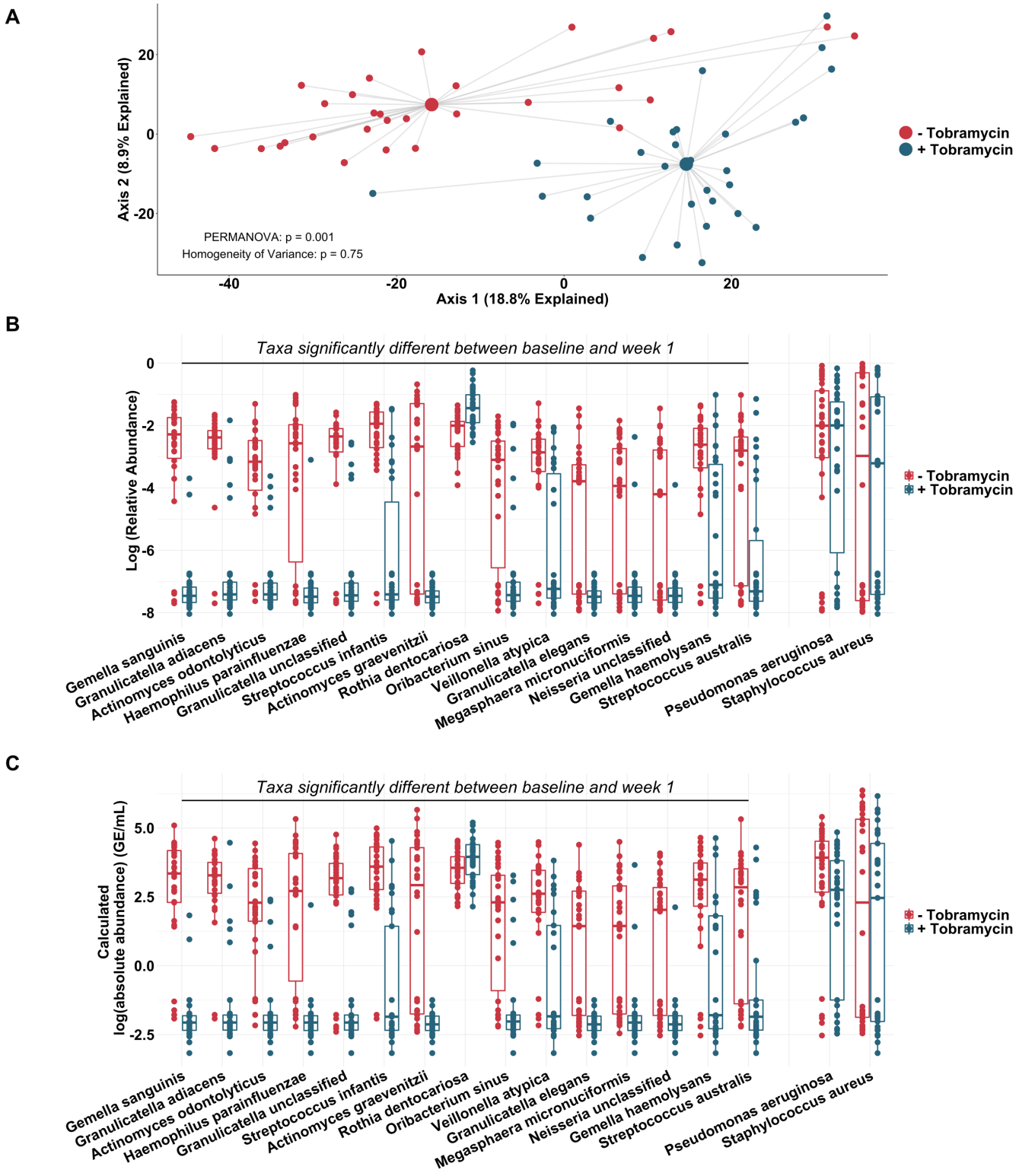
A
Axis 2 (8.9% Explained)
20
0
-20
PERMANOVA: p = 0.001
Homogeneity of Variance: p = 0.75
-40
-20
0
20
Axis 1 (18.8% Explained)
- Tobramycin
+ Tobramycin
B
Taxa significantly different between baseline and week 1
Log (Relative Abundance)
0
-2
-4
-6
-8
Gemella sanguinis
Granulicatella adiacens
Actinomyces odontolyticus
Haemophilus parainfluenzae
Granulicatella unclassified
Streptococcus infantis
Actinomyces graevenitzii
Rothia dentocariosa
Oribacterium sinus
Veillonella atypica
Granulicatella elegans
Megasphaera micronuciformis
Neisseria unclassified
Gemella haemolysans
Streptococcus australis
Pseudomonas aeruginosa
Staphylococcus aureus
- Tobramycin
+ Tobramycin
C
Taxa significantly different between baseline and week 1
Calculated log(absolute abundance) (GE/mL)
5.0
2.5
0.0
-2.5
Gemella sanguinis
Granulicatella adiacens
Actinomyces odontolyticus
Haemophilus parainfluenzae
Granulicatella unclassified
Streptococcus infantis
Actinomyces graevenitzii
Rothia dentocariosa
Oribacterium sinus
Veillonella atypica
Granulicatella elegans
Megasphaera micronuciformis
Neisseria unclassified
Gemella haemolysans
Streptococcus australis
Pseudomonas aeruginosa
Staphylococcus aureus
- Tobramycin
+ Tobramycin

Figure 9. Non-dominant taxa contribute substantially to taxonomic shift with therapy

Taxonomic profiles of all samples defined by metagenomic sequencing and phylogenetic analysis. (A) Principal component analysis using the Aitchison dissimilarity metric of all baseline and Week 1 samples (small dots) colored and grouped by antibiotic treatment status. Large dots represent the centroid of each treatment category, with lines connecting individual sample dots to their respective centroids. (B) Log-relative abundances and (C) calculated absolute abundances of the 15 taxa that contributed most to the differences between baseline and week 1 samples identified in (A) ordered by effect size as well as *P. aeruginosa* and *S. aureus* for comparison. PERMANOVA is a statistical test for difference between centroids, Homogeneity of Variance is a statistical test which assessed the difference in spread between two groups. Absolute abundances were calculated by multiplying relative abundances via MetaPhlAn2 by total bacterial loads as determined via universal 16S qPCR. Boxes represent interquartile region and middle represents the median.

decrease). We identified no significant differences between groups in either baseline sputum microbiota (Fig. 10a) or their change after Week 1 (Fig. 10b); parallel analyses identified no associations with symptomatic responses (Fig. 11a-b). Interestingly, we did observe trends towards correlations between responder status and higher baseline sputum abundances of P. aeruginosa, despite minimal change in this species with therapy using either $ppFEV_1$ (Fig. 10c-d) or symptomatic response (Fig. 11c-d).

Discussion

Using metagenomic sequencing, qPCR, and standard culture, we defined the CF sputum microbiota response to a cycle of maintenance inhaled tobramycin treatment. While this therapy is primarily intended to target *P. aeruginosa,* the most significant abundance changes with treatment occurred among non-dominant taxa, none of which are considered classic CF pathogens. As expected in this study of maintenance treatment in clinical stable individuals, the study population experienced modest changes in both objective (spirometry) and subjective (symptom score) clinical measures during treatment and we identified no significant correlations between clinical and microbiological responses in this study. These findings suggest that the primary sputum microbiological effect of tobramycin maintenance therapy is on non-targeted bacteria, offering avenues both to better define the microbial determinants of clinical response to therapy and to devise more effective treatment strategies for a broad range of people with CF and other chronic infections.

The majority of the sputum microbiological change occurred by the first week of therapy. By culture, absolute sputum abundances of traditional CF pathogens (*S. aureus* and gram-negative organisms, predominantly *P. aeruginosa*) changed most by week 1 and were followed by either a plateau or gradual return towards baseline levels, suggesting gains made earlier during treatment subsequently diminished. Sputum microbial changes identified by molecular methods were qualitatively similar to culture, but with lower magnitudes; absolute sputum loads of all bacteria and of common taxa (*S. aureus*, *P. aeruginosa*, *Streptococcus spp.* and *Prevotella spp.*) changed very little with treatment. Quantitative discrepancies between culture and molecular analyses of CF sputum have been observed before [166]

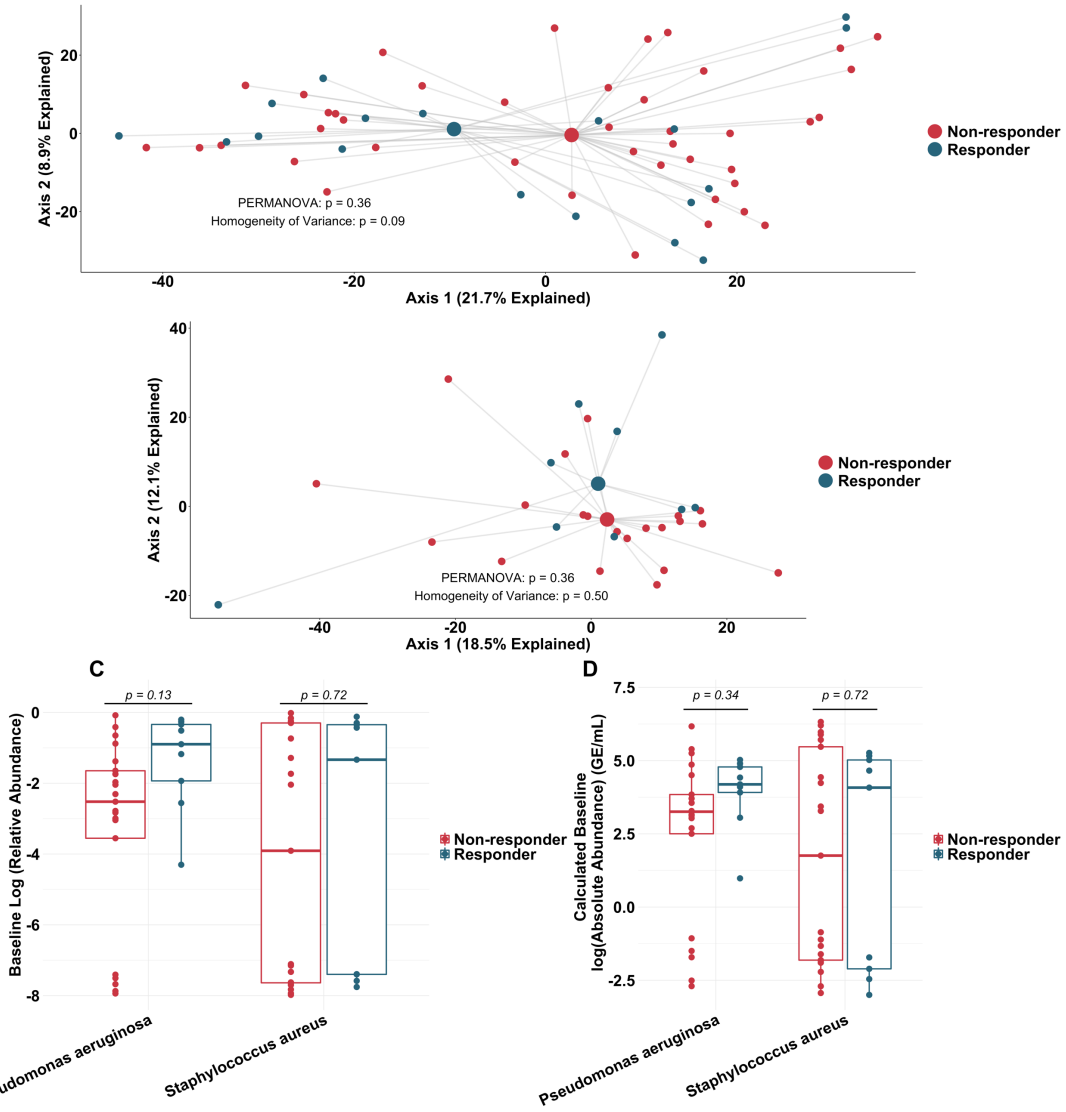

Figure 10. Similarity in microbial community between "Responders" and "Non-responders" to tobramycin using ppFEV1

A) Principal component analysis using the Aitchison dissimilarity metric of all baseline samples (small dots) colored and grouped by response status. Large dots represent the centroid of each group, with lines connecting individual sample dots to their respective centroids. (B) Principal component analysis using the Aitchison dissimilarity metric of changes in relative abundance between baseline and week 1 samples (small dots) colored and grouped by response status. Large dots represent the centroid of each group, with lines connecting individual sample dots to their respective centroids. PERMANOVA is a statistical test for difference between centroids, Homogeneity of Variance is a statistical test which assessed the difference in spread between two groups. (C-D) Difference in relative abundance (C) and calculated absolute abundance (D) of P. aeruginosa and S. aureus at baseline between "Responders" and "Non-responders." Boxes represent interquartile region and middle represents the median.

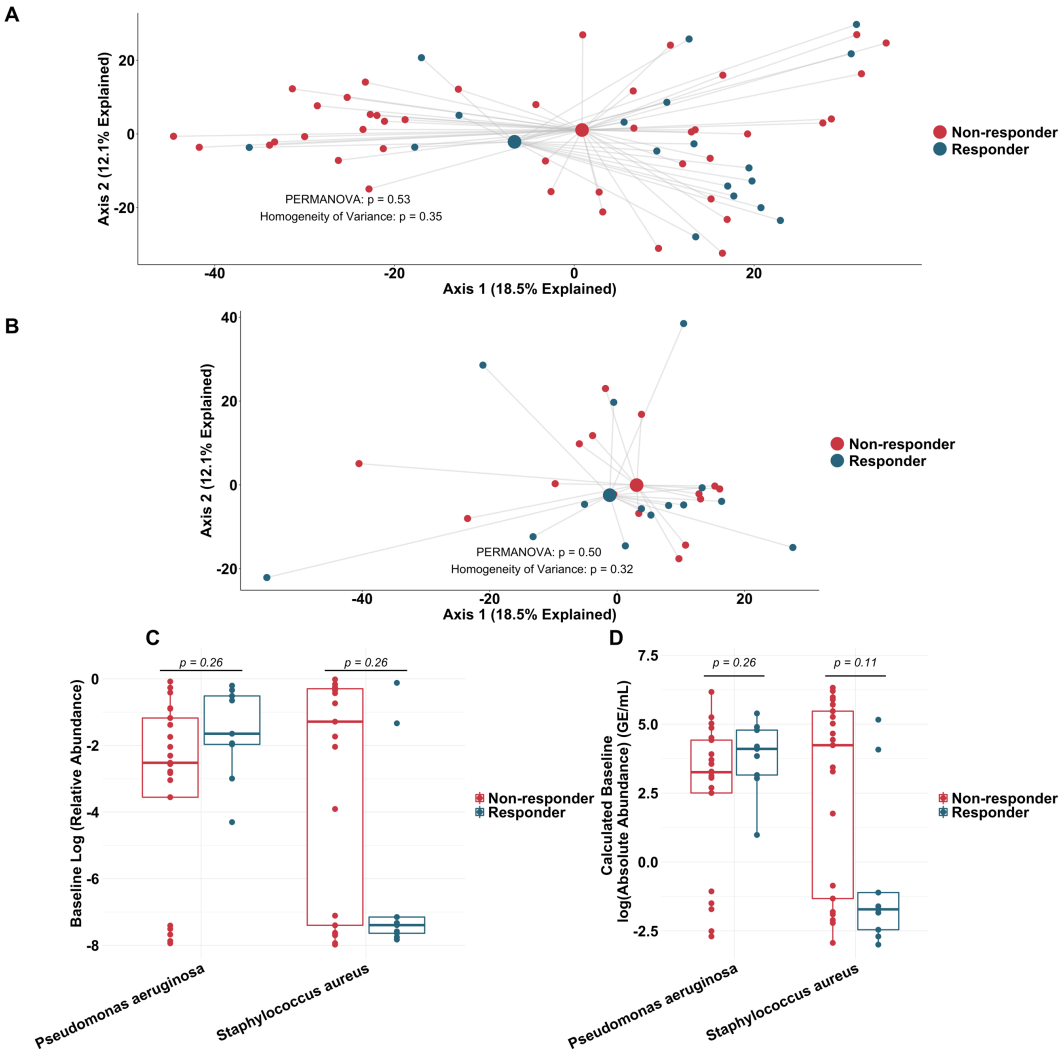

Figure 11. Similarity in microbial community between "Responders" and "Non-responders" to tobramycin using CRISS symptomatic score

A) Principal component analysis using the Aitchison dissimilarity metric of all baseline samples (small dots) colored and grouped by response status. Large dots represent the centroid of each group, with lines connecting individual sample dots to their respective centroids. (B) Principal component analysis using the Aitchison dissimilarity metric of changes in relative abundance between baseline and week 1 samples (small dots) colored and grouped by response status. Large dots represent the centroid of each group, with lines connecting individual sample dots to their respective centroids. PERMANOVA is a statistical test for difference between centroids, Homogeneity of Variance is a statistical test which assessed the difference in spread between two groups. (C-D) Difference in relative abundance (C) and calculated absolute abundance (D) of P. aeruginosa and S. aureus at baseline between "Responders" and "Non-responders." Boxes represent interquartile region and middle represents the median.

and could be attributable to viable, nonculturable bacteria (undermining the accuracy of culture) or DNA from dead cells (particularly problematic for qPCR and sequencing). Our use here of a method that depletes extracellular DNA, including that from host and lysed bacterial cells, mitigates the latter contribution. It is also possible primer-probe competition could have blunted observed changes in individual taxa in this quadraplex assay. Further research will be required to quantify and distinguish the contributions of these two effects; however, prior work showing that antibiotics can induce bacteria to enter unculturable states [167] suggest this mechanism could contribute to our observations.

As found in studies of antibiotic therapy for CF exacerbations, the sputum microbiota in this population differed more between than within subjects, even during treatment. Nevertheless, we identified significant, treatment-emergent changes in sputum microbiota that were not evident in prior studies, likely due to our study design focusing on a single antibiotic and a defined treatment and specimen collection schedule [29,30]. Surprisingly, non-dominant, prevalent taxa contributed most to sputum taxonomic shifts with maintenance tobramycin treatment. Non-dominant taxa therefore represent a conserved feature among CF sputum microbiomes; while the taxa involved differed between individuals, their presence and dynamic responses to maintenance tobramycin varied much less than did those of the dominant taxa. A relationship between taxonomic and functional microbiome characteristics has been observed for diverse microbial communities [168,169], suggesting CF sputum microbiota share characteristics with other well-established microbiota, a relationship we will further explore in Chapter 4.

The non-dominant taxa that shifted most in abundance with maintenance tobramycin therapy were facultative and obligate anaerobes; most are also known oral microbiota constituents. These taxa have been associated with healthy airways [26], and their abundances have been noted to rise in CF respiratory samples prior to pulmonary exacerbations [170,171]. Some of these taxa have also been implicated in worsening CF lung disease [59,172]. It is especially interesting to note that both the relative and absolute abundances of most anaerobic taxa we identified decreased after one week of therapy, rather than simply rising in relative abundance as those of dominant taxa decrease. Aminoglycosides, including tobramycin, have traditionally been thought to primarily impact aerobic taxa, as aerobic respiration efficiently drives electron transport-dependent aminoglycoside uptake [63]. However, prior

work has identified activity of aminoglycosides, albeit limited, under low-oxygen conditions [66,67], which are known to occur in CF secretions [170], including in *Haemophilus influenzae* and *Streptococcus spp* [173]. There is little specific data on aminoglycoside susceptibility in many of the taxa we identified as have significantly lower relative abundance after one week of antibiotic therapy and it is possible susceptibility of clinical CF isolates differs from that of reference strains. Further *in vitro* antibiotic susceptibility testing of such non-canonical CF pathogens is required to determine the effect tobramycin may have on these taxa. Our data do not distinguish whether tobramycin had direct antimicrobial effects on specific taxa, as opposed to indirect effects through other, interacting microbes. If tobramycin is able to directly target these lower-abundance taxa, it remains unclear from our data how these effects are related, if at all, to the long-term morbidity and mortality benefits of this common therapy. It is possible that such effects are directly related to the pathogenesis of CF lung disease, that reducing their numbers has secondary effects on canonical pathogen loads with downstream ramifications on pathogenesis or that they are entirely unrelated to clinical benefit. Regardless, the clinical consequences—both short- and long-term—of these unintended effects require further study in order to translate these findings into clinical practice.

Our work has a number of limitations. Although we did detect a number of DNA viruses (primarily phage) and fungi (primarily *Candida* and *Aspergillus*) our DNA extraction method was not optimized to detect either of these kingdoms and our sequencing platform cannot detect RNA viruses. Thus, it is likely we did not detect the full complement of these taxa in our samples and thus our analysis primarily focuses on bacteria. Any bioinformatic method for taxonomic classification is limited by the fidelity of its database. Further work should be done to verify the species identities of the taxa we found to be significantly different between baseline and week 1 samples. Sputum variably samples the entire respiratory tract (including the oral cavity and upper respiratory tract). There is debate regarding how well sputum reflects lower airway microbiota, whether sputum always samples the same anatomic locations, despite day-to-day reproducibility [29], the extent to which oral bacteria are introduced to sputum during expectoration or whether all oral taxa detected in sputum represent microbes residing in the lungs, transiently aspirated from the upper respiratory tract, or oropharyngeal contamination during expectoration [35–37]. Tobramycin concentrations in the oral cavity are likely high with inhaled therapy

and this treatment therefore likely impacts the oropharyngeal microbiome. Because this study was not designed to answer these questions, we report our results with respect to sputum microbiomes, without inferring relationships with lower airways. Samples were stored for 2-48 hours at 4℃ prior to processing and long-term storage at -80℃; a schema which we chose to maximize subject participation. Prior studies have indicated minimal changes in viable counts of *P. aeruginosa* [174] or microbiota composition [175] with extended incubation at 4℃, but alterations in taxonomic or functional constituency of our samples with this extended collection time remains a possibility, although this is unlikely to have affected differences between baseline and Week 1 samples. Further studies may benefit from immediate sample processing after collection. Furthermore, all subjects were clinically stable during the study, and the majority (24/30) were not naïve to tobramycin maintenance. Of the 30 participants recruited to the study, 4 did not complete the study, of whom 2 discontinued participation due to increased symptoms requiring alternative antibiotic therapy. We chose these exclusion criteria to ensure all participants were clinically stable at the time of sample collection (limiting potential microbiota changes not attributable to antibiotic therapy). It is possible that these missing week 4 samples spuriously skewed the clinical response to tobramycin towards the null (i.e. by removing individuals with adverse responses to drug, we artificially enriched for individuals that had minimal or positive clinical response to tobramycin), which in turn may have further limited our power to detect microbial metrics associated with differential clinical response. Therefore, results may not reflect the effects of tobramycin on respiratory sample microbiomes from people during initial treatment or at later stages of disease.

Our study also has notable strengths compared with prior work in this area. All participants were on only tobramycin, apart from maintenance azithromycin, for the duration of the study. Samples and data were collected at standardized time points with respect to treatment period, with weekly sampling during treatment. Variation in treatment and collection procedures were significant limitations of previous studies and may have limited resolution [29,59,91,176,177], and the relative homogeneity of our study design likely augmented our discriminatory power. While this focus may limit generalizability to other antibiotics or diseases, tobramycin is the antibiotic used most often to treat *P. aeruginosa* infections in people with CF, representing a large population of affected patients. Our results indicate the power of using

standardized study designs to identify significant microbiome changes with treatment despite inter-patient variability.

Finally, this study represents one of the largest reported analysis of CF sputum metagenomes, and our data complement previous smaller studies which have characterized the CF sputum microbiota in cross-sectional samples via metagenomic sequencing and found it to harbor diverse taxa not typically identified by classic clinical culture and marked inter-subject diversity [90,96,97,178]. In this study, we used metagenomic sequencing to specifically define the CF sputum microbial and functional genomic capacities (Chapter 4) for a large number of samples and subjects over a single maintenance course of inhaled tobramycin, providing a rich view of the effects of tobramycin and identifying significant similarities across samples and with antibiotic perturbation. Metagenomic sequencing also allows for species-level resolution, compared to genus-level with 16S amplicon sequencing, and we also identified significant treatment-emergent abundance changes in non-dominant taxa, perhaps underlying the well-known resilience of CF respiratory microbiota to therapy [29,58,177] and suggesting a potential role for tobramycin therapy for a broader population of patients with CF and related infections. While our study did not suggest any mechanisms of the clinical effects of maintenance inhaled tobramycin, our data showed the largest microbiological impact of this therapy to be on non-dominant taxa, suggesting a causal role for antibiotics for the decreasing microbial diversity that has been observed in CF respiratory samples with increasing age, disease severity, and antibiotic exposure [59,179]. These findings indicate the power of metagenomic sequencing and longitudinal, controlled study design during treatments such as antibiotics, anti-inflammatories, and CFTR modulators to better define the microbial determinants of CF lung disease and treatment responses.

Methods

Study design and Description of Subjects

The TIP study population comprised 30 people with CF who were spontaneously expectorating at three clinical sites, the University of Washington Medical Center, Seattle Children's Hospital and University of Michigan Health System. Spontaneously expectorated sputum was collected prospectively from these

subjects at commencement of TIP (“baseline”) and weekly during the month-long treatment period, which occurred after at least 28 days without exposure to any antibiotics other than maintenance azithromycin. A subset of participants also provided a follow-up sample one month after completion of the 28-day-long TIP treatment cycle (“follow-up”). Study inclusion criteria were as follows:

- Documented diagnosis of CF, based on standard criteria established by the Cystic Fibrosis Foundation
- Age > 6 years
- Able to reproducibly perform spirometry for clinical care
- Able to expectorate at least daily
- History of recurrent P. aeruginosa positive respiratory cultures, defined as 2 or more occurrences prior to enrollment
- Respiratory culture positive for P. aeruginosa at least once within the 12 months prior to enrollment
- Abstinence or documented use of contraception by females of child-bearing potential at baseline
- Signed informed consent, and assent, as applicable
- FEV_1 greater than 25% predicted at baseline
- Study exclusion criteria were as follows:
- History of intolerance of inhaled tobramycin
- In the opinion of CF care provider, judged not to be able to tolerate 28 days without antibiotics other than chronic azithromycin
- Treatment with any antibiotics (other than chronic azithromycin) within the prior 28 days to first sample collection
- Initiation of chronic azithromycin within the prior 28 days of first sample collection
- Requiring hospitalization or treatment with antibiotics in addition to inhaled tobramycin during the study period
- Positive urine or serum pregnancy test at baseline

- Participating in an investigational drug trial that is determined by the investigators to conflict with this study's goals

Of the 30 participants recruited to the study, 4 did not complete the study: one was lost to follow up, one decided to leave the study early for unknown reasons, one did not tolerate tobramycin and was switched to alternative inhaled and oral antibiotics, one was started on intravenous antibiotics due to increased symptoms. We chose these exclusion criteria to ensure all participants were clinically stable at the time of sample collection (limiting potential microbiota changes not attributable to antibiotic therapy). Sputum samples not collected during stability were excluded from the study.

Sputum specimen processing

All samples were shipped on ice within 48 hours of collection [175] and sputum was homogenized with dithiothreitol to liquefy the specimen as described below. After culture, the remainder of the specimen was frozen for DNA analysis. Storage at 4C has been shown to preserve CF sputum microbiota profiles compared with -80℃ [175]. The 48-hour window was chosen to allow participants to send in samples from home (maximizing recruitment in the study) as well as to allow samples to be available by mail from three different clinical sites in two cities, (University of Washington Medical Center, Seattle Children's Hospital and University of Michigan Health System), broadening generalizability. Sputum was also frozen in glycerol to allow for future culturing of non-canonical organisms where volume allowed. Subject details are present in Table 1.

Sputum culture for standard pathogens

An aliquot of each sample was cultured on differential media according to standard practice [180] to calculate total culturable loads of gram negative organisms (predominantly P. aeruginosa) and S. aureus. Viable loads of each taxon between baseline and Week 1 were compared using a Wilcoxon rank sum test.

Initial sample processing and bacterial culture

All samples were either collected at a study visit followed by transfer to the lab or mailed directly the lab by the study participant on ice. An aliquot of each sample was cultured on differential media according to standard practice [180] to calculate total culturable loads of gram negative organisms and of S. aureus as follows. Samples were diluted 1:1 in 0.1% Sputolysin (DTT; Millipore Sigma), serially diluted in PBS and plated on blood agar (Fisher), MacConkey agar (BD) and mannitol salt agar (BD) according to standard practice [180]. Plates were incubated at 35°C for 24-48 hrs prior to quantifying the abundances of both S. aureus and select gram negative organisms (MacConkey agar, predominantly P. aeruginosa but counts also including Achromobacter xylosoxidans, Stenotrophomonas maltophila, Serratia marcescens and an unidentified yeast taxon determined by microbiota analysis). Colonies on mannitol salt agar were confirmed as *S. aureus* either by mannitol fermentation or nucA PCR [181]. All colonies on MacConkey agar were counted as gram-negative organisms. Representative isolates were collected and frozen for future analysis.

Spirometry and symptom scores

Standard spirometric measures of lung function, including forced expiratory volume in 1 second (FEV_1) and forced vital capacity (FVC), were recorded before and after therapy according to standard methods [182]. Participants' subjective responses were collected before, during and one month after treatment, using the standardized Chronic Respiratory Infection Symptom Score (CRISS) [183], which quantifies symptoms such as cough and fatigue.

DNA extraction

All samples were extracted from frozen, DTT-treated sputum using a method designed to minimize human DNA and extracellular bacterial DNA as detailed previously [161]. 300-450mL of DTT-treated sputum was suspended in 7 mL dH_2O and incubated at room temperature for one hour with gentle agitation. 10x strength Benzonase buffer (200 mM Tris-HCl, 10 mM $MgCl_2$) to a final 1X and 250U Benzonase (Sigma E-1014) were added and the sample was incubated at 37°C for 2 h with gentle agitation. 1.5M NaCl (150 mM final) was added and the resulting solution was centrifuged at 8,000g for

10min and the pellet washed once with PBS, centrifuged at 13,000g for 4min, the pellet resuspended in 400µL TE, and 0.5M EDTA (5 mM final) was added to inactivate the endonuclease. A mixture of 1 mm and 0.1 mm silica:zirconia beads and a single tungsten-carbide bead was added to the TE solution, followed by bead-beating for one minute in a BioSpec MiniBeadBeater. The resulting solution was boiled for 5 min at 95°C. Lysozyme (Sigma L6876, 3 mg/mL final) and lysostaphin (Ambi LSPN, 0.14 mg/mL final) were added, and the sample was incubated for one hour at 37°C. Proteinase K (Invitrogen 25530049, 1.4 mg/mL final) and SDS (1.8% final) were added, and the resulting solution incubated at 56°C for 30 min before cooling to room temperature. The solution was removed to a separate tube and 5 M NaCl was added (2 M final) before adding phenol:chloroform:isoamylalcohol (25:24:1) at a 1:1 volume. The solution was then incubated for 20 min at room temperature, centrifuged at 13,000 g for 20 min[17] and the top aqueous layer was collected. 0.133 volume equivalent of 7.5 M ammonium acetate was added to the aqueous layer, and the resulting solution diluted 1:1 with cold 100% ethanol to precipitate DNA. DNA product was cleaned with a spin column (BioBasic BS423, skipping the cell lysis steps and proceeding directly to DNA clean-up steps). Reagent blanks consisting of PBS alone were processed and sequenced via 16S amplicon sequencing to assess contamination.

Quantitative PCR

Total bacterial load was determined by universal 16S quantitative PCR with PowerUp SYBR Green Master mix (Applied Biosystems A25742) using previously published primers and reaction conditions [139]. Proportions of human and bacterial DNA in samples was determined by calculating the genome equivalents (GE) of each, using $5x10^6$ bp for the average microbial genome size. Based on the following commonly-used definition of limit of detection (LOD) for qPCR [141,142], “the lowest copy number associated with the serial dilution that gave a positive PCR response on 95% of occasions,” the LOD of our universal 16S qPCR assay was approximately 8.24 GE per uL. In addition to all sputum samples, we calculated total bacterial loads of 12 extraction blanks. Average loads in extraction blanks were ~4 logs lower than bacterial loads in samples and the lowest sample bacterial load was >2 logs higher than the highest extraction blank bacterial load. Taxon-specific qPCR was determined using previously published primers-probe combinations against the femA gene for S. aureus [184], the gyrB gene for P. aeruginosa

[185] and genus-specific primers against the 16S genes for both Streptococcus [186] and Prevotella [187]. All primer-probe combinations were present in the same reaction well and assessed in the same machine run.

Data were analyzed with the Bio-Rad CFX Maestro 1.1 (Version 4.1.2434.0124), using software-defined Cq thresholds and total bacterial load between baseline and Week 1 was compared using a Wilcoxon rank sum test.

Phylogenetic composition from metagenomic shotgun sequencing

Next-generation sequencing libraries were prepared for all samples using the Nextera DNA Sample Prep Kit (Illumina FC-121-1031) following manufacturer's instructions. All samples were sequenced on the Illumina HiSeq platform, producing an average of 3.3 $x10^7$ total reads per sample after quality filtering. Sequencing data from all samples were de-duplicated using SeqUniq (version 0.1) https://github.com/standage/sequniq) and quality filtered using KneadData (version 0.6.1) and Trimmomatic (version 0.33) [143], producing an average of 2.9 $x10^7$ reads per sample. Human reads, comprising 12%-95% of our samples based on the efficiency of the human DNA depletion step in our DNA extraction method described in Chapter 2 (Table 2), were identified and removed with BMTagger (version 3.101), resulting in 1.3$x10^7$ microbial reads per sample on average. Due to the range of human DNA load across our samples, we were unable to use this diploid genome to normalize our microbial read abundances as has been proposed elsewhere [188]. Community phylogenetic composition was determined using MetaPhlAn2 [111,131] (version 2.2.0) to produce a MetaPhlAn2 taxa table. MetaPhlAn2 utilizes a small set of taxon-specific bacterial marker genes for taxonomic assignment and excludes genes from accessory genomic elements or paralogs, which cannot be reliably attributed to a specific taxon; thus it prioritizes specificity over sensitivity in contrast to other bioinformatic pipelines which determine the taxonomic origin of all reads [189–191]. We utilized such a marker-gene approach for taxonomic designation to reduce the spurious detection of low-abundance taxa. All commands were executed with default settings, with the exception of KneadData, which was used with the "–run-bmtagger" flag. For all samples, the Achromobacter genus represents reads originally annotated either as

Achromobacter or Bordetella, the latter of which proved to reflect misidentified Achromobacter reads: Using mock communities, we found MetaPhlAn2 consistently splits this taxon into Achromobacter and Bordetella, even when this latter taxon was not present, as confirmed by 16S amplicon sequencing of the same samples [161]. As the lowest sample bacterial load was >2 logs higher than the highest extraction blank bacterial load, we pooled all taxa below 1% in an “other” category. Those taxa pooled into the other category are listed in Table 3.

Phylogenetic composition from 16S amplicon sequencing

As a control for our metagenomic sequencing, we performed 16S amplicon sequencing (the most common microbiota sequencing method) for 155 out of the 157 study sputum samples as well as 12 extraction blank controls, one for each day of DNA extractions. The V4 region of the 16S rRNA gene was amplified using primers from the Earth Microbiome Project [131] and barcodes adapted from Kozich et al [144]. (detailed at https://github.com/SchlossLab/MiSeq_WetLab_SOP/blob/master/MiSeq_WetLab_SOP_v4.md). 16S amplicons were made under the following conditions: 94°C for 3min, 30 cycles of the following sequence: [94°C for 45s, 50°C for 60s, 72°C for 90s], and then 72°C for 10min. Libraries were constructed by pooling equimolar amounts of each sample or of each blank at the volume of the least concentrated sample. Libraries were sequenced on the Illumina MiSeq platform producing paired 300 bp reads.

Table 2. Number of metagenomic sequences remaining after each filtering step by sample

Sample	Raw Sequences	Post Deduplication	Post QC Trim	Post Human Removal	Human Reads	Prop. Human	Microbial Reads
N101S0	3.97E+07	3.89E+07	3.53E+07	3.09E+07	4.38E+06	0.12	6.42E+06
N101S1	5.50E+07	5.36E+07	4.93E+07	3.58E+07	1.34E+07	0.27	1.55E+07
N101S2	5.39E+07	5.26E+07	4.85E+07	3.35E+07	1.50E+07	0.31	1.68E+07
N101S3	3.00E+07	2.92E+07	2.65E+07	7.44E+06	1.90E+07	0.72	1.96E+07
N101S4	4.20E+07	4.12E+07	3.76E+07	2.50E+07	1.26E+07	0.34	1.42E+07
N102S0	4.58E+07	4.43E+07	4.08E+07	2.82E+07	1.25E+07	0.31	1.41E+07
N102S1	3.79E+07	3.67E+07	3.24E+07	2.19E+07	1.05E+07	0.32	1.25E+07
N102S2	3.68E+07	3.58E+07	3.20E+07	2.24E+07	9.55E+06	0.30	1.14E+07
N102S3	4.97E+07	4.85E+07	4.35E+07	2.46E+07	1.89E+07	0.44	2.09E+07
N102S4	3.75E+07	3.69E+07	3.16E+07	2.39E+07	7.74E+06	0.24	1.03E+07
N201S0	3.42E+07	3.40E+07	3.00E+07	1.85E+07	1.16E+07	0.39	1.32E+07
N201S1	3.24E+07	3.19E+07	2.86E+07	3.38E+06	2.52E+07	0.88	2.54E+07
N201S2	3.03E+07	2.98E+07	2.71E+07	4.72E+06	2.24E+07	0.83	2.27E+07
N201S3	3.98E+07	3.92E+07	3.58E+07	1.50E+07	2.09E+07	0.58	2.18E+07
N201S4	5.03E+07	4.94E+07	4.44E+07	4.97E+06	3.95E+07	0.89	3.99E+07
N201S5	3.95E+07	3.89E+07	3.50E+07	6.30E+06	2.87E+07	0.82	2.93E+07
N202S0	3.17E+07	3.12E+07	2.80E+07	6.92E+06	2.10E+07	0.75	2.16E+07
N202S1	3.28E+07	3.20E+07	2.97E+07	4.88E+06	2.48E+07	0.84	2.51E+07
N202S2	2.81E+07	2.76E+07	2.53E+07	3.20E+06	2.21E+07	0.87	2.23E+07
N202S3	3.90E+07	3.84E+07	3.62E+07	1.60E+07	2.03E+07	0.56	2.09E+07
N202S4	3.58E+07	3.52E+07	3.17E+07	1.66E+07	1.52E+07	0.48	1.63E+07
N202S5A	3.25E+07	3.19E+07	2.96E+07	1.24E+07	1.73E+07	0.58	1.79E+07
N202S5B	3.06E+07	3.02E+07	2.79E+07	1.42E+07	1.38E+07	0.49	1.45E+07
N203S0	2.68E+07	2.62E+07	2.43E+07	1.04E+07	1.39E+07	0.57	1.45E+07
N203S1	2.93E+07	2.83E+07	2.63E+07	5.00E+06	2.13E+07	0.81	2.16E+07
N203S2	3.33E+07	3.24E+07	3.00E+07	6.98E+06	2.30E+07	0.77	2.35E+07
N203S3	4.55E+07	4.30E+07	3.98E+07	1.20E+07	2.78E+07	0.70	2.87E+07
N203S4	1.85E+07	1.80E+07	1.65E+07	4.17E+06	1.23E+07	0.75	1.27E+07
N203S5	2.48E+07	2.42E+07	2.20E+07	1.40E+07	8.02E+06	0.36	8.99E+06
N204S0	2.13E+07	2.11E+07	1.99E+07	1.58E+07	4.10E+06	0.21	4.69E+06
N204S1	2.91E+07	2.87E+07	2.70E+07	1.57E+07	1.13E+07	0.42	1.20E+07
N204S2	2.42E+07	2.38E+07	2.26E+07	1.01E+07	1.24E+07	0.55	1.28E+07
N204S3	2.53E+07	2.49E+07	2.33E+07	1.31E+07	1.01E+07	0.44	1.08E+07

N204S4	2.71E+07	2.66E+07	2.50E+07	1.35E+07	1.15E+07	0.46	1.20E+07
N204S5	3.24E+07	3.19E+07	2.97E+07	2.09E+07	8.83E+06	0.30	9.84E+06
N205S0	3.67E+07	3.57E+07	3.30E+07	1.19E+07	2.12E+07	0.64	2.20E+07
N205S1	2.65E+07	2.63E+07	2.39E+07	1.56E+07	8.32E+06	0.35	9.46E+06
N205S2	1.30E+07	1.29E+07	1.14E+07	8.48E+06	2.88E+06	0.25	3.64E+06
N205S3	2.66E+07	2.63E+07	2.41E+07	1.54E+07	8.62E+06	0.36	9.62E+06
N205S4	2.36E+07	2.33E+07	2.15E+07	8.77E+06	1.27E+07	0.59	1.34E+07
N205S5	3.20E+07	3.16E+07	2.89E+07	2.37E+07	5.20E+06	0.18	6.60E+06
N206S0	3.24E+07	3.11E+07	2.83E+07	8.02E+06	2.03E+07	0.72	2.11E+07
N206S1	4.78E+07	4.54E+07	4.01E+07	7.17E+06	3.29E+07	0.82	3.37E+07
N206S2	3.15E+07	3.07E+07	2.79E+07	1.85E+06	2.60E+07	0.93	2.62E+07
N206S3	3.50E+07	3.39E+07	3.06E+07	3.15E+06	2.75E+07	0.90	2.78E+07
N206S4	3.17E+07	3.05E+07	2.74E+07	4.55E+06	2.29E+07	0.83	2.33E+07
N206S5	4.13E+07	3.96E+07	3.55E+07	7.39E+06	2.81E+07	0.79	2.90E+07
N207S0	2.72E+07	2.69E+07	2.30E+07	1.42E+07	8.83E+06	0.38	1.05E+07
N207S1	3.00E+07	2.96E+07	2.63E+07	8.91E+06	1.74E+07	0.66	1.84E+07
N208S0	2.81E+07	2.78E+07	2.56E+07	2.18E+07	3.83E+06	0.15	4.91E+06
N208S1	3.46E+07	3.36E+07	3.00E+07	1.06E+07	1.94E+07	0.65	2.03E+07
N208S2	2.86E+07	2.78E+07	2.52E+07	1.50E+07	1.02E+07	0.40	1.12E+07
N208S3	2.23E+07	2.21E+07	1.93E+07	1.56E+07	3.75E+06	0.19	5.13E+06
N208S4	2.40E+07	2.37E+07	2.08E+07	1.04E+07	1.04E+07	0.50	1.15E+07
N208S5	2.55E+07	2.50E+07	2.20E+07	9.57E+06	1.24E+07	0.57	1.35E+07
N209S0	3.43E+07	3.39E+07	3.08E+07	7.10E+06	2.37E+07	0.77	2.41E+07
N209S1	2.55E+07	2.47E+07	2.24E+07	7.91E+06	1.45E+07	0.65	1.50E+07
N209S2	3.20E+07	3.14E+07	2.84E+07	6.16E+06	2.22E+07	0.78	2.27E+07
N209S3	2.85E+07	2.72E+07	2.49E+07	5.24E+06	1.96E+07	0.79	1.99E+07
N209S4A	4.07E+07	4.00E+07	3.58E+07	5.52E+06	3.03E+07	0.85	3.07E+07
N209S4B	3.48E+07	3.43E+07	3.12E+07	2.73E+06	2.85E+07	0.91	2.87E+07
N209S5	3.31E+07	3.28E+07	2.94E+07	1.44E+06	2.79E+07	0.95	2.80E+07
N210S0	3.81E+07	3.74E+07	3.35E+07	1.62E+07	1.73E+07	0.52	1.86E+07
N210S1	4.97E+07	4.85E+07	4.43E+07	2.23E+07	2.20E+07	0.50	2.33E+07
N210S2	2.84E+07	2.72E+07	2.46E+07	4.62E+06	2.00E+07	0.81	2.04E+07
N210S3	3.50E+07	3.45E+07	3.08E+07	2.15E+07	9.26E+06	0.30	1.10E+07
N210S4	2.66E+07	2.63E+07	2.30E+07	1.56E+07	7.42E+06	0.32	8.88E+06
N210S5	3.40E+07	3.35E+07	3.03E+07	2.47E+07	5.60E+06	0.18	7.20E+06
N211S0	3.13E+07	3.07E+07	2.78E+07	1.11E+07	1.67E+07	0.60	1.75E+07
N211S1	3.34E+07	3.29E+07	2.87E+07	9.33E+06	1.93E+07	0.67	2.03E+07
N211S2	2.29E+07	2.27E+07	2.00E+07	8.62E+06	1.14E+07	0.57	1.22E+07
N211S3	3.09E+07	3.04E+07	2.74E+07	9.14E+06	1.83E+07	0.67	1.90E+07

N211S4	3.13E+07	3.06E+07	2.71E+07	1.71E+07	9.98E+06	0.37	1.15E+07
N211S5	6.01E+07	5.81E+07	5.25E+07	2.95E+07	2.31E+07	0.44	2.50E+07
N212S0	3.41E+07	3.37E+07	3.16E+07	2.41E+07	7.52E+06	0.24	8.58E+06
N212S1	3.18E+07	3.11E+07	2.78E+07	2.15E+07	6.36E+06	0.23	7.98E+06
N212S2	2.19E+07	2.14E+07	1.90E+07	1.50E+07	4.01E+06	0.21	5.20E+06
N212S3	3.38E+07	3.31E+07	3.03E+07	2.22E+07	8.07E+06	0.27	9.39E+06
N212S4	3.28E+07	3.23E+07	2.97E+07	1.77E+07	1.20E+07	0.40	1.30E+07
N212S5	3.28E+07	3.22E+07	2.90E+07	2.53E+07	3.73E+06	0.13	5.40E+06
N213S0	4.21E+07	4.11E+07	3.81E+07	2.31E+07	1.50E+07	0.39	1.61E+07
N213S1	2.87E+07	2.81E+07	2.64E+07	1.80E+07	8.39E+06	0.32	9.06E+06
N213S2	3.50E+07	3.44E+07	3.26E+07	2.18E+07	1.09E+07	0.33	1.16E+07
N213S3	3.45E+07	3.38E+07	3.14E+07	2.56E+07	5.71E+06	0.18	6.99E+06
N213S4	3.58E+07	3.53E+07	3.31E+07	2.48E+07	8.25E+06	0.25	9.29E+06
N213S5	2.46E+07	2.41E+07	2.18E+07	1.19E+07	9.90E+06	0.45	1.08E+07
N214S0	2.82E+07	2.76E+07	2.39E+07	1.24E+07	1.16E+07	0.48	1.30E+07
N214S1	3.54E+07	3.48E+07	3.15E+07	2.01E+06	2.95E+07	0.94	2.97E+07
N214S2	3.26E+07	3.20E+07	2.84E+07	5.47E+06	2.30E+07	0.81	2.36E+07
N214S3	3.45E+07	3.38E+07	2.98E+07	1.12E+07	1.86E+07	0.62	1.96E+07
N214S4	3.70E+07	3.61E+07	3.23E+07	4.40E+06	2.79E+07	0.86	2.84E+07
N214S5	4.46E+07	4.34E+07	3.89E+07	2.51E+07	1.39E+07	0.36	1.58E+07
N215S0	2.91E+07	2.83E+07	2.52E+07	1.26E+07	1.26E+07	0.50	1.36E+07
N215S1	3.10E+07	3.04E+07	2.78E+07	8.83E+06	1.89E+07	0.68	1.95E+07
N215S2	2.62E+07	2.57E+07	2.29E+07	9.99E+06	1.29E+07	0.56	1.38E+07
N215S3	3.67E+07	3.57E+07	3.30E+07	1.31E+07	1.99E+07	0.60	2.06E+07
N215S4	3.41E+07	3.32E+07	3.04E+07	1.43E+07	1.62E+07	0.53	1.69E+07
N215S5	4.10E+07	4.04E+07	3.60E+07	2.52E+07	1.08E+07	0.30	1.26E+07
N216S0	3.50E+07	3.46E+07	3.06E+07	1.42E+07	1.64E+07	0.54	1.77E+07
N216S1	3.21E+07	3.16E+07	2.84E+07	6.33E+06	2.21E+07	0.78	2.26E+07
N216S2	3.05E+07	2.99E+07	2.60E+07	1.18E+07	1.42E+07	0.55	1.55E+07
N216S3	3.45E+07	3.36E+07	2.92E+07	1.49E+07	1.43E+07	0.49	1.59E+07
N216S4	3.25E+07	3.17E+07	2.85E+07	1.72E+07	1.13E+07	0.40	1.26E+07
N216S5	3.13E+07	3.07E+07	2.74E+07	1.11E+07	1.63E+07	0.59	1.72E+07
N217S0	3.00E+07	2.94E+07	2.73E+07	8.28E+06	1.90E+07	0.70	1.95E+07
N217S1	2.96E+07	2.90E+07	2.67E+07	1.31E+07	1.36E+07	0.51	1.44E+07
N301S0	3.18E+07	3.13E+07	2.83E+07	1.54E+07	1.29E+07	0.45	1.42E+07
N301S1	3.54E+07	3.51E+07	3.22E+07	8.04E+06	2.41E+07	0.75	2.47E+07
N301S2	2.22E+07	2.20E+07	2.01E+07	3.98E+06	1.61E+07	0.80	1.64E+07
N301S3	3.95E+07	3.90E+07	3.56E+07	7.45E+06	2.82E+07	0.79	2.87E+07
N301S4	2.23E+07	2.20E+07	1.97E+07	4.80E+06	1.49E+07	0.76	1.54E+07

N305S0	3.34E+07	3.28E+07	3.01E+07	4.59E+06	2.55E+07	0.85	2.58E+07
N305S1	3.30E+07	3.20E+07	2.85E+07	3.55E+06	2.49E+07	0.88	2.53E+07
N305S2	3.34E+07	3.27E+07	2.96E+07	2.90E+06	2.67E+07	0.90	2.69E+07
N305S3	3.26E+07	3.20E+07	2.90E+07	6.63E+06	2.23E+07	0.77	2.28E+07
N305S4	3.12E+07	3.05E+07	2.77E+07	5.41E+06	2.23E+07	0.80	2.27E+07
N306S0	3.20E+07	3.16E+07	2.76E+07	1.83E+07	9.29E+06	0.34	1.11E+07
N306S1	3.48E+07	3.42E+07	3.04E+07	3.94E+06	2.64E+07	0.87	2.68E+07
N306S2	4.22E+07	4.14E+07	3.78E+07	8.74E+06	2.90E+07	0.77	2.96E+07
N306S3	3.60E+07	3.54E+07	3.25E+07	7.09E+06	2.55E+07	0.78	2.58E+07
N306S4	4.29E+07	4.21E+07	3.91E+07	1.26E+07	2.65E+07	0.68	2.71E+07
N307S0	4.17E+07	4.05E+07	3.64E+07	6.07E+06	3.03E+07	0.83	3.10E+07
N307S1	2.95E+07	2.92E+07	2.53E+07	1.89E+07	6.46E+06	0.26	8.40E+06
N307S2	3.71E+07	3.65E+07	3.25E+07	1.02E+07	2.23E+07	0.69	2.32E+07
N307S3	2.56E+07	2.51E+07	2.22E+07	1.64E+07	5.89E+06	0.26	7.24E+06
N307S4	2.77E+07	2.71E+07	2.46E+07	2.47E+06	2.22E+07	0.90	2.24E+07
N308S0	2.14E+07	2.10E+07	1.81E+07	8.13E+06	1.00E+07	0.55	1.10E+07
N308S1	2.88E+07	2.75E+07	2.44E+07	9.42E+06	1.50E+07	0.61	1.61E+07
N308S2	4.06E+07	3.95E+07	3.41E+07	1.13E+07	2.28E+07	0.67	2.45E+07
N308S3	2.88E+07	2.84E+07	2.47E+07	1.59E+07	8.81E+06	0.36	1.04E+07
N308S4	4.66E+07	4.31E+07	3.85E+07	1.36E+07	2.49E+07	0.65	2.61E+07
N309S0	4.78E+07	4.66E+07	4.20E+07	1.63E+07	2.57E+07	0.61	2.69E+07
N309S1	5.70E+07	5.46E+07	5.10E+07	1.11E+07	4.00E+07	0.78	4.04E+07
N309S2	3.08E+07	3.00E+07	2.83E+07	6.98E+06	2.13E+07	0.75	2.16E+07
N309S3	2.83E+07	2.74E+07	2.60E+07	5.67E+06	2.03E+07	0.78	2.05E+07
N310S0	4.43E+07	4.36E+07	3.82E+07	2.79E+07	1.03E+07	0.27	1.28E+07
N310S1	2.70E+07	2.64E+07	2.35E+07	1.15E+07	1.19E+07	0.51	1.29E+07
N310S2	2.42E+07	2.39E+07	2.07E+07	1.41E+07	6.55E+06	0.32	8.01E+06
N312S0	2.25E+07	2.21E+07	2.08E+07	1.39E+07	6.92E+06	0.33	7.43E+06
N312S1	2.39E+07	2.36E+07	2.16E+07	1.43E+07	7.30E+06	0.34	8.19E+06
N312S2	2.42E+07	2.39E+07	2.21E+07	1.63E+07	5.81E+06	0.26	6.62E+06
N312S3	2.21E+07	2.17E+07	2.03E+07	1.38E+07	6.53E+06	0.32	7.11E+06
N312S4	3.82E+07	3.74E+07	3.44E+07	2.20E+07	1.24E+07	0.36	1.37E+07
N313S0	2.55E+07	2.51E+07	2.37E+07	1.41E+07	9.67E+06	0.41	1.02E+07
N313S1	3.01E+07	2.97E+07	2.74E+07	2.33E+07	4.10E+06	0.15	5.29E+06
N313S2	1.58E+07	1.57E+07	1.41E+07	1.03E+07	3.83E+06	0.27	4.55E+06
N313S4	2.18E+07	2.16E+07	1.98E+07	1.45E+07	5.30E+06	0.27	6.13E+06
N315S0	2.93E+07	2.84E+07	2.51E+07	4.18E+06	2.09E+07	0.83	2.14E+07
N315S1	3.86E+07	3.70E+07	3.18E+07	1.11E+07	2.07E+07	0.65	2.23E+07
N315S2	3.02E+07	2.93E+07	2.51E+07	1.58E+07	9.25E+06	0.37	1.12E+07

N315S3	3.54E+07	3.40E+07	3.07E+07	7.06E+06	2.37E+07	0.77	2.44E+07
N315S4	4.13E+07	3.99E+07	3.64E+07	3.70E+06	3.27E+07	0.90	3.30E+07
N316S0	3.15E+07	3.08E+07	2.80E+07	1.37E+07	1.43E+07	0.51	1.52E+07
N316S1	3.65E+07	3.61E+07	3.32E+07	1.79E+06	3.14E+07	0.95	3.15E+07
N316S2	3.34E+07	3.24E+07	2.94E+07	4.35E+06	2.51E+07	0.85	2.54E+07
N316S3	2.70E+07	2.66E+07	2.45E+07	9.52E+06	1.50E+07	0.61	1.56E+07
N316S4	3.31E+07	3.25E+07	2.95E+07	2.08E+07	8.68E+06	0.29	1.01E+07

Raw Sequence Number: Total number of metagenomic sequencing reads.

Number after deduplication: Total number of reads remaining after removing duplicate reads.

Number after QC filter: Total number of reads remaining after quality control filtering.

Number after human DNA removal: Total number of reads remaining after removal all reads mapping to the human genome.

Human reads: Total number of reads mapping to the human genome.

Proportion human reads: Proportion of post-QC filter reads mapping to the human genome.

Microbial reads: Total number of microbial reads used for all downstream analysis.

Table 3. Those taxa less than 1% relative abundance in all samples, pooled into the "other" category

Metagenomic sequencing (left column) and 16S amplicon sequencing (right column).

MetaPhlAn2	16S Amplicon Sequencing
Achromobacter unclassified	*Acidovorax*
Acinetobacter pittii calcoaceticus nosocomialis	*Acinetobacter*
Actinobacillus unclassified	*Actinobaculum*
Actinobaculum unclassified	*Actinomycetospora*
Actinomyces georgiae	*Aerococcus*
Actinomyces oris	*Afipia*
Actinomyces sp oral taxon 171	*Aggregatibacter*
Actinomyces sp oral taxon 448	*Alcanivorax*
Actinomyces sp ph3	*Anaeroglobus*
Actinomyces timonensis	*Anoxybacillus*
Actinomyces viscosus	*Arsenicicoccus*
Afipia unclassified	*Aspergillus*
Aggregatibacter aphrophilus	*Bergeyella*
Aggregatibacter unclassified	*Bilophila*
Alloprevotella rava	*Blautia*
Alloprevotella unclassified	*Bordetella*
Anaerococcus lactolyticus	*Bosea*
Anaerococcus obesiensis	*Bradyrhizobium*
Anaerococcus sp PH9	*Brevundimonas*
Anaeroglobus geminatus	*Brochothrix*
Aspergillus terreus	*Burkholderia-Paraburkholderia*
Bacteroides fragilis	*Butyrivibrio*
Bilophila unclassified	*Caenimonas*
Bilophila wadsworthia	*Campylobacter*
Brevibacterium unclassified	*Candidatus_Saccharimonas*
Burkholderia cenocepacia	*Cardiobacterium*
Burkholderia lata	*Catonella*
Burkholderia multivorans	*Centipeda*
Burkholderia unclassified	*Chroococcidiopsis*
C2likevirus unclassified	*Chryseobacterium*
Campylobacter concisus	*Cloacibacterium*
Campylobacter gracilis	*Comamonas*
Campylobacter rectus	*Conchiformibius*
Campylobacter showae	*Cryptobacterium*
Candida unclassified	*Defluviitaleaceae_UCG-011*
candidate division TM7 single cell isolate TM7b	*Delftia*

candidate division TM7 single cell isolate TM7c
Candidatus Prevotella conceptionensis
Capnocytophaga granulosa
Capnocytophaga ochracea
Capnocytophaga sp oral taxon 338
Cardiobacterium hominis
Cardiobacterium valvarum
Catonella morbi
Clavispora lusitaniae
Collinsella unclassified
Comamonas unclassified
Corynebacterium accolens
Corynebacterium matruchotii
Cryptobacterium curtum
Deinococcus unclassified
Delftia unclassified
Dialister micraerophilus
Dietzia unclassified
Dolosigranulum pigrum
Eggerthia catenaformis
Eikenella corrodens
Enterococcus faecalis
Eremothecium unclassified
Escherichia coli
Escherichia unclassified
Eubacteriaceae bacterium CM5
Eubacteriaceae noname unclassified
Eubacterium infirmum
Eubacterium saphenum
Eubacterium yurii
Facklamia unclassified
Finegoldia magna
Fretibacterium fastidiosum
Fusobacterium periodonticum
Gardnerella vaginalis
Gemella unclassified
Haemophilus aegyptius
Haemophilus haemolyticus
Haemophilus parahaemolyticus
Haemophilus paraphrohaemolyticus
Dietzia
Distigma
Dolosigranulum
Eikenella
Enhydrobacter
Enterococcus
Erysipelotrichaceae_UCG-006
Escherichia/Shigella
Ezakiella
Fastidiosipila
Filifactor
Flavobacterium
Fretibacterium
Gardnerella
Halomonas
Helcococcus
Herbaspirillum
hgcI_clade
Howardella
Hydrogenophaga
Hymenobacter
Janthinobacterium
Johnsonella
Jonquetella
Kingella
Klebsiella
Lactococcus
Lautropia
Lawsonella
Leuconostoc
Mannheimia
Megasphaera
Mesorhizobium
Methylobacillus
Methylobacterium
Microvirga
Mitochondrion
Mobiluncus
Mogibacterium
Moraxella

Haemophilus sputorum
Herbaspirillum unclassified
Human herpesvirus 4
Human herpesvirus 7
Johnsonella ignava
Jonquetella anthropi
Jonquetella unclassified
KI polyomavirus
Kingella denitrificans
Kingella oralis
Kingella unclassified
Lachnospiraceae oral taxon 107
Lactobacillus crispatus
Lactobacillus phage J 1
Lactobacillus phage Lc Nu
Lactobacillus vaginalis
Lautropia mirabilis
Leptotrichia buccalis
Leptotrichia goodfellowii
Leptotrichia hofstadii
Leptotrichia shahii
Leptotrichia unclassified
Leptotrichia wadei
Lymphocryptovirus unclassified
Meyerozyma guilliermondii
Mitsuokella unclassified
Mycobacterium avium
Naumovozyma unclassified
Neisseria elongata
Neisseria macacae
Neisseria meningitidis
Neisseria subflava
Olsenella unclassified
Oribacterium sp ACB1
Oribacterium sp oral taxon 078
Parvimonas sp oral taxon 110
Peptoniphilus harei
Peptoniphilus sp oral taxon 836
Porcine type C oncovirus
Porphyromonas asaccharolytica
Mycobacterium
Olsenella
Ornithobacterium
Paracoccus
Parascardovia
Pelomonas
Peptococcus
Photobacterium
Phyllobacterium
Prevotellaceae_YAB2003_group
Pseudopropionibacterium
Pseudoramibacter
Psychrobacter
Ralstonia
Ramlibacter
Rheinheimera
Rhizobium
Rhodococcus
Rikenellaceae_RC9_gut_group
Roseateles
Ruminococcaceae_UCG-014
Selenomonas
Senegalimassilia
Shinella
Slackia
Sneathia
Sphingomonas
Sphingopyxis
Streptobacillus
Suttonella
Tannerella
Thermus
Thioalkalispira
Ureaplasma
Weissella
Xanthomonas
Yersinia

Porphyromonas catoniae
Porphyromonas uenonis
Prevotella baroniae
Prevotella bivia
Prevotella buccae
Prevotella dentalis
Prevotella intermedia
Prevotella loescheii
Prevotella marshii
Prevotella micans
Prevotella saccharolytica
Prevotella sp oral taxon 473
Propionibacterium acnes
Propionibacterium phage P100D
Propionibacterium phage P101A
Propionibacterium phage PAD20
Pseudomonas geniculata
Pseudomonas phage F116
Pseudomonas unclassified
Pseudoramibacter alactolyticus
Pusillimonas unclassified
Ralstonia pickettii
Riemerella unclassified
Roseolovirus unclassified
Ruminococcus torques
Saccharomyces cerevisiae
Scardovia unclassified
Selenomonas artemidis
Selenomonas flueggei
Selenomonas noxia
Selenomonas sputigena
Serratia marcescens
Slackia exigua
Slackia unclassified
Staphylococcus hominis
Staphylococcus phage PVL
Staphylococcus warneri
Streptococcus agalactiae
Streptococcus cristatus
Streptococcus massiliensis

Streptococcus oligofermentans
Streptococcus phage Cp 1
Streptococcus pseudopneumoniae
Streptococcus sp F0442
Subdoligranulum unclassified
Tannerella forsythia
Thiomonas unclassified
Torque teno virus
Treponema denticola
Treponema maltophilum
Treponema medium
Treponema socranskii
Treponema vincentii
Ureaplasma unclassified
Ureaplasma urealyticum
Veillonella dispar
Veillonella sp 3 1 44
Veillonella sp 6 1 27
Veillonella sp oral taxon 780

16S amplicon sequencing data were analyzed using the denoising program DADA2 [133] (version 1.6.0). Briefly, we computationally trimmed 10bp off the beginning of the both the forward and reverse reads and truncated the forward read to 200bp and the reverse read to 100bp. We used our entire data set to define an error rate at each base pair and then denoised all sequences. Forward and reverse reads were merged and any pair without perfect overlap was removed. Finally, chimeric sequences were removed. This program produces a list of "Amplicon Sequence Variants" (ASV) analogous to OTUs generated with a 97% clustering method. Each ASV was annotated with the RDP Bayesian classifier [133] against the SILVA database [145] to produce a 16S amplicon taxa table. ASVs identified as Pseudomonas, Staphylococcus and Achromobacter were analyzed with BLASTn to determine species identity.

After quality filtering, sputum samples produced an average of 83,332 reads per sample, extraction blanks produced an average of 575 reads and amplicon blanks produced an average of 126 reads. Most reads in the blanks were taxa in common among samples with similar sequencing barcodes. As the absolute abundance of these reads was 2-3 logs lower in blanks than in neighboring samples during sequencing, we found it unlikely they contributed to the taxonomic profiles of our samples, but that these reads were more likely due to errors in barcode reading [146,147]. The remaining taxa were those noted by Salter et al. to be common reagent contaminants [148] and were found at <1% relative abundance in samples. As the number of reads from blanks were substantially outnumbered by those detected in sputum samples, we did not analyze the reads from these controls further.

As the lowest sample bacterial load was >2 logs higher than the highest extraction blank bacterial load, we pooled all taxa below 1% in an "other" category. Those taxa pooled into the other category are listed in File 3. The phylogenetic profiles of extraction blanks and amplicon PCR no template controls are shown in Fig. 12.

Microbiota composition analysis

Analysis of ecological sputum microbiota composition was performed in R [149] (version 3.4.2) and visualized using ggplot2 [134] (version 2.2.1). For sputum samples, all taxa with <1% relative abundance

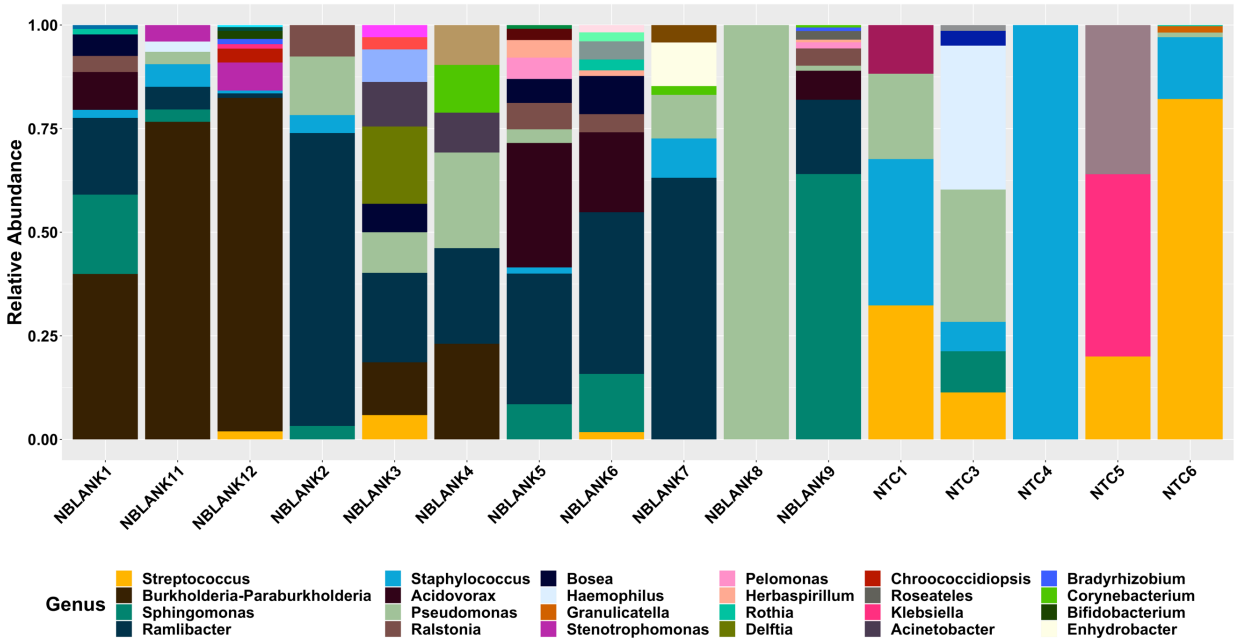

Figure 12. Phylogenetic profile of extraction blanks and 16S amplicon PCR no-template-controls

Extraction blanks (NBLANK) and no template controls (NTC) produced an average of 526 reads and 126 reads on average. NTC2 did not produce any reads and thus is not displayed.

in all samples or >1% relative abundance solely in extraction blanks were pooled into the “Other” category, as the largest difference in bacterial loads between samples and extraction blanks was 2 logs.

As MetaPhlAn2 output data are expressed as proportions, we multiplied the proportion of each taxon in a sample by that sample’s total number of microbial reads, and rounded to the nearest integer value, to obtain calculated absolute number of reads per sample contributed by each taxon. Pseudocounts were added to all reads using the cmultRepl function in the zCompositions package [192] to remove zero values, and a centered log transformation using the clr function was then performed in the compositions package [193]. PERMANOVA calculations and PCoA plots using Euclidean distance of transformed data were performed using the adonis function in the vegan package [135]. Homogeneity of variance was calculated and PCA coordinates determined using the betadisper function in the vegan package. Differences in taxonomic abundances between sample groups were identified by Wilcoxon signed rank test performed using the aldex function in the ALDEx2 package, a probabilistic model that implements statistical testing using a Dirichlet distribution designed for compositional data [194,195]. Estimated absolute abundances for all taxa were calculated by multiplying relative abundance data, after addition of pseudocounts as detailed above, by the total bacterial load determined by universal 16S qPCR. Strain-level analysis of P. aeruginosa and S. aureus was performed using StrainPhlAn2 [196] applied to metagenomic data, using default settings with the exception of setting the “--bootstrap_raxml” flag to 999.

Statistical analysis: Comparison of changes in sputum microbiota and clinical outcome measures

For a priori comparison of all microbial community characteristics to therapeutic response, we calculated change in $ppFEV_1$ and CRISS symptom score from baseline to week 4 for all subjects who had week 4 visits (26 subject total). To maximize analytical power, we treated subjective and objective clinical responses as continuous, rather than dichotomous, variables. Log-transformed viable counts of gram-negative organisms and *S. aureus*, and total bacterial load via 16S qPCR, both at baseline and change in these features from baseline to week 4, were compared to clinical outcomes using linear regression. Centered log transformation of relative abundance of all taxa, as detailed above, at baseline and change

in each taxon from baseline to week 4 were compared to clinical outcomes using linear regression. Shannon diversity at baseline and change in this metric from baseline to week 4 were compared to clinical outcomes using linear regression. Each KEGG module as baseline was compared to clinical outcomes using EdgeR [197]. A Benjamini-Hochberg correction for multiple comparisons was performed for all analyses described above, and p-values smaller than 0.05 were considered significant. Relative abundances of all taxa, normalized abundances of functional groups and diversity metrics were compared to change in cultured viable count of *P. aeruginosa* and *S. aureus* from baseline to week 1 in the same manner.

Chapter 4

Effects of Maintenance Tobramycin in the CF Sputum Metagenome

A portion of the work presented in this chapter is under review as:

Nelson MT, Walter DJ, Eng A, Weiss EJ, Vo AT, Brittnacher MJ, Hayden HS, Ravishankar S, Bautista G, Ratjen A, Blackledge M, McNamara S, Nay L, Majors C, Miller SI, Borenstein E, Simon R, LiPuma JJ, Hoffman LR. Maintenance Tobramycin Primarily Affects Untargeted Bacteria in the CF Sputum Microbiome. *Thorax. in revisions, Dec. 2019.*

Introduction

As discussed in chapter 1, it is unclear how maintenance inhaled tobramycin, an antibiotic therapy with demonstrated long-term morbidity and mortality benefits [79], improves respiratory disease outcomes or even if such effects are due to its anti-microbial action. In chapter 3, we introduced the Tobramycin Inhaled Powder (TIP) study, wherein samples were collected from thirty individuals with CF before, during and one month after a standard 28-day maintenance course of inhaled tobramycin. We hypothesized that changes in CF sputum microbiomes (both microbiota and metagenomes) with inhaled tobramycin therapy would involve taxa beyond the intended target, *P. aeruginosa*, and that defining these microbiome dynamics would identify candidate mechanisms by which chronic infections in CF persist despite antibiotic treatment. While we identified no microbial metrics (taxonomic or functional) that correlated with differential clinical response to tobramycin therapy (defined by spirometric or symptomatic change), we observed a number of significant taxonomic changes after just one week of therapy. Using calculated absolute abundances, we determined that the abundance changes in non-targeted taxa, including anaerobes, were of a greater magnitude than those observed in canonical CF pathogens *Pseudomonas aeruginosa* and *Staphylococcus aureus* with inhaled tobramycin treatment in our cohort.

Because we performed shotgun metagenomic sequencing, we were able to define not only the taxonomic changes in bacterial abundances with this therapy, but also changes in the predicted sputum microbial functional capacity, or the metagenome. In this chapter, we describe preliminary data from investigating sputum bacterial metagenomic changes with this routine maintenance medication, further exploring the functions that may explain the effect of tobramycin on both targeted and non-targeted taxa. Furthermore, previous studies have demonstrated that the metagenome of a microbial community can be "disconnected" from its microbiota: Communities can have markedly different functional capacities despite similar taxonomic profiles, and taxonomically distinct communities from similar niches can have similar functional capacities [168,169,198], raising the possibility that specific functional groups attributable to dominant pathogens may indeed be affected by tobramycin even when their species abundances change little. Our goal was to specifically investigate the functional changes with a single course of maintenance inhaled tobramycin in CF sputum.

Results

CF sputum microbial predicted functional capacity is largely maintained with inhaled tobramycin therapy despite changes in taxonomic constituency

To determine whether the significant taxonomic shifts we observed after one week of inhaled tobramycin resulted in significant changes in sputum metagenomes, we first annotated all sputum sequencing reads with the Kyoto Encyclopedia of Genes and Genomes (KEGG) database at the module level. Principal coordinate analysis (using the Aitchison dissimilarity metric, which is optimized for sparse, compositional data such as microbiota [164,165]) found no significant difference in overall functional gene content between samples collected at baseline compared to week 1 samples (Fig. 1a), suggesting that while the taxonomic composition of CF sputum microbiota may shift with antibiotic therapy, functional redundancy dampened metagenomic changes (Chapter 3, Fig. 9a vs. Chapter 4, Fig. 1a). However, thirteen individual modules differed significantly in abundance between samples collected on and off therapy (Fig. 1b, Table 1), primarily composed of metabolic, antibiotic resistance and stress response modules.

The contributions of dominant taxa to metagenomic data can mask those of less abundant taxa, such as the species we identified as significantly different after one week of therapy in Chapter 3. To further investigate whether the sputum abundance changes we observed in non-dominant taxa with tobramycin therapy (Chapter 3, Fig. 9b-c) resulted in predicted functional changes among those unconventional microbiota, we computationally removed all reads from dominant taxa, defined here as any taxon with relative abundance twice that of the next most dominant taxon at baseline in each subject [199], and repeated the same KEGG functional analysis. Principal coordinate analysis (Fig. 2a) using the Aitchison dissimilarity metric identified significant differences in overall functional gene content among these non-dominant microbiota. Significant differences were identified in 141 modules of nondominant microbiota between sputum samples collected on and off therapy (Fig. 2b, Table 2), particularly among efflux pump and stress response genes, both of which are known to be involved in bacterial responses to antibiotics [200,201]. Notably, all modules identified as significantly different after one week of therapy

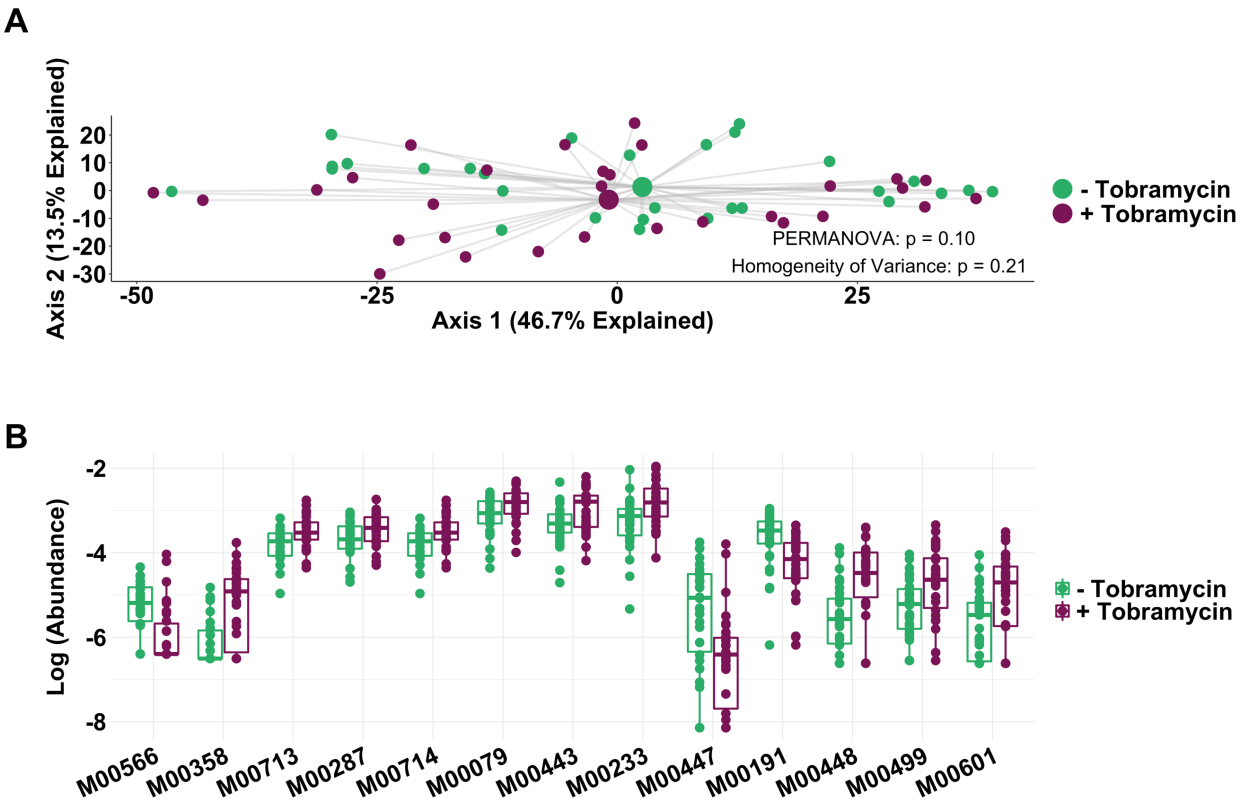

Figure 1. Genetically conferred functional capacity changes relatively little with antibiotic therapy
Genetically conferred functional capacity was determined by mapping all shotgun sequencing reads to the KEGG database at the module level. (A) Principal component analysis using the Aitchison dissimilarity metric – Euclidean distance between samples after a centered log-transformation of relative abundance data – of Baseline and Week 1 samples without removal of dominant taxa (small dots) colored and grouped by antibiotic status. (B) Log-abundances of all significantly different modules between baseline and week 1 samples without dominant taxa removed, ordered by effect size. PERMANOVA is a statistical test for difference between centroids, Homogeneity of Variance is a statistical test which assesses the difference. Modules are listed in Table 1.

Table 1. Description of functional features shown in Figure 1

All KEGG modules found to be significantly different between baseline and week 1 samples; all are shown for reads with dominant taxa present. Modules colored in red decreased after one week of therapy and modules in blue increased after one week of therapy.

Module	Description
M00079	Keratan sulfate degradation
M00191	Thiamine transport system
M00233	Glutamate transport system
M00287	PTS system, galactosamine-specific II component
M00358	Coenzyme M biosynthesis
M00443	SenX3-RegX3 (phosphate starvation response) two-component regulatory system
M00447	CpxA-CpxR (envelope stress response) two-component regulatory system
M00448	CssS-CssR (secretion stress response) two-component regulatory system
M00499	HydH-HydG (metal tolerance) two-component regulatory system
M00566	Dipeptide transport system
M00601	Putative chitobiose transport system
M00713	Fluoroquinolone resistance, efflux pump LfrA
M00714	Multidrug resistance, efflux pump QacA

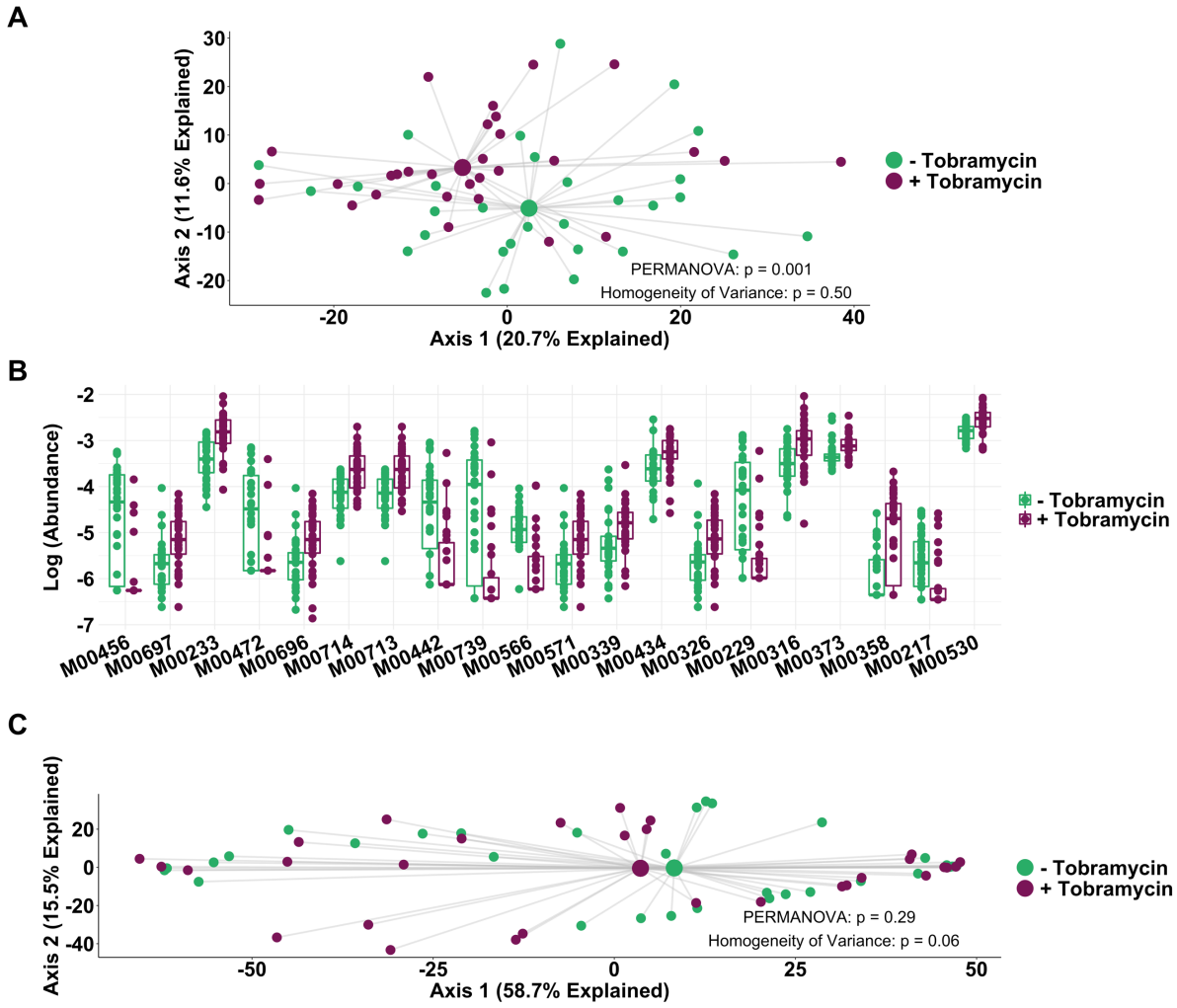

Figure 2. Genetically conferred functional capacity changes primarily in non-dominant taxa

(A) Principal component analysis of KEGG modules using the Aitchison dissimilarity metric– Euclidean distance between samples after a centered log-transformation of relative abundance data – of all Baseline and Week 1 samples colored and grouped by antibiotic treatment status after removal of reads from dominant taxa. (B) Log-abundances of the 20 modules with the largest effect size of those significantly different KEGG modules between baseline and week 1 samples after removal of reads from dominant taxa, ordered by effect size. (C) Principal component analysis using the Aitchison dissimilarity metric of all Baseline and Week 1 samples (small dots) colored and grouped by antibiotic treatment status using only reads from dominant taxa. Large dots represent the centroid of each group, with lines connecting individual sample dots to their respective centroids. Boxes represent interquartile region and middle represents the median. PERMANOVA is a statistical test for difference between centroids, Homogeneity of Variance is a statistical test which assesses the difference. Module descriptions are listed in Table 2.

Table 2: Descriptions of modules identified in Fig. 2b

All modules found to be significantly different between baseline and week 1 samples; all are shown for reads with dominant taxa removed. Modules in red decreased on average after one week of therapy and pathways in blue increased on average after one week of therapy. Asterisks indicate modules detected both with and without reads from dominant taxa present in Fig. 1b and 2b.

Module	Description
M00007	Pentose phosphate pathway, non-oxidative phase, fructose 6P => ribose 5P
M00009	Citrate cycle (TCA cycle, Krebs cycle)
M00011	Citrate cycle, second carbon oxidation, 2-oxoglutarate => oxaloacetate
M00016	Lysine biosynthesis, succinyl-DAP pathway, aspartate => lysine
M00020	Serine biosynthesis, glycerate-3P => serine
M00022	Shikimate pathway, phosphoenolpyruvate + erythrose-4P => chorismate
M00027	GABA (gamma-Aminobutyrate) shunt
M00032	Lysine degradation, lysine => saccharopine => acetoacetyl-CoA
M00033	Ectoine biosynthesis, aspartate => ectoine
M00036	Leucine degradation, leucine => acetoacetate + acetyl-CoA
M00045	Histidine degradation, histidine => N-formiminoglutamate => glutamate
M00060	Lipopolysaccharide biosynthesis, KDO2-lipid A
M00063	CMP-KDO biosynthesis
M00064	ADP-L-glycero-D-manno-heptose biosynthesis
M00076	Dermatan sulfate degradation
M00077	Chondroitin sulfate degradation
M00079*	Keratan sulfate degradation
M00082	Fatty acid biosynthesis, initiation
M00083	Fatty acid biosynthesis, elongation
M00086	beta-Oxidation, acyl-CoA synthesis
M00093	Phosphatidylethanolamine (PE) biosynthesis, PA => PS => PE
M00096	C5 isoprenoid biosynthesis, non-mevalonate pathway
M00097	beta-Carotene biosynthesis, GGAP => beta-carotene
M00115	NAD biosynthesis, aspartate => NAD
M00116	Menaquinone biosynthesis, chorismate => menaquinone
M00119	Pantothenate biosynthesis, valine/L-aspartate => pantothenate
M00121	Heme biosynthesis, glutamate => protoheme/siroheme
M00124	Pyridoxal biosynthesis, erythrose-4P => pyridoxal-5P
M00125	Riboflavin biosynthesis, GTP => riboflavin/FMN/FAD
M00126	Tetrahydrofolate biosynthesis, GTP => THF

M00127	Thiamine biosynthesis, AIR => thiamine-P/thiamine-2P
M00129	Ascorbate biosynthesis, animals, glucose-1P => ascorbate
M00133	Polyamine biosynthesis, arginine => agmatine => putrescine => spermidine
M00135	GABA biosynthesis, eukaryotes, putrescine => GABA
M00136	GABA biosynthesis, prokaryotes, putrescine => GABA
M00141	C1-unit interconversion, eukaryotes
M00149	Succinate dehydrogenase, prokaryotes
M00151	Cytochrome bc1 complex respiratory unit
M00153	Cytochrome d ubiquinol oxidase
M00155	Cytochrome c oxidase, prokaryotes
M00159	V/A-type ATPase, prokaryotes
M00169	CAM (Crassulacean acid metabolism), light
M00170	C4-dicarboxylic acid cycle, phosphoenolpyruvate carboxykinase type
M00172	C4-dicarboxylic acid cycle, NADP - malic enzyme type
M00173	Reductive citrate cycle (Arnon-Buchanan cycle)
M00185	Sulfate transport system
M00191*	Thiamine transport system
M00196	Raffinose/stachyose/melibiose transport system
M00204	Trehalose/maltose transport system
M00209	Osmoprotectant transport system
M00217	D-Allose transport system
M00218	Fructose transport system
M00229	Arginine transport system
M00233*	Glutamate transport system
M00254	ABC-2 type transport system
M00255	Lipoprotein-releasing system
M00267	PTS system, N-acetylglucosamine-specific II component
M00268	PTS system, alpha-glucoside-specific II component
M00272	PTS system, beta-glucoside (arbutin/salicin/cellobiose)-specific II component
M00287*	PTS system, galactosamine-specific II component
M00303	PTS system, N-acetylmuramic acid-specific II component
M00315	Uncharacterized ABC transport system
M00316	Manganese transport system
M00317	Manganese/iron transport system
M00320	Lipopolysaccharide export system
M00324	Dipeptide transport system

M00325	alpha-Hemolysin/cyclolysin transport system
M00326	RTX toxin transport system
M00330	Adhesin protein transport system
M00339	RaxAB-RaxC type I secretion system
M00344	Formaldehyde assimilation, xylulose monophosphate pathway
M00346	Formaldehyde assimilation, serine pathway
M00349	Microcin C transport system
M00358*	Coenzyme M biosynthesis
M00361	Nucleotide sugar biosynthesis, eukaryotes
M00362	Nucleotide sugar biosynthesis, prokaryotes
M00364	C10-C20 isoprenoid biosynthesis, bacteria
M00365	C10-C20 isoprenoid biosynthesis, archaea
M00366	C10-C20 isoprenoid biosynthesis, plants
M00368	Ethylene biosynthesis, methionine => ethylene
M00373	Ethylmalonyl pathway
M00374	Dicarboxylate-hydroxybutyrate cycle
M00377	Reductive acetyl-CoA pathway (Wood-Ljungdahl pathway)
M00391	Exosome, eukaryotes
M00434	PhoR-PhoB (phosphate starvation response) two-component regulatory system
M00442	Putative hydroxymethylpyrimidine transport system
M00443*	SenX3-RegX3 (phosphate starvation response) two-component regulatory system
M00447*	CpxA-CpxR (envelope stress response) two-component regulatory system
M00448*	CssS-CssR (secretion stress response) two-component regulatory system
M00453	QseC-QseB (quorum sensing) two-component regulatory system
M00456	ArcB-ArcA (anoxic redox control) two-component regulatory system
M00467	SasA-RpaAB (circadian timing mediating) two-component regulatory system
M00472	NarQ-NarP (nitrate respiration) two-component regulatory system
M00478	DegS-DegU (multicellular behavior control) two-component regulatory system
M00479	DesK-DesR (membrane lipid fluidity regulation) two-component regulatory system
M00493	AlgZ-AlgR (alginate production) two-component regulatory system
M00499*	HydH-HydG (metal tolerance) two-component regulatory system
M00501	PilS-PilR (type 4 fimbriae synthesis) two-component regulatory system
M00526	Lysine biosynthesis, DAP dehydrogenase pathway, aspartate => lysine
M00527	Lysine biosynthesis, DAP aminotransferase pathway, aspartate => lysine
M00529	Denitrification, nitrate => nitrogen
M00530	Dissimilatory nitrate reduction, nitrate => ammonia

M00531	Assimilatory nitrate reduction, nitrate => ammonia
M00532	Photorespiration
M00546	Purine degradation, xanthine => urea
M00552	D-galactonate degradation, De Ley-Doudoroff pathway, D-galactonate => glycerate-3P
M00554	Nucleotide sugar biosynthesis, galactose => UDP-galactose
M00555	Betaine biosynthesis, choline => betaine
M00566*	Dipeptide transport system, Firmicutes
M00571	AlgE-type Mannuronan C-5-Epimerase transport system
M00572	Pimeloyl-ACP biosynthesis, BioC-BioH pathway, malonyl-ACP => pimeloyl-ACP
M00575	Pertussis pathogenicity signature 2, T1SS
M00581	Biotin transport system
M00601*	Putative chitobiose transport system
M00620	Incomplete reductive citrate cycle, acetyl-CoA => oxoglutarate
M00628	beta-Lactam resistance, AmpC system
M00631	D-Galacturonate degradation (bacteria), D-galacturonate => pyruvate + D-glyceraldehyde 3P
M00642	Multidrug resistance, efflux pump MexJK-OprM
M00646	Multidrug resistance, efflux pump AcrAD-TolC
M00669	gamma-Hexachlorocyclohexane transport system
M00670	Mce transport system
M00696	Multidrug resistance, efflux pump AcrEF-TolC
M00697	Multidrug resistance, efflux pump MdtEF-TolC
M00701	Multidrug resistance, efflux pump EmrAB
M00713*	Fluoroquinolone resistance, efflux pump LfrA
M00714*	Multidrug resistance, efflux pump QacA
M00715	Lincosamide resistance, efflux pump LmrB
M00720	Multidrug resistance, efflux pump VexEF-TolC
M00721	Cationic antimicrobial peptide (CAMP) resistance, arnBCADTEF operon
M00729	Fluoroquinolone resistance, gyrase-protecting protein Qnr
M00739	Cationic peptide transport system
M00741	Propanoyl-CoA metabolism, propanoyl-CoA => succinyl-CoA
M00742	Aminoglycoside resistance, protease FtsH
M00747	Bacitracin transport system
M00769	Multidrug resistance, efflux pump MexPQ-OpmE
M00793	dTDP-L-rhamnose biosynthesis
M00813	Lantibiotic transport system

M00821	Multidrug resistance, efflux pump TriABC-TolC
M00822	Multidrug resistance, efflux pump MexMN-OprM
M00840	Tetrahydrofolate biosynthesis, mediated by ribA and trpF, GTP => THF
M00841	Tetrahydrofolate biosynthesis, mediated by PTPS, GTP => THF

with dominant taxa present remained significantly different after removal of dominant taxa, indicating that non-dominant taxa largely drove the observed changes in microbial functional capacity with tobramycin treatment. In support of this interpretation, repeating the analysis in Fig. 2a using only reads from dominant taxa identified no significant differences between samples collected on and off therapy, either in total functional capacity or in individual modules (Fig. 2c).

Changes in co-abundance gene groups driven by non-dominant taxa

A number of the modules with significantly different normalized abundances between baseline and week one belonged to similar functional categories, largely efflux pumps or amino acid transport systems (Table 2). In addition, not all modules within a given category changed in the same direction; as an example, M00447 decreased and M00448 increased after one week of tobramycin therapy, although both are stress response genes (Table 1). KEGG primarily reduces dimensionality of large metagenomic sequencing datasets by grouping genes into categories with similar functions [202]. This type of clustering has been shown to mask variation in metagenomic datasets, where many functions are ubiquitous or common among bacterial taxa and different functions within the same grouping can move in opposite directions in response to stimuli, dampening detectable signals [203]. Therefore, we again characterized the functional genetic capacity of all samples using the KEGG database, but this time at the higher-granularity KEGG orthology (KO) level. We then processed all KOs detected in our samples using EMPANADA, a computational program which reduces dimensionality by clustering genes that commonly co-occur and/or contribute to similar functions into groups referred to as “pathways” [203]. Similar to KEGG modules, we did not identify a significant difference in overall pathway constituency via principal component analysis using the Aitchison dissimilarity metric between baseline and Week 1 with reads from dominant taxa present (Fig. 3a); however, we did identify 16 individual pathways that changed significantly on average after one week of tobramycin therapy (Fig. 3b). Among these were increases in pathways related to glycan and isoprenoid (terpenoid) metabolism, both of which have been linked to aminoglycoside resistance [204,205]. We also observed paradoxical decreases in a number of pathways related to xenobiotic metabolism genes after one week of tobramycin therapy, although enzymatic modification of aminoglycosides is a known mechanism of resistance to these antibiotics

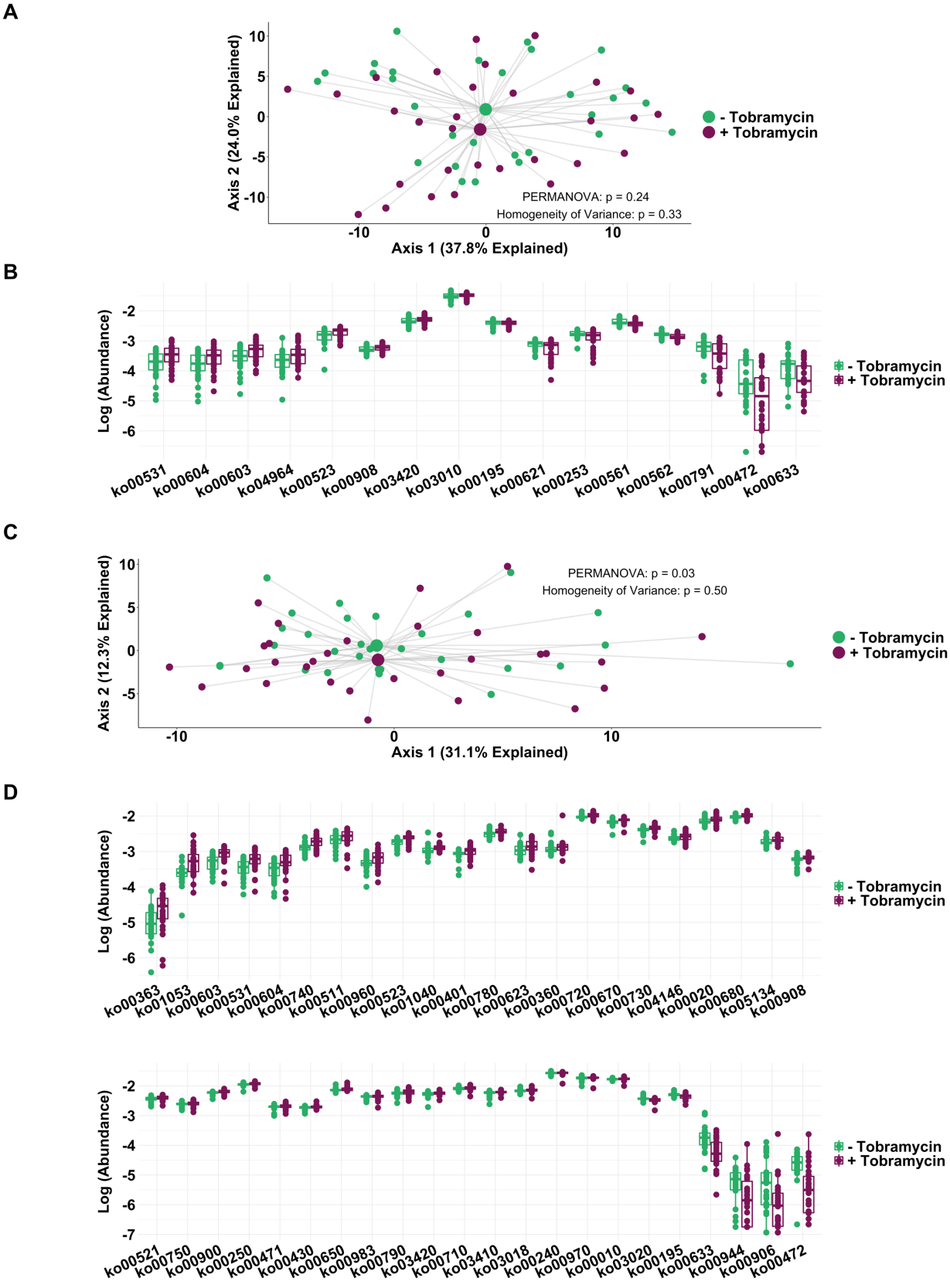
A
Axis 2 (24.0% Explained)
Axis 1 (37.8% Explained)
PERMANOVA: p = 0.24
Homogeneity of Variance: p = 0.33
- Tobramycin
+ Tobramycin
B
Log (Abundance)
ko00531
ko00604
ko00603
ko04964
ko00523
ko00908
ko03420
ko03010
ko00195
ko00621
ko00253
ko00561
ko00562
ko00791
ko00472
ko00633
C
Axis 2 (12.3% Explained)
Axis 1 (31.1% Explained)
PERMANOVA: p = 0.03
Homogeneity of Variance: p = 0.50
D
ko00363
ko01053
ko00603
ko00531
ko00604
ko00740
ko00511
ko00960
ko00523
ko01040
ko00401
ko00780
ko00623
ko00360
ko00720
ko00670
ko00730
ko04146
ko00020
ko00680
ko05134
ko00908
ko00521
ko00750
ko00900
ko00250
ko00471
ko00430
ko00650
ko00983
ko00790
ko03420
ko00710
ko03410
ko03018
ko00240
ko00970
ko00010
ko03020
ko00195
ko00633
ko00944
ko00906
ko00472

Figure 3. Changes in co-abundant gene pathways primarily driven by non-dominant taxa
Genetically conferred functional capacity was determined by mapping all shotgun sequencing reads to the KEGG database at the KO level followed by processing with EMPANADA. (A) Principal component analysis using the Aitchison dissimilarity metric – Euclidean distance between samples after a centered log-transformation of relative abundance data – of Baseline (-Tobramycin) and Week 1 (+ Tobramycin) samples without removal of dominant taxa (small dots) colored and grouped by antibiotic status. (B) Log-abundances of all significantly different pathways between baseline and week 1 samples without dominant taxa removed, ordered by effect size. (C) Principal component analysis using the Aitchison dissimilarity metric of all Baseline and Week 1 samples colored and grouped by antibiotic treatment status after removal of reads from dominant taxa. (D) Log-abundances of those significantly different pathways between baseline and week 1 samples after removal of dominant taxa, ordered by effect size. Large dots represent the centroid of each group, with lines connecting individual sample dots to their respective centroids. Boxes represent interquartile region and middle represents the median. PERMANOVA is a statistical test for difference between centroids, Homogeneity of Variance is a statistical test which assesses the difference. Pathway descriptions are listed in Table 3.

Table 3: Co-abundance gene pathways detected in Fig. 3b

All pathways found to be significantly different between baseline and week 1 samples; all are shown for reads with dominant taxa present. Pathways in red decreased on average after one week of therapy and pathways in blue increased on average after one week of therapy. The class of each pathway is indicated below the table.

Pathway	Description
ko00195	Photosynthesis proteins[1]
ko00253	Tetracycline biosynthesis[2]
ko00472	D-Arginine and D-ornithine metabolism[3]
ko00523	Polyketide sugar unit biosynthesis[2]
ko00531	Glycosaminoglycan degradation[5]
ko00561	Glycerolipid metabolism[4]
ko00562	Inositol phosphate metabolism[6]
ko00603	Glycosphingolipid biosynthesis - globo and isoglobo series[5]
ko00604	Glycosphingolipid biosynthesis - ganglio series[5]
ko00621	Dioxin degradation[7]
ko00633	Nitrotoluene degradation[7]
ko00791	Atrazine degradation[7]
ko00908	Zeatin biosynthesis[2]
ko03010	Ribosome[8]
ko03420	Nucleotide excision repair[9]
ko04964	Proximal tubule bicarbonate reclamation[10]

[1] Energy Metabolism

[2] Metabolism of terpenoids and polyketides

[3] Metabolism of other amino acids

[4] Lipid metabolism

[5] Glycan biosynthesis and metabolism

[6] Carbohydrate metabolism

[7] Xenobiotic degradation and metabolism

[8] Transcription or Translation

[9] Replication and repair

[10] Excretory system

Table 4: Co-abundance gene pathways detected in Fig. 3d

All pathways found to be significantly different between baseline and week 1 samples; all are shown for reads with dominant taxa removed. Pathways in red decreased on average after one week of therapy and pathways in blue increased on average after one week of therapy. The class of each pathway is indicated below the table.

Pathway	Description
ko00010	Glycolysis / Gluconeogenesis[1]
ko00020	TCA Cycle[1]
ko00195	Photosynthesis proteins[2]
ko00240	Pyrimidine metabolism[3]
ko00250	Alanine, aspartate and glutamate metabolism[4]
ko00360	Phenylalanine metabolism[4]
ko00363	Bisphenol degradation[5]
ko00401	Novobiocin biosynthesis[6]
ko00430	Taurine and hypotaurine metabolism[4]
ko00471	D-Glutamine and D-glutamate metabolism[4]
ko00472	D-Arginine and D-ornithine metabolism[4]
ko00511	Other glycan degradation[7]
ko00521	Streptomycin biosynthesis[6]
ko00523	Polyketide sugar unit biosynthesis[8]
ko00531	Glycan biosynthesis and metabolism[7]
ko00603	Glycosphingolipid biosynthesis[7]
ko00604	Glycosphingolipid biosynthesis[7]
ko00623	Toluene degradation[5]
ko00633	Nitrotoluene Degradation[5]
ko00650	Butanoate metabolism[1]
ko00670	One carbon pool folate[9]
ko00680	Methane metabolism[2]
ko00710	Carbon fixation in photosynthetic organisms[2]
ko00720	Carbon fixation pathways in prokaryotes[2]
ko00730	Thiamine metabolism[9]
ko00740	Riboflavin metabolism[9]
ko00750	Vitamin B6 metabolism[9]
ko00780	Biotin metabolism[9]
ko00790	Folate biosynthesis[9]
ko00900	Terpenoid backbone biosynthesis[8]

ko00906	Carotenoid biosynthesis[8]
ko00908	Zeatin biosynthesis[8]
ko00944	Flavone and flavonol biosynthesis[6]
ko00960	Tropane, piperidine and pyridine alkaloid biosynthesis[6]
ko00970	Aminoacyl tRNA biosynthesis[10]
ko00983	Drug metabolism - other enzymes[5]
ko01040	Biosynthesis of unsaturated fatty acids[11]
ko01053	Biosynthesis of siderophore group nonribosomal peptides[8]
ko03018	RNA degradation[12]
ko03020	RNA Polymerase[13]
ko03410	Base excision repair[14]
ko03420	Nucleotide excision repair[14]
ko04146	Peroxisome[15]
ko05134	Legionellosis

[1] Carbohydrate metabolism

[2] Energy metabolism

[3] Nucleotide metabolism

[4] Amino acid metabolism

[5] Xenobiotic degradation and metabolism

[6] Biosynthesis of other secondary metabolites

[7] Glycan biosynthesis and metabolism

[8] Metabolism of terpenoids and polyketides

[9] Metabolism of co-factors and vitamins

[10] Transcription or Translation

[11] Lipid metabolism

[12] Folding, sorting and degradation

[13] Translation and transcription

[14] Replication and repair

[15] Transport and catabolism

[64,206] and these functions would fit within this class. It is also possible such changes reflect changes in the inter-bacterial competitive landscape due to taxonomic shifts with tobramycin therapy. Further analysis into specific aminoglycoside-modifying genes would be required to resolve these intricacies.

As with the analysis of KEGG modules above, we computationally removed reads from dominant taxa and repeated the co-abundant pathway analysis detailed above. Principal component analysis using the Aitchison dissimilarity metric of this modified dataset indicated a significant difference between baseline and week 1 samples in overall co-abundant pathway constituency on average (Fig. 3c) and identified 44 co-abundant pathways that were significantly different on average after one week of tobramycin therapy after removal of reads from dominant taxa (Fig. 3d). Similar to analysis at the KEGG module level, nearly all of the pathways identified with reads from dominant taxa present were also identified after their removal, consistent with the primary effect of maintenance tobramycin being on non-targeted taxa. We again observed abundance increases in a number of glycan and isoprenoid metabolic pathways, but both increases and decreases in pathways related to xenobiotic degradation, in contrast to those pathways detected with reads from dominant taxa present in which we saw only decreases. These data may indicate the presence of degradative pathways present in non-dominant taxa which are selected for by tobramycin perturbation. Unique to analysis after removal of reads from dominant taxa, we identified increases in abundances of a number of co-abundant pathways relevant for amino acid and vitamin metabolism. These results may suggest that such functions are more vital to life in the CF lung for non-dominant taxa compared than for dominant pathogens, may reflect a changing metabolic landscape after tobramycin-induced taxonomic shifts or are bystander changes all together.

Changes in anaerobe-associated antibiotic resistance gene abundances after one week of tobramycin therapy

Finally, we annotated all metagenomic sequencing reads using the Comprehensive Antibiotic Resistance Database (CARD) to define and compare the antibiotic resistance complement of all samples before and after one week of tobramycin therapy. Principal coordinate analysis using the Aitchison dissimilarity metric again indicated no significant differences in total antibiotic resistance gene constituency between

baseline and Week 1 (Fig. 4a). However, we identified ten individual antibiotic resistance genes with significantly different normalized abundances between baseline and Week 1 (Fig. 4b). Half of the differentially-abundant genes were in the cfxA beta-lactamase family and were annotated as belonging to gastrointestinal and oral anaerobic taxa [207]. Four of these resistance determinants were tetracycline efflux pumps annotated as belonging to intestinal resistomes [208]. As before, we repeated this principal component analysis after removing reads from dominant taxa; as shown in Fig. 5a, while the difference between these microbiota was not as visually evident as for other analyses above, this analysis identified a significant difference in total sputum microbial antibiotic resistance gene constituency between baseline and Week 1 sample (Fig. 5a). In contrast with the prior analyses, the genes identified as significantly different in abundance between baseline and week one in this analysis included only one additional antibiotic resistance gene after removal of reads from dominant taxa (Fig. 5b).

Discussion

In Chapter 3, we investigated the effects of a single, maintenance course of inhaled tobramycin on the CF sputum microbiota. Surprisingly, non-dominant, prevalent taxa contributed most to sputum taxonomic shifts with maintenance tobramycin treatment, and these changes largely occurred after just one week of tobramycin therapy. Treatment-emergent changes in non-dominant taxa therefore represent a conserved feature among these CF sputum microbiomes, and their presence and dynamic responses to maintenance tobramycin varied much less than did those of the dominant taxa. Here, we extended these findings to define the effect of maintenance inhaled tobramycin on the CF sputum metagenome, the genetically conferred functional capacity of the microbial community, an analysis rarely used in clinical microbiology. Again, we observed a shift with tobramycin exposure in genetically conferred functional capacity, again primarily at one week of therapy, and again driven largely by non-dominant taxa, but of a much smaller magnitude than that seen in taxonomic constituency, implying functional redundancy and stability within sputum microbiota. This relationship between taxonomic and functional microbiome characteristics has been observed for diverse microbial communities [168,169], suggesting CF sputum microbiota share characteristics with other well-established microbiota.

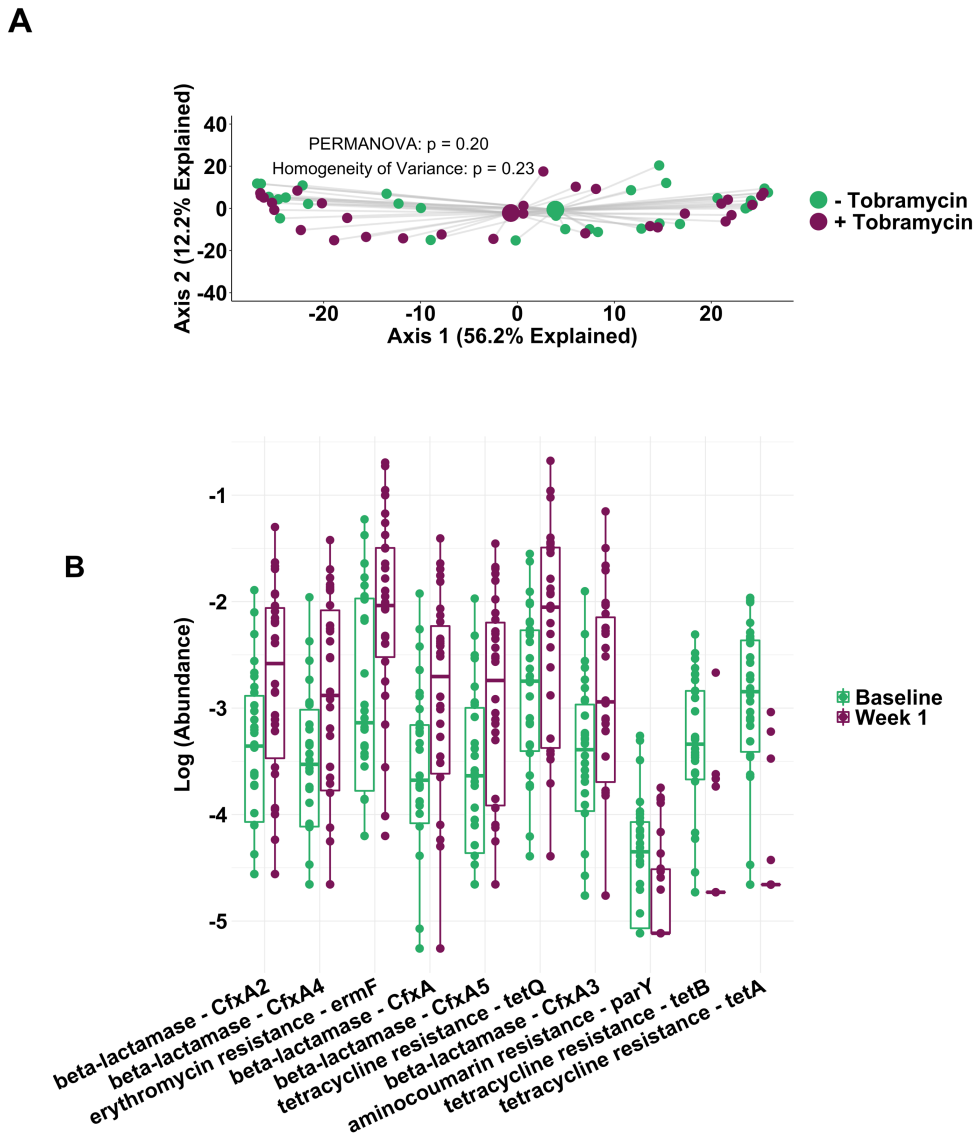

Figure 4. Change in antibiotic resistance gene complement after one week of tobramycin therapy
All shotgun sequencing reads were annotated using the comprehensive antibiotic resistance gene database (CARD). (A) Principal component analysis using the Aitchison dissimilarity metric – Euclidean distance between samples after a centered log-transformation of relative abundance data – of Baseline (- Tobramycin) and Week 1 (+ Tobramycin) samples without removal of dominant taxa (small dots) colored and grouped by antibiotic status. Large dots represent the centroid of each group, with lines connecting individual sample dots to their respective centroids. PERMANOVA is a statistical test for difference between centroids, Homogeneity of Variance is a statistical test which assesses the difference. (B) Log-abundances of all resistance genes with significantly different abundances between baseline and week 1 samples without reads from dominant taxa removed. Boxes represent interquartile region and middle represents the median.

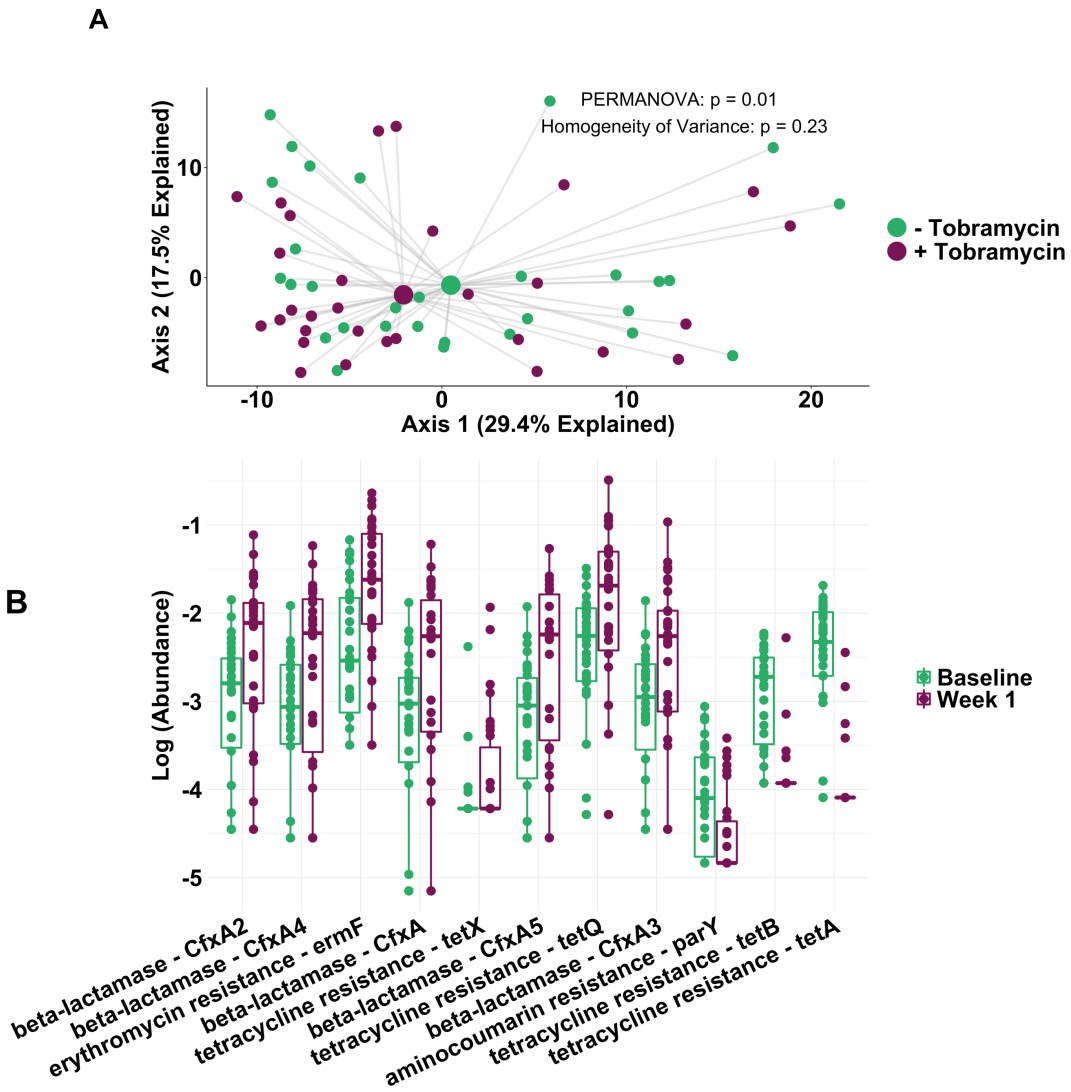

Figure 5. Change in antibiotic resistance gene complement after one week of tobramycin therapy after removal of reads from dominant taxa

Reads from dominant taxa were removed from all shotgun sequencing reads and then annotated using the comprehensive antibiotic resistance gene database (CARD). (A) Principal component analysis using the Aitchison dissimilarity metric – Euclidean distance between samples after a centered log-transformation of relative abundance data – of Baseline (-Tobramycin) and Week 1 (+ Tobramycin) samples with removal of dominant taxa (small dots) colored and grouped by antibiotic status. Large dots represent the centroid of each group, with lines connecting individual sample dots to their respective centroids. PERMANOVA is a statistical test for difference between centroids, Homogeneity of Variance is a statistical test which assesses the difference. (B) Log-abundances of all significantly different resistance genes between baseline and week 1 samples with reads from dominant taxa removed. Boxes represent interquartile region and middle represents the median.

We first analyzed all baseline and week 1 samples using the KEGG database at the module level and found that the microbial functions predicted to be altered by this treatment involved efflux pumps and responses to environmental stress, providing clues to the remarkable resilience of CF sputum microbiota to the many pressures encountered within the CF respiratory environment, which includes antibiotics and other treatments. Repeating this analysis after removal of reads from dominant taxa identified even more pronounced differences between the functional constituency of baseline and week 1 samples, further suggesting that tobramycin predominantly affects clinically non-targeted and nondominant taxa.

When we analyzed this dataset for co-abundant gene pathways, we again identified differences in abundances of a number of pathways between baseline and week 1 samples that were more pronounced after removal of reads from dominant taxa. Most of these changes involve functions that have been shown previously to play roles in antibiotic responses, including increases in abundances of pathways related to glycan and isoprenoid metabolism and decreases in those for degrading xenobiotic compounds. Glycans form part of the extracellular matrix for some bacteria cultured from CF airways, including *P. aeruginosa* [204,209] and *S. aureus* [210], and are thought to be involved in resistance to antibiotics in those taxa [211,212]. Genes involved in glycan synthesis have specifically been shown to be necessary for this resistance phenotype *in vitro* in *P. aeruginosa* [213]. While it has not been shown whether these mechanisms truly confer resistance *in vivo*, our data suggest a role for these molecules in community responses to tobramycin pressure. Decreased susceptibility to aminoglycoside antibiotics *in vitro* has also been reported in clinical isolates of *P. aeruginosa* with inherent “impermeability-type” resistance due to alterations in glycosylated structures of the cell wall and lipopolysaccharide [205]. Isoprenoids are a diverse class of organic compounds involved in a number of prokaryotic processes, including bacterial wall integrity and electron transport, both of which play role in the physical barrier and permeability to antibiotics. Therefore, it is plausible that increased abundances of pathways involved in isoprenoid metabolism may reflect cells more able to exclude antibiotics. As isoprenoid biosynthesis has been identified in a previous analysis of CF metagenomes [178] as being commonly carried by “keystone” taxa, including anaerobes, that were also common in our samples, the changes we observed likely reflect selection for these taxa during tobramycin treatment. Alternatively, specific isoprenoid molecules have been demonstrated *in vitro* to enhance tobramycin activity against *P. aeruginosa* [214] and *S.*

aureus[215]; thus increased biosynthesis of such small molecules may also be a mechanism of resistance in our samples. We also observed a decrease in co-abundant pathways involved in xenobiotic degradation for the entire microbiome; if we assume these pathways also degrade tobramycin, this change would be predicted to paradoxically *increase* community susceptibility to this drug. However, when we removed reads from dominant taxa and repeated the above analysis, we identified significant increases in abundances of a number of xenobiotic degradative pathways. Therefore, selection for microbes that encode tobramycin degradation pathways may occur among non-dominant taxa. Similarly, we identified the same isoprenoid and glycan metabolism pathways both before and after removal of reads from dominant taxa, again indicating that these functional changes are primarily driven by these non-targeted taxa and suggesting that they may confer resistance to tobramycin to their source taxa. Alternatively, the changes we observed in the abundances in these pathways may not signify that they play any role in response to tobramycin itself, but rather that they simply reflect changes in abundance of taxa that carry the “bystander” genes involved; in this case, other mechanisms not identified through our metagenomic analysis need not have conferred resistance to this treatment. This question awaits further studies.

Finally, we annotated all shotgun metagenomic reads for genes known to confer antibiotic resistance. This analysis identified ten canonical antibiotic resistance genes that differed in abundance after versus before tobramycin exposure. While these findings did not differ depending on whether dominant taxa were included or excluded, most of these ten genes were annotated as originating from anaerobic genomes, again suggesting the outsize role of non-dominant taxa in the sputum microbial response to tobramycin. Interestingly, we did not identify any aminoglycoside resistance determinants, suggesting such genes may simply reflect the changing taxonomic constituency of our samples. It is also a possibility that the genes we identified increasing in abundance after tobramycin therapy co-occur with aminoglycoside resistance genes which we were unable to detect in these taxa. As discussed in chapter 3, aminoglycosides are not canonically thought to have action against anaerobic taxa so it is possible such taxa simply do not carry resistance genes against them, although aminoglycoside-resistance genes have been detected in anaerobic environments [216,217]. It is possible such genes are not as well represented in resistance-gene database and thus more difficult to detect than the more common beta-

lactamase genes we did identify. Regardless, this functional analysis not only provides further evidence of the effects of maintenance tobramycin on non-dominant, non-targeted anaerobic taxa, but it identifies potentially unknown mechanisms of aminoglycoside response for future study.

Our work has a number of limitations, most of which are shared with the analysis in chapter 3; among these are the caveats that sputum is not a perfect sample of lower respiratory microbiology, and that using samples collected from clinically stable individuals limited our ability to correlate metagenomic characteristics with changes in clinical outcome measures. Conversely, the standardized sample collection and antibiotic exposure regimens of this study remain a strength of this analysis. Furthermore, this study represents one of largest CF sputum metagenomic analyses to date, and our data complement previous smaller studies. For example, Quinn et al. defined the sputum metagenomes from 19 individuals with CF to identify keystone functional groups, highly connected functions which can have dramatic influences on community characteristics when removed [218], and their relationship to taxonomic composition and identified isoprenoid biosynthesis, similar to the majority of isoprenoid metabolism functions we identified, as a shared, required functional pathway among many taxa found in CF respiratory samples [178] although that study did not specifically investigate changes in response to community perturbation. In this study, metagenomic sequencing allowed us to define the CF sputum microbial functional genomic capacities for a large number of samples and subjects, providing a rich view of the effects of tobramycin, and identifying significant similarities across samples with antibiotic perturbation. We also identified significant treatment-emergent abundance changes in functional groups related to antibiotic resistance, stress responses, metabolism and cell wall integrity, all carried largely by non-dominant taxa, perhaps underlying the well-known resilience of CF respiratory microbiota to therapy [29,58,177] and suggesting a potential adjuvant targets which may augment tobramycin therapy patients with CF and related infections. While our study did not identify the mechanisms of the clinical effects of maintenance inhaled tobramycin, these functional data add to the taxonomic findings in Chapter 3 to demonstrate that the largest microbiological impact of this therapy is on non-dominant taxa, suggesting a causal role for antibiotics for the decreasing microbial diversity that has been observed in CF respiratory samples with increasing age, disease severity, and antibiotic exposure [59,179]. We also identified a number of functional groups with little direct relevance to antibiotic resistance that were increased after

one week of tobramycin therapy. Such genes represent novel avenues for understanding response to tobramycin beyond canonical antibiotic resistance genes as well as potential targets of adjuvant anti-microbial therapy beyond those currently clinically available. These findings indicate the power of metagenomic sequencing during treatments such as antibiotics, anti-inflammatories, and CFTR modulators to better define the microbial determinants of CF lung disease and treatment responses.

Methods

Study design, sample collection and processing, DNA extraction and sequencing are the same as in Chapter 3 but are repeated here for ease of viewing.

Study design and Description of Subjects

The TIP study population comprised 30 people with CF who were spontaneously expectorating at three clinical sites, the University of Washington Medical Center, Seattle Children's Hospital and University of Michigan Health System. Spontaneously expectorated sputum was collected prospectively from these subjects at commencement of TIP ("baseline") and weekly during the month-long treatment period, which occurred after at least 28 days without exposure to any antibiotics other than maintenance azithromycin. A subset of participants also provided a follow-up sample one month after completion of the 28-day-long TIP treatment cycle ("follow-up"). Study inclusion criteria were as follows:

- Documented diagnosis of CF, based on standard criteria established by the Cystic Fibrosis Foundation
- Age > 6 years
- Able to reproducibly perform spirometry for clinical care
- Able to expectorate at least daily
- History of recurrent P. aeruginosa positive respiratory cultures, defined as 2 or more occurrences prior to enrollment
- Respiratory culture positive for P. aeruginosa at least once within the 12 months prior to enrollment

- Abstinence or documented use of contraception by females of child-bearing potential at baseline
- Signed informed consent, and assent, as applicable
- FEV_1 greater than 25% predicted at baseline

Study exclusion criteria were as follows:

- History of intolerance of inhaled tobramycin
- In the opinion of CF care provider, judged not to be able to tolerate 28 days without antibiotics other than chronic azithromycin
- Treatment with any antibiotics (other than chronic azithromycin) within the prior 28 days to first sample collection
- Initiation of chronic azithromycin within the prior 28 days of first sample collection
- Requiring hospitalization or treatment with antibiotics in addition to inhaled tobramycin during the study period
- Positive urine or serum pregnancy test at baseline
- Participating in an investigational drug trial that is determined by the investigators to conflict with this study's goals

Of the 30 participants recruited to the study, 4 did not complete the study: one was lost to follow up, one decided to leave the study early for unknown reasons, one did not tolerate tobramycin and was switched to alternative inhaled and oral antibiotics, one was started on intravenous antibiotics due to increased symptoms. We chose these exclusion criteria to ensure all participants were clinically stable at the time of sample collection (limiting potential microbiota changes not attributable to antibiotic therapy). Sputum samples not collected during stability were excluded from the study.

Sputum specimen processing

All samples were shipped on ice within 48 hours of collection [175] and sputum was homogenized with dithiothreitol to liquefy the specimen as described below. After culture, the remainder of the specimen was frozen for DNA analysis. Storage at 4C has been shown to preserve CF sputum microbiota profiles

compared with -80℃ [175]. The 48-hour window was chosen to allow participants to send in samples from home (maximizing recruitment in the study) as well as to allow samples to be available by mail from three different clinical sites in two cities, (University of Washington Medical Center, Seattle Children's Hospital and University of Michigan Health System), broadening generalizability. Sputum was also frozen in glycerol to allow for future culturing of non-canonical organisms where volume allowed. Subject details are present in Table 1.

DNA extraction

All samples were extracted from frozen, DTT-treated sputum using a method designed to minimize human DNA and extracellular bacterial DNA as detailed previously [161]. 300-450mL of DTT-treated sputum was suspended in 7 mL dH_2O and incubated at room temperature for one hour with gentle agitation. 10x strength Benzonase buffer (200 mM Tris-HCl, 10 mM $MgCl_2$) to a final 1X and 250U Benzonase (Sigma E-1014) were added and the sample was incubated at 37°C for 2 h with gentle agitation. 1.5M NaCl (150 mM final) was added and the resulting solution was centrifuged at 8,000g for 10min and the pellet washed once with PBS, centrifuged at 13,000g for 4min, the pellet resuspended in 400µL TE, and 0.5M EDTA (5 mM final) was added to inactivate the endonuclease. A mixture of 1 mm and 0.1 mm silica:zirconia beads and a single tungsten-carbide bead was added to the TE solution, followed by bead-beating for one minute in a BioSpec MiniBeadBeater. The resulting solution was boiled for 5 min at 95°C. Lysozyme (Sigma L6876, 3 mg/mL final) and lysostaphin (Ambi LSPN, 0.14 mg/mL final) were added, and the sample was incubated for one hour at 37°C. Proteinase K (Invitrogen 25530049, 1.4 mg/mL final) and SDS (1.8% final) were added, and the resulting solution incubated at 56°C for 30 min before cooling to room temperature. The solution was removed to a separate tube and 5 M NaCl was added (2 M final) before adding phenol:chloroform:isoamylalcohol (25:24:1) at a 1:1 volume. The solution was then incubated for 20 min at room temperature, centrifuged at 13,000 g for 20 min[17] and the top aqueous layer was collected. 0.133 volume equivalent of 7.5 M ammonium acetate was added to the aqueous layer, and the resulting solution diluted 1:1 with cold 100% ethanol to precipitate DNA. DNA product was cleaned with a spin column (BioBasic BS423, skipping the cell lysis steps and proceeding directly to DNA clean-up steps).

Quality filtering

Next-generation sequencing libraries were prepared for all samples using the Nextera DNA Sample Prep Kit (Illumina FC-121-1031) following manufacturer's instructions. All samples were sequenced on the Illumina HiSeq platform, producing an average of 3.3 $x10^7$ total reads per sample after quality filtering. Sequencing data from all samples were de-duplicated using SeqUniq (version 0.1) https://github.com/standage/sequniq) and quality filtered using KneadData (version 0.6.1) and Trimmomatic (version 0.33) [143], producing an average of 2.9 $x10^7$ reads per sample. Human reads, comprising 12%-95% of our samples based on the efficiency of the human DNA depletion step in our DNA extraction method described in Chapter 2 (Table 2), were identified and removed with BMTagger (version 3.101), resulting in $1.3x10^7$ microbial reads per sample on average.

KEGG module analysis

All metagenomic sequencing reads were mapped to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [202] and normalized using MUSiCC [219], a computational program that uses single-copy genes to normalize functional abundances in each sample, as described previously [220]. Samples were collapsed at the KEGG module level. All modules with a variance less than 0.2 and those present in less than 25% of samples were removed.

Co-abundant pathway analysis

All metagenomic sequencing reads were mapped to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [202] and normalized using MUSiCC [219]. KEGG KOs were then processed with EMPANADA to determine co-abundant functional pathways [203].

Antibiotic resistance gene analysis

The "nucleotide protein homolog model" collection from the Comprehensive Antibiotic Resistance Database [123] was used to detect antibiotic resistance genes. This model comprises 2,239 unique genes for which presence is sufficient to confer resistance (i.e. not including ubiquitous genes that only

confer resistance when mutated such as DNA gyrase). All reads were mapped to this database with Bowtie2 (Version 2.2.6) [151]. Abundances of detected genes were normalized using MUSiCC [219] as detailed above and all genes present in less than 25% of samples were removed.

Statistical analysis, principal coordinate analysis and data visualization

Analysis of ecological sputum microbiota composition was performed in R [149] (version 3.4.2) and visualized using ggplot2 [134] (version 2.2.1). PERMANOVA calculations and PCoA plots using Euclidean distance of transformed data were performed using the adonis function in the vegan package [135]. Homogeneity of variance was calculated and PCA coordinates determined using the betadisper function in the vegan package. Differences in KEGG modules between sample groups were identified by paired Wilcoxon signed rank test performed using the aldex function in the ALDEx2 package, a probabilistic model that implements statistical testing using a Dirichlet distribution designed for compositional data [194,195]. Differences in co-abundant pathways and antibiotic resistance genes were identified with a paired Wilcoxon signed rank test using the “stats” package [149]. A Benjamini-Hochberg correction for multiple comparisons was performed for all analyses described above, and p-values smaller than 0.05 were considered significant. Pseudocounts were added to all reads using the cmultRepl function in the zCompositions package [192] to remove zero values for graphing.

Chapter 5

Conclusions and Future Directions

Overview

Respiratory infections are the most significant contributors to morbidity and mortality in individuals living with cystic fibrosis (CF) today. Decades of research into the microbial constituents of CF respiratory infections have provided remarkable insight into pathogenic determinants of this disease. Despite these advances, as well as improvements in symptomatic clinical management and development of new therapies targeting the causative protein mutations, these respiratory infections persist. In the early years of scientific investigation into CF, *Staphylococcus aureus* were frequently seen in pathological post-mortem investigations in infant lungs, leading to the introduction of inhaled anti-*Staphylococcal* antibiotic therapy. Although these therapies were effective in reducing complications of such infections, a new pathogen soon emerged, *Pseudomonas aeruginosa*, which would come to be a common, and even the prototypical, pathogen in CF. *P. aeruginosa* infections have been associated with reduced quality and length of life, typically increasing in abundance in end-stage disease. With the common administration of aggressive anti-*Pseudomonal* antibiotic therapy, the pendulum has now swung back: *S. aureus* is again the most common pathogen detected in the respiratory tracts of children with CF in the US [2]. These historical dynamics exemplify the complex, nature of respiratory infections, and what is commonly referred to as the “whack-a-mole” character of their treatment.

A major caveat of the historical characterization of CF respiratory communities is that they are primarily defined using standard clinical culturing methods, and typically using sputum, a commonly-collected respiratory specimen that variably samples the entire respiratory tract [25,26]. While standard culture techniques usually yield a small number of classic pathogens such as *Staphylococcus aureus* and *Pseudomonas aeruginosa* in CF respiratory samples, culture-free sequencing methods frequently identify diverse, uncultured taxa, including a number of anaerobic taxa associated with the oral cavity [36,170,221]. These newer techniques also suggest CF sputum microbiota are relatively stable over time, with only transient changes during exacerbations and antibiotic treatments [29,30]. While one might predict that antibiotic-induced reductions in sputum abundances of “pathogenic” bacteria (such as *P. aeruginosa*) are responsible for clinical improvement, previous studies have not supported this hypothesis [29,30], usually identifying minimal and/or transient abundance changes during treatment or exacerbations

that do not consistently correlate with changes in either symptoms, lung function or *in vitro* clinical susceptibility testing [31,32]. Furthermore, while most CF microbiome research has focused on the microbial taxa ("microbiota") in sputum, the full complements of sputum microbial genes ("metagenomes") have not been well studied. Metagenomic analysis provides an opportunity to study several poorly-understood features of CF lung disease, including the mechanisms underlying the resilience of CF respiratory infections, during potent antibiotic treatments [76,80–82]. Previous studies assessing the microbial determinants of clinical status have been limited, however, by the diversity of treatment regimens studied and by their retrospective approaches: study samples are typically collected from individuals on different antibiotics and at different time points, making controlling for confounding variables challenging, particularly when considering the relatively low population prevalence of CF (limiting sample size and power). Furthermore, metagenomic sequencing is difficult in such complex infection samples due to the presence of overwhelming quantities of human and extracellular bacterial DNA. The goal of this work was to begin to address the association between the microbiota and metagenome (the taxonomic and functional genetic) constituency of CF sputum and clinical response to a single, standard maintenance course of inhaled tobramycin, a commonly used antibiotic in CF. Before we could do this, however, we first worked to optimize our sputum processing and DNA extraction methods to deplete DNA from sources that would limit sequencing coverage of viable bacteria.

Chapter 2

Development of a method for optimized characterization of microbiota and metagenomic profiles in complex clinical samples

As the price of next generation sequencing (NGS) technology continues to drop, the prospect of metagenomic analysis is becoming a realistic goal for a wide variety of clinical samples. Metagenomic sequencing offers particularly enticing advantages for complex samples, including those from chronic, inflammatory infections of the lung, nose, sinuses, skin, and other surfaces. In Chapter 2, we described an optimized method for DNA extraction from CF sputum microbiomes, a prototypical complex clinical sample, that enriches for microbial DNA from viable cells by (1) selectively lysing human cells with hypotonic treatment, and (2) digesting both human and microbial extracellular DNA using an

endonuclease. Our analysis demonstrated this approach to have superior ability to deplete human DNA from CF sputum compared with four other published or commercially available methods, resulting in a 14-fold increase in microbial reads without affecting total bacterial load measured by qPCR. Using bacterial mock communities, we were also able to demonstrate that this method does not affect culture absolute abundances of canonical CF pathogens, allowing for an optimized picture of the viable bacterial constituency of these community.

There are several important considerations from this work that may extend beyond the samples tested in this study. Focusing on the viable microbial fraction is particularly important in studies examining samples exposed to perturbations likely to broadly kill bacterial cells, which could then leave behind DNA that may mask genetic features of surviving taxa and strains. Thus, while our method was shown here to work well for metagenomic studies of CF sputum, this method is likely equally suited for other types of samples from inflamed, infected tissues that have overwhelming amounts of human DNA such as wound, vaginal, blood, oral, sinus or upper airway samples; respiratory specimens that sample more directly from the lower airways (such as bronchoalveolar lavage); or other chronic infection samples characterized by chronic antibiotic exposure. It is also easy to append the Benzonase2 methodologic steps to a variety of DNA extraction protocols, whether custom or kit-based. We also found little change in calculated taxonomic profile in samples extracted before vs. after freezing, indicating utility for previously banked samples, further widening applicability. Nevertheless, it is important to consider that while the mock communities tested here were constructed to reflect the viable microbiota found in CF sputum, unlike sputum itself, these mock communities contained neither added human DNA nor the complex matrix of proteins, glycans and cellular debris that binds extracellular DNA *in vivo*. Interestingly, when we did construct bacterial mock communities with exogenous human DNA (constituting approximately 95% of total DNA), we found that even Standard extraction efficiently removed >99% of this host DNA, a much greater efficiency than that achieved with sputum samples, highlighting the differences between purulent sputum and liquid culture. Therefore, it is possible that this method will not exhibit the same efficiency with non-sputum sample types, and this method should be tested and optimized on samples and mock communities based on other samples and infections to determine the generalizability of the results presented here. At least for sputum, we show in Chapters 3-

4 that this method was easily scalable for a large sample set; in the larger TIP study collection of 157 sputum samples, we achieved an average 11-fold enrichment in microbial DNA compared with standard extraction across this large sample set.

The ramifications of removing DNA from dead bacterial cells remain incompletely defined by this work when correlating microbial changes and clinical outcomes (including in response to therapies). DNA is an important constituent of bacterial extracellular matrices [85] and is thought to aid in shielding cells from antibiotic and host pressure [112,222,223] as well as fortifying biofilm structure [224]. Extracellular DNA is also known to be inherently immunostimulatory. Therefore, while removing extracellular DNA increases sequencing coverage of viable bacterial cells, this process eliminates the possibility of investigating potential associations between this extracellular DNA itself and clinical outcomes. Extracellular bacterial DNA may also have a clinical relevance; for example, it is also possible that focusing on DNA from cells that did *not* survive would also provide valuable insight into the pathogenesis of chronic infection; this possibility could be investigated by "subtracting" reads from analyses performed with and without Benzonase2. Further studies are needed to fully understand the utility of this method in a variety of clinical contexts and for diverse clinical questions.

Chapter 3

Investigation of the microbiota changes with a standard month-long course of maintenance antibiotic therapy

In Chapter 3, we introduced the Tobramycin Inhaled Powder (TIP) study, which yielded 157 sputum samples collected from 30 individuals with CF before, during and after a standard one-month course of maintenance, inhaled tobramycin, the antibiotic used most commonly in the US to treat individuals with CF [2]. All individuals were free of antibiotic for at least one month prior to starting therapy, with the exception of azithromycin (which is given as an anti-inflammatory in CF) and received no other antibiotics for the duration of the treatment period. All individuals were also clinically stable, limiting confounding microbiota or metagenomic variables from changes in symptom severity. Using metagenomic sequencing, qPCR, and standard culture, we defined the changes in CF sputum

microbiota (the taxonomic constituency) during to a cycle of maintenance inhaled tobramycin treatment. As expected in this study of maintenance treatment, the study population experienced modest changes in both objective (spirometry) and subjective (symptom score) clinical measures during treatment; accordingly, we identified no significant correlations between clinical and microbiological responses in this study. There are a number of reasons for these results. Our study population was relatively small, particularly in light of the complexity of metagenomic data sets (hundreds of individual taxa and thousands of potential functional groups) and the baseline heterogeneity in microbial constituency known to exist in CF respiratory samples [29], making finding unifying features of differential response difficult. The individuals in our study were also clinically stable, resulting in relatively little change in clinical parameters over the month-long study period (fewer than half of participants had a change in respiratory function greater than the technical variation in measurement). Furthermore, most subjects were not naïve to inhaled tobramycin, increasing the possibility that their respiratory microbial communities were habituated to chronic tobramycin perturbation and thus would change little with a single course. Although these inclusion criteria were chosen to limit confounding variables in our study, they may also have limited our power to detect associations between microbial and clinical changes. Conducting a study with similar standardized collection and treatment protocols before, during and after a pulmonary exacerbation may provide more insight into such associations between microbiological and clinical factors, if they in fact do exist.

Despite these limitations, we were able to identify significant treatment-emergent changes in the CF sputum microbiota. The majority of sputum microbiological changes we identified (in absolute, viable loads of *S. aureus* and *P. aeruginosa,* total absolute bacterial load, diversity metrics and overall taxonomic constituency) occurred after just one week of tobramycin therapy. Inhaled tobramycin is generally administered in alternating, one month-long cycles, a treatment schedule initially designed and tested to limit the emergence of antibiotic resistance [76] and that remains the standard clinical practice in the US, in some cases alternating with month-long cycles of inhaled aztreonam [1], which we did not study here; similarly, we did not investigate the effect of either new tobramycin exposure in a naïve patient, or of continuous administration. In this relatively homogeneous population, however, our data suggest that CF microbial communities change largely by one week of this therapy, and that the

microbial community trends back towards its pretreatment state over the following three weeks of therapy. These results suggest that continuous tobramycin (rather than cycled monthly) would not likely alter microbial communities differently than demonstrated here (and the effect could continue to diminish over time); conversely, it is possible that alternating antibiotic therapy weekly would lead to more pronounced reduction in bacterial load or alteration in taxonomic constituency of these samples, with potential clinical benefits.

Tobramycin has demonstrated beneficial effects on long-term mortality outcomes in CF [79]. While this therapy is primarily intended to target *P. aeruginosa,* the most significant abundance changes with treatment occurred among non-dominant taxa, none of which are considered classic CF pathogens. These taxa have an unclear and controversial role in CF respiratory infections: they are associated with healthy airways [26], and their abundances have been noted to rise in CF respiratory samples prior to pulmonary exacerbations [170,171]. Conversely, some of these taxa have also been implicated in worsening CF lung disease [59,172]. Aminoglycosides, including tobramycin, are most efficiently taken up through an electron transport-dependent active process exhibited by aerobically respirating cells [63]; therefore, these drugs are considered to primarily affect anaerobic taxa. However, prior work has demonstrated aminoglycosides to be active against some anaerobic taxa, albeit to a lesser extent than for aerobic species [65–67], and anoxic conditions are known to occur in CF secretions [170], suggesting that these drugs could be working at low oxygen tensions. Our study cannot determine whether tobramycin had direct antimicrobial effects on specific taxa, as opposed to indirect effects through other interacting microbes. Isolation of specific anaerobic species from clinical samples and *in vitro* antibiotic susceptibility testing will be needed to determine the potential direct anti-anaerobe effect of this drug.

Non-antimicrobial effects of tobramycin could also contribute to its long-term clinical benefits irrespective of any changes in community structure. Aminoglycosides can promote read-through of CFTR stop codons [83] and increase expression of CFTR protein *in vitro*, raising the possibility that the clinical benefit of these drugs may not derive entirely from their antimicrobial effects but from increasing expression of functional protein. Data supporting increased CFTR expression *in vivo* after

aminoglycoside therapy has been contradictory [227], however, and stop codons represent approximately 10% of CFTR mutations in the US [2], suggesting that the potential increased protein expression is contributing marginally to tobramycin's clinical effect. Aminoglycosides have also been suggested to have immunomodulatory effects separate from their antimicrobial actions [228]. Thus, tobramycin's clinical benefits may result from reducing overall inflammatory status within the conducting airways, with potential indirect impacts on microbial community structure. It is thus possible the microbiota changes we observed are entirely incidental and tobramycin's extra-antimicrobial effects are most relevant.

Even if tobramycin is able to directly target the lower-abundance taxa shown to change most here, it remains unclear from our data how these effects are related, if at all, to the long-term morbidity and mortality benefits of this common therapy. It is possible that periodic reduction in anaerobic taxa with monthly therapy reduces the total "pathogenic load" in the respiratory tract; if so, this would indicate the importance of microbes beyond standard pathogens in respiratory disease pathogenesis. It is also possible that reducing the abundances of these nontraditional taxa has secondary effects on canonical pathogen loads or behaviors, which in turn directly relate to clinical response; alternatively, these taxonomic changes may be entirely unrelated to clinical benefit and simply represent off-target effects of this therapy. Nevertheless, these findings suggest that the primary sputum microbiological effect of tobramycin maintenance therapy is on non-targeted bacteria, either directly or indirectly, offering avenues both to better define the microbial determinants of clinical response to therapy and to devise more effective treatment strategies for a broad range of people with CF and other chronic infections. Inhaled tobramycin itself is currently only approved for treatment of *P. aeruginosa* by the Federal Drug Administration; our data suggest the potential to expand the use of this drug as well.

Chapter 4

Investigation of the effects of maintenance inhaled tobramycin on the genetic functional capacity of CF sputum microbial communities

Previous studies investigating the genetic functional capacity of CF sputa are sparse [90,91,96,178], have employed small samples sets and primarily focused on species-level characterization of microbial communities. Here, we utilized functional metagenomic sequencing to characterize the functional shifts occurring after a single week of tobramycin therapy. In support of our findings in Chapter 3, we found that treatment-emergent changes in sputum microbial functional gene groups primarily occurred in non-dominant, predominantly anaerobic, taxa, a conclusion strongly supported by repeating all analyses before and after removal of reads from dominant taxa. However, our preliminary analyses also underscore the complexity of functional metagenomic studies: performing our analyses using two "levels" of functional annotation, those for KEGG modules and co-abundant gene pathways, yielded markedly different sets of functional genetic groups that significantly differed in abundance before and after therapy. With KEGG modules, we identified functions related to efflux pumps and response to environmental stresses, whereas with co-abundant gene pathways, we primarily identified metabolic and xenobiotic degradation groups. Any or all of these functions could reflect direct adaption to perturbation from antibiotic therapy, changes in overall community constituency and resulting alterations to inter-bacterial interactions or even adaptive functions present in a small number of taxa that are required for existence in the complex CF respiratory environment and are unrelated to taxonomic shifts resulting from tobramycin pressure but revealed by such shifts. It is also possible that these functional changes simply reflect overarching taxonomic shifts due to different, undetected, functional groups all together. This interpretation is supported by our antibiotic resistance gene analysis which identified a number of resistance genes with significant changes in abundance after one week of therapy associated with anaerobic taxa, but none related specifically to aminoglycoside resistance.

Although many of the identified functional categories have been noted in the literature to be involved in antibiotic resistance, primarily *in vitro*, further work is needed to understand how our results fit in with this prior literature. For instance, we identified increased gene abundances for a number of pathways related to biosynthesis of glycosylated moieties which are also known constituents of the biofilm matrices constructed by bacteria in complex infections and are involved in antibiotic resistance [204,210,212,213], but our data do not differentiate between increased production of biofilm-associated exopolysaccharide and glycosylation of surface-associated signal transduction molecules. Thus,

increases in these pathways may simply reflect changes in inter-bacterial interactions after shifts in taxonomic structure. More detailed genomic studies investigating individual genes within these functional groups and what taxa may be harboring them and *in vitro* testing of clinical isolates is required to fully explore the clinical and pathogenesis ramifications of these findings. We also identified significant increases after one week of inhaled tobramycin in genetic capacity for isoprenoid biosynthesis, which has been identified as a keystone function in CF sputum [178], that is present in and required by a number of taxa within the community and could therefore be a candidate drug target, as was proposed by Quinn et al. [178]. Our data suggest that antibiotics may select for taxa or strains which are more dependent on these pathways, further supporting such a drug target. It is not clear from our data, however, whether such pathways are being utilized by microbes within the community to resist antibiotic pressure, and thus represent a potential microbial community Achilles heel, or if they simply reflect changes in community structure due to other compensatory functions in response to tobramycin therapy. In broader sense, our data suggest selection for number of functional categories after one week of tobramycin therapy which represent potential drug targets and topics for further investigation. For example, we observed a significant increase in a number of multidrug efflux pump modules, potentially leading to increased resistance to a number of antimicrobials besides tobramycin, suggesting following tobramycin therapy with drugs susceptible to these pumps would not be effective. We also identified increases in a number of beta-lactamase genes, suggesting following tobramycin with beta-lactam class antibiotics would be similarly ineffective. Conversely, we identified increases in a number of metabolic genes in response to tobramycin therapy which may point to adjuvant therapies which would be effective. For example, we observed an increase in folate biosynthesis functions after tobramycin therapy suggesting combined therapy with a folate-synthesis inhibitor may be effective. We also observed decreases in tetracycline resistance genes, again suggesting a potential adjuvant. Again, further bioinformatic and *in vitro* studies with mock communities investigating whether tobramycin may select for taxa with changes in these functions and others would be required to flesh out these details.

Final remarks

In this work, we optimized a method for metagenomic sequencing and analysis of CF sputum, a highly viscous and purulent sample from a site characterized by chronic polymicrobial infections, marked bacterial and human cell turnover and resultingly high levels of extracellular DNA. We then applied this method to a large collection of CF sputa before, during and after standard antibiotic therapy, uncovering notable and surprising effects on non-targeted taxa as well as changes in a number of intriguing functional pathways. While we do not definitively identify determinants of differential clinical or microbiological response to therapy, this study identified a number of promising taxa and functional pathways for further investigation. We hope this work can serve as a jumping off point to better understand this and other equally complex and recalcitrant infections.

Citations

1 Rochholz, E.L. (2018) *Alemannisches kinderlied und kinderspiel aus der schweiz*, Creative Media Partners, LLC.

2 *Cystic Fibrosis Foundation Patient Registry: Annual Data Report to the Center Directors, 2015*,

3 Andersen, D.H. (1938) Cystic fibrosis of the pancreas and its relation to celiac disease: a clinical and pathologic study. *American journal of Diseases of Children* 56, 344–399

4 Andersen, D.H. and others (1939) Cystic flbrosis of the pancreas, vitamin A deficiency, and bronchiectasis. *Journal of Pediatrics* 15, 763–771

5 Andersen, D.H. and Hodges, R.G. (1946) Celiac syndrome: V. genetics of cystic fibrosis of the pancreas with a consideration of etiology. *American journal of diseases of children* 72, 62–80

6 KESSLER, W.R. and Andersen, D.H. (1951) Heat prostration in fibrocystic disease of the pancreas and other conditions. *Pediatrics* 8, 648–656

7 di Sant'Agnese, P.A. *et al.* (1953) Abnormal electrolyte composition of sweat in cystic fibrosis of the pancreas: clinical significance and relationship to the disease. *Pediatrics* 12, 549–563

8 di Sant'Tagnese, P. *et al.* (1946) Chemotherapy in infections of the respiratory tract associated with cystic fibrosis of the pancreas; observations with penicillin and drugs of the sulfonamide group, with special reference to penicillin aerosol. *Am. J. Dis. Child* 72, 17–61

9 Shwachman, H. *et al.* (1949) Aureomycin therapy in the pulmonary involvement of pancreatic fibrosis (mucoviscidosis). *New England Journal of Medicine* 241, 185–192

10 GARRARD, S.D. *et al.* (1951) Pseudomonas aeruginosa infection as a complication of therapy in pancreatic fibrosis (mucoviscidosis). *Pediatrics* 8, 482–488

11 Høiby, N. and Axelsen, N.H. (1973) Identification and quantitation of precipitins against Pseudomonas aeruginosa in patients with cystic fibrosis by means of crossed immunoelectrophoresis with intermediate gel. *Acta Pathologica Microbiologica Scandinavica Section B Microbiology and Immunology* 81, 298–308

12 Hoiby, N. *et al.* (1977) Pseudomonas aeruginosa infection in cystic fibrosis. Diagnostic and prognostic significance of Pseudomonas aeruginosa precipitins determined by means of crossed immunoelectrophoresis. *Scandinavian journal of respiratory diseases* 58, 65–79

13 Szaff, M. *et al.* (1983) Frequent antibiotic therapy improves survival of cystic fibrosis patients with chronic Pseudomonas aeruginosa infection. *Acta Paediatrica* 72, 651–657

14 Ramsey, B.W. *et al.* (2011) A CFTR potentiator in patients with cystic fibrosis and the G551D mutation. *New England Journal of Medicine* 365, 1663–1672

15 Middleton, P.G. *et al.* (2019) Elexacaftor–tezacaftor–ivacaftor for cystic fibrosis with a single Phe508del allele. *New England Journal of Medicine* 381, 1809–1819

16 Sawicki, G.S. *et al.* (2015) Sustained benefit from ivacaftor demonstrated by combining clinical trial and cystic fibrosis patient registry data. *American journal of respiratory and critical care medicine* 192, 836–842

17 McKone, E.F. *et al.* (2014) Long-term safety and efficacy of ivacaftor in patients with cystic fibrosis who have the Gly551Asp-CFTR mutation: a phase 3, open-label extension study (PERSIST). *The lancet Respiratory medicine* 2, 902–910

18 Heltshe, S.L. *et al.* (2015) Pseudomonas aeruginosa in cystic fibrosis patients with G551D-CFTR treated with ivacaftor. *Clinical Infectious Diseases* 60, 703–712

19 Hisert, K.B. *et al.* (2017) Restoring CFTR Function Reduces Airway Bacteria and Inflammation in People With Cystic Fibrosis and Chronic Lung Infections. *American Journal of Respiratory and Critical Care Medicine* DOI: 10.1164/rccm.201609-1954OC

20 Emerson, J. *et al.* (2002) Pseudomonas aeruginosa and other predictors of mortality and morbidity in young children with cystic fibrosis. *Pediatric Pulmonology* 34, 91–100

21 Gibson, R.L. *et al.* (2003) Pathophysiology and Management of Pulmonary Infections in Cystic Fibrosis. *American Journal of Respiratory and Critical Care Medicine* 168, 918–951

22 Schwensen, H.F. *et al.* (2020) Pseudomonas aeruginosa antibody response in cystic fibrosis decreases rapidly following lung transplantation. *Journal of Cystic Fibrosis*

23 Konstan, M.W. *et al.* (2007) Risk Factors For Rate of Decline in Forced Expiratory Volume in One Second in Children and Adolescents with Cystic Fibrosis. *The Journal of Pediatrics* 151, 134-139.e1

24 Sanders, D.B. *et al.* (2011) Pulmonary exacerbations are associated with subsequent FEV1 decline in both adults and children with cystic fibrosis. *Pediatric Pulmonology* 46, 393–400

25 Carmody, L.A. *et al.* (2018) Fluctuations in airway bacterial communities associated with clinical states and disease stages in cystic fibrosis. *PLOS ONE* 13, e0194060

26 Muhlebach, M.S. *et al.* (2018) Initial acquisition and succession of the cystic fibrosis lung microbiome is associated with disease progression in infants and preschool children. *PLOS Pathogens* 14, e1006798

27 Tunney, M.M. *et al.* (2011) Use of culture and molecular analysis to determine the effect of antibiotic treatment on microbial community diversity and abundance during exacerbation in patients with cystic fibrosis. *Thorax* DOI: 10.1136/thx.2010.137281

28 Fodor, A.A. *et al.* (2012) The Adult Cystic Fibrosis Airway Microbiota Is Stable over Time and Infection Type, and Highly Resilient to Antibiotic Treatment of Exacerbations. *PLoS ONE* DOI: 10.1371/journal.pone.0045001

29 Carmody, L.A. *et al.* (2015) The daily dynamics of cystic fibrosis airway microbiota during clinical stability and at exacerbation. *Microbiome* 3,

30 Price, K.E. *et al.* (2013) Unique microbial communities persist in individual cystic fibrosis patients throughout a clinical exacerbation. *Microbiome* 1, 27

31 O'Toole, G.A. (2017) Cystic Fibrosis Airway Microbiome: Overturning the Old, Opening the Way for the New. *Journal of Bacteriology* 200, e00561-17

32 Regelmann, W.E. *et al.* (1990) Reduction of SputumPseudomonas aeruginosaDensity by Antibiotics Improves Lung Function in Cystic Fibrosis More than Do Bronchodilators and Chest Physiotherapy Alone. *American Review of Respiratory Disease* 141, 914–921

33 Stressmann, F.A. *et al.* (2011) Does bacterial density in cystic fibrosis sputum increase prior to pulmonary exacerbation? *Journal of Cystic Fibrosis* DOI: 10.1016/j.jcf.2011.05.002

34 O'Toole, G.A. (2018) Cystic fibrosis airway microbiome: Overturning the old, opening the way for the new. *Journal of Bacteriology* DOI: 10.1128/JB.00561-17

35 Regelmann, W.E. *et al.* (1990) Reduction of Sputum *Pseudomonas aeruginosa* Density by Antibiotics Improves Lung Function in Cystic Fibrosis More than Do Bronchodilators and Chest Physiotherapy Alone. *American Review of Respiratory Disease* 141, 914–921

36 Cox, M.J. *et al.* (2010) Airway Microbiota and Pathogen Abundance in Age-Stratified Cystic Fibrosis Patients. *PLoS ONE* 5, e11044

37 Rogers, G.B. *et al.* (2004) characterization of bacterial community diversity in cystic fibrosis lung infections by use of 16s ribosomal DNA terminal restriction fragment length polymorphism profiling. *J Clin Microbiol* 42, 5176–83

38 Rudkjøbing, V.B. *et al.* (2011) True Microbiota Involved in Chronic Lung Infection of Cystic Fibrosis Patients Found by Culturing and 16S rRNA Gene Analysis ▿ . *Journal of Clinical Microbiology* 49, 4352–4355

39 Zhao, J. *et al.* (2012) Decade-long bacterial community dynamics in cystic fibrosis airways. *Proceedings of the National Academy of Sciences of the United States of America* 109, 5809–5814

40 Tunney, M.M. *et al.* (2011) Use of culture and molecular analysis to determine the effect of antibiotic treatment on microbial community diversity and abundance during exacerbation in patients with cystic fibrosis. *Thorax* DOI: 10.1136/thx.2010.137281

41 Fodor, A.A. *et al.* (2012) The Adult Cystic Fibrosis Airway Microbiota Is Stable over Time and Infection Type, and Highly Resilient to Antibiotic Treatment of Exacerbations. *PLoS ONE* DOI: 10.1371/journal.pone.0045001

42 Stressmann, F.A. *et al.* (2012) Long-term cultivation-independent microbial diversity analysis demonstrates that bacterial communities infecting the adult cystic fibrosis lung show stability and resilience. *Thorax* DOI: 10.1136/thoraxjnl-2011-200932

43 Heirali, A.A. *et al.* (2017) The effects of inhaled aztreonam on the cystic fibrosis lung microbiome. *Microbiome* DOI: 10.1186/S40168-017-0265-7

44 Carmody, L.A. *et al.* (2015) The daily dynamics of cystic fibrosis airway microbiota during clinical stability and at exacerbation. *Microbiome* DOI: 10.1186/s40168-015-0074-9

45 Cystic Fibrosis Foundation (2016) *Patient Registry - Annual Data Report to the centers directors 2015*,

46 Coburn, B. *et al.* (2015) Lung microbiota across age and disease stage in cystic fibrosis. *Scientific Reports* DOI: 10.1038/srep10241

47 Brown, P.S. *et al.* (2014) Directly sampling the lung of a young child with cystic fibrosis reveals diverse microbiota. *Annals of the American Thoracic Society* DOI: 10.1513/AnnalsATS.201311-383OC

48 Klepac-Ceraj, V. *et al.* (2010) Relationship between cystic fibrosis respiratory tract bacterial communities and age, genotype, antibiotics and Pseudomonas aeruginosa. *Environmental Microbiology* DOI: 10.1111/j.1462-2920.2010.02173.x

49 Zemanick, E.T. *et al.* (2017) Airway microbiota across age and disease spectrum in cystic fibrosis. *European Respiratory Journal* DOI: 10.1183/13993003.00832-2017

50 Van Der Gast, C.J. *et al.* (2011) Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. *ISME Journal* DOI: 10.1038/ismej.2010.175

51 Flight, W.G. *et al.* (2015) Rapid detection of emerging pathogens and loss of microbial diversity associated with severe lung disease in cystic fibrosis. *Journal of Clinical Microbiology* DOI: 10.1128/JCM.00432-15

52 Hisert, K.B. *et al.* (2017) Restoring cystic fibrosis transmembrane conductance regulator function reduces airway bacteria and inflammation in people with cystic fibrosis and chronic lung infections. *American Journal of Respiratory and Critical Care Medicine* DOI: 10.1164/rccm.201609-1954OC

53 Carmody, L.A. *et al.* (2013) Changes in cystic fibrosis airway microbiota at pulmonary exacerbation. *Annals of the American Thoracic Society* DOI: 10.1513/AnnalsATS.201211-107OC

54 Carmody, L.A. *et al.* (2018) Fluctuations in airway bacterial communities associated with clinical states and disease stages in cystic fibrosis. *PLOS ONE* 13, e0194060

55 Acosta, N. *et al.* (2018) Sputum microbiota is predictive of long-term clinical outcomes in young adults with cystic fibrosis. *Thorax* DOI: 10.1136/thoraxjnl-2018-211510

56 Heirali, A.A. *et al.* (2017) The effects of inhaled aztreonam on the cystic fibrosis lung microbiome. *Microbiome* DOI: 10.1186/S40168-017-0265-7

57 Zemanick, E.T. *et al.* (2013) Inflammation and Airway Microbiota during Cystic Fibrosis Pulmonary Exacerbations. *PLoS ONE* DOI: 10.1371/journal.pone.0062917

58 Stressmann, F.A. *et al.* (2011) Does bacterial density in cystic fibrosis sputum increase prior to pulmonary exacerbation? *Journal of Cystic Fibrosis* 10, 357–365

59 Zhao, J. *et al.* (2012) Decade-long bacterial community dynamics in cystic fibrosis airways. *Proceedings of the National Academy of Sciences* 109, 5809–5814

60 Dickson, R.P. *et al.* (2015) Spatial variation in the healthy human lung microbiome and the adapted island model of lung biogeography. *Annals of the American Thoracic Society* 12, 821–830

61 Dickson, R.P. *et al.* (2017) Bacterial topography of the healthy human lower respiratory tract. *MBio* 8, e02287-16

62 Kramer, R. *et al.* (2015) High Individuality of Respiratory Bacterial Communities in a Large Cohort of Adult Cystic Fibrosis Patients under Continuous Antibiotic Treatment. *PLOS ONE* 10, e0117436

63 Brogden, R.N. *et al.* (1976) Tobramycin: A Review of its Antibacterial and Pharmacokinetic Properties and Therapeutic Use. *Drugs* 12, 166–200

64 Krause, K.M. *et al.* (2016) Aminoglycosides: an overview. *Cold Spring Harbor perspectives in medicine* 6, a027029

65 Martin, W.J. *et al.* (1972) In vitro antimicrobial susceptibility of anaerobic bacteria isolated from clinical specimens. *Antimicrobial Agents and Chemotherapy* 1, 148–158

66 Fraimow, H.S. *et al.* (1991) Tobramycin uptake in Escherichia coli is driven by either electrical potential or ATP. *Journal of Bacteriology* 173, 2800–2808

67 Smith, H. *et al.* (2013) “Affect of anaerobiosis on the antibiotic susceptibility of H. influenzae.” *BMC Research Notes* 6, 241

68 Meylan, S. *et al.* (2017) Carbon Sources Tune Antibiotic Susceptibility in Pseudomonas aeruginosa via Tricarboxylic Acid Cycle Control. *Cell Chemical Biology* 24, 195–206

69 Baran, D. *et al.* (1975) Concentration of gentamicin in bronchial secretions of children with cystic fibrosis of tracheostomy.(Comparison between the intramuscular route, the endotracheal instillation and aerosolization). *International journal of clinical pharmacology and biopharmacy* 12, 336–341

70 Louria, D.B. *et al.* (1969) Gentamicin in the treatment of pulmonary infections. *The Journal of infectious diseases* 119, 483–485

71 Weintraub, R.G. *et al.* (1982) Comparative nephrotoxicity of two aminoglycosides: Gentamicin and tobramycin. *Medical Journal of Australia* 2, 129–132

72 Martin, A. *et al.* (1980) Gentamicin and tobramycin compared in the treatment of mucoid pseudomonas lung infections in cystic fibrosis. *Archives of disease in childhood* 55, 604–607

73 Melby, K. *et al.* (1979) A Comparison of the in vitro Activity of Tobramycin and Gentamicin against 6,042 Clinical Isolates. *Chemotherapy* 25, 286–295

74 Smith, A.L. *et al.* (1989) Safety of aerosol tobramycin administration for 3 months to patients with cystic fibrosis. *Pediatric pulmonology* 7, 265–271

75 Steinkamp, G. *et al.* (1989) Long-term tobramycin aerosol therapy in cystic fibrosis. *Pediatric Pulmonology* 6, 91–98

76 Ramsey, B.W. *et al.* (1999) Intermittent administration of inhaled tobramycin in patients with cystic fibrosis. *New England Journal of Medicine* 340, 23–30

77 Ramsey, B.W. *et al.* (1993) Efficacy of Aerosolized Tobramycin in Patients with Cystic Fibrosis. *New England Journal of Medicine* 328, 1740–1746

78 MacLusky, I.B. *et al.* (1989) Long-term effects of inhaled tobramycin in patients with cystic fibrosis colonized with Pseudomonas aeruginosa. *Pediatric pulmonology* 7, 42–8

79 Sawicki, G.S. *et al.* (2011) Reduced mortality in cystic fibrosis patients treated with tobramycin inhalation solution. *Pediatric Pulmonology* 47, 44–52

80 Konstan, M.W. *et al.* (2011) Safety, efficacy and convenience of tobramycin inhalation powder in cystic fibrosis patients: The EAGER trial. *Journal of Cystic Fibrosis* 10, 54–61

81 Konstan, M.W. *et al.* (2011) Tobramycin inhalation powder for P. aeruginosa infection in cystic fibrosis: The EVOLVE trial. *Pediatric Pulmonology* 46, 230–238

82 Harrison, M.J. *et al.* (2014) Inhaled versus nebulised tobramycin: A real world comparison in adult cystic fibrosis (CF). *Journal of Cystic Fibrosis* 13, 692–698

83 Wilschanski, M. *et al.* (2003) Gentamicin-induced correction of CFTR function in patients with cystic fibrosis and CFTR stop mutations. *New England Journal of Medicine* 349, 1433–1441

84 Howard, M. *et al.* (1996) Aminoglycoside antibiotics restore CFTR function by overcoming premature stop mutations. *Nature medicine* 2, 467–469

85 Jakubovics, N.S. *et al.* (2013) Life after death: The critical role of extracellular DNA in microbial biofilms. *Letters in Applied Microbiology* 57, 467–475

86 Nguyen, L.D.N. *et al.* (2016) Effects of Propidium Monoazide (PMA) Treatment on Mycobiome and Bacteriome Analysis of Cystic Fibrosis Airways during Exacerbation. *PLoS One* 11, e0168860

87 Carini, P. *et al.* (2016) Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nature Microbiology* 2, 16242

88 Rogers, G.B. *et al.* (2013) Reducing bias in bacterial community analysis of lower respiratory infections. *ISME J* 7, 697–706

89 Rogers, G.B. *et al.* (2010) The exclusion of dead bacterial cells is essential for accurate molecular analysis of clinical samples. *Clinical Microbiology and Infection* 16, 1656–1658

90 Lim, Y.W. *et al.* (2013) Clinical Insights from Metagenomic Analysis of Sputum Samples from Patients with Cystic Fibrosis. *Journal of Clinical Microbiology* 52, 425–437

91 Lim, Y.W. *et al.* (2013) Metagenomics and metatranscriptomics: windows on CF-associated viral and microbial communities. *J Cyst Fibros* 12, 154–64

92 Feigelman, R. *et al.* (2017) Sputum DNA sequencing in cystic fibrosis: non-invasive access to the lung microbiome and to pathogen details. *Microbiome* 5, 20

93 Langille, M.G.I. *et al.* (2013) Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology* 31, 814–821

94 Yatsunenko, T. *et al.* (2012) Human gut microbiome viewed across age and geography. *Nature* DOI: 10.1038/nature11053

95 Lloyd-Price, J. *et al.* (2017) Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature* DOI: 10.1038/nature23889

96 Feigelman, R. *et al.* (2017) Sputum DNA sequencing in cystic fibrosis: non-invasive access to the lung microbiome and to pathogen details. *Microbiome* 5, 20

97 Moran Losada, P. *et al.* (2016) The cystic fibrosis lower airways microbial metagenome. *ERJ Open Research* 2, 00096–02015

98 Horz, H.-P. *et al.* (2010) New methods for selective isolation of bacterial DNA from human clinical specimens. *Anaerobe* 16, 47–53

99 Ferretti, P. *et al.* (2017) Experimental metagenomics and ribosomal profiling of the human skin microbiome. *Experimental Dermatology* 26, 211–219

100 Goltsman, D.S.A. *et al.* (2018) Metagenomic analysis with strain-level resolution reveals fine-scale variation in the human pregnancy microbiome. DOI: 10.1101/266700

101 Bright, A.T. *et al.* (2012) Whole genome sequencing analysis of Plasmodium vivax using whole genome capture. *BMC Genomics* 13, 262–262

102 Jervis-Bardy, J. *et al.* (2015) Deriving accurate microbiota profiles from human samples with low bacterial content through post-sequencing processing of Illumina MiSeq data. *Microbiome* 3,

103 Bhatt, A.S. *et al.* (2014) In search of a candidate pathogen for giant cell arteritis: sequencing-based characterization of the giant cell arteritis microbiome. *Arthritis Rheumatol* 66, 1939–44

104 Chiodini, R.J. *et al.* (2013) Crohn's disease may be differentiated into 2 distinct biotypes based on the detection of bacterial genomic sequences and virulence genes within submucosal tissues. *Journal of clinical gastroenterology* 47, 612–620

105 Rogers, G.B. *et al.* (2010) Revealing the dynamics of polymicrobial infections: implications for antibiotic therapy. *Trends in Microbiology* 18, 357–364

106 Hasan, M.R. *et al.* (2016) Depletion of Human DNA in Spiked Clinical Specimens for Improvement of Sensitivity of Pathogen Detection by Next-Generation Sequencing. *Journal of Clinical Microbiology* 54, 919–927

107 Lim, Y.W. *et al.* (2014) Purifying the Impure: Sequencing Metagenomes and Metatranscriptomes from Complex Animal-associated Samples. *Journal of Visualized Experiments* DOI: 10.3791/52117

108 Marotz, C.A. *et al.* (2018) Improving saliva shotgun metagenomics by chemical host DNA depletion. *Microbiome* 6,

109 Thoendel, M. *et al.* (2016) Comparison of microbial DNA enrichment tools for metagenomic whole genome sequencing. *Journal of Microbiological Methods* 127, 141–145

110 Zhou, L. and Pollard, A.J. (2012) A novel method of selective removal of human DNA improves PCR sensitivity for detection of Salmonella Typhi in blood samples. *BMC Infectious Diseases* 12, 164

111 Truong, D.T. *et al.* (2015) MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat Methods* 12, 902–3

112 Mulcahy, H. *et al.* (2008) Extracellular DNA Chelates Cations and Induces Antibiotic Resistance in Pseudomonas aeruginosa Biofilms. *PLoS Pathogens* 4, e1000213

113 van Tilburg Bernardes, E. *et al.* (2017) Exopolysaccharide-Repressing Small Molecules with Antibiofilm and Antivirulence Activity against Pseudomonas aeruginosa. *Antimicrobial Agents and Chemotherapy* 61, e01997-16

114 Haque, H. and Russell, A.D. (1974) Effect of Ethylenediaminetetraacetic Acid and Related Chelating Agents on Whole Cells of Gram-Negative Bacteria. *Antimicrobial Agents and Chemotherapy* 5, 447–452

115 Banin, E. *et al.* (2006) Chelator-induced dispersal and killing of Pseudomonas aeruginosa cells in a biofilm. *Appl Environ Microbiol* 72, 2064–9

116 Burns, J.L. and Rolain, J.-M. (2014) Culture-based diagnostic microbiology in cystic fibrosis: Can we simplify the complexity? *Journal of Cystic Fibrosis* 13, 1–9

117 Glassing, A. *et al.* (2015) Changes in 16s RNA Gene Microbial Community Profiling by Concentration of Prokaryotic DNA. *Journal of Microbiological Methods* 119, 239–242

118 Nocker, A. *et al.* (2007) Molecular monitoring of disinfection efficacy using propidium monoazide in combination with quantitative PCR. *Journal of Microbiological Methods* 70, 252–260

119 Bellehumeur, C. *et al.* (2015) Propidium monoazide (PMA) and ethidium bromide monoazide (EMA) improve DNA array and high-throughput sequencing of porcine reproductive and respiratory syndrome virus identification. *Journal of Virological Methods* 222, 182–191

120 Exterkate, R.A.M. *et al.* (2014) The effect of propidium monoazide treatment on the measured bacterial composition of clinical samples after the use of a mouthwash. *Clinical Oral Investigations* 19, 813–822

121 Rogers, G.B. *et al.* (2008) Assessing the diagnostic importance of nonviable bacterial cells in respiratory infections. *Diagnostic Microbiology and Infectious Disease* 62, 133–141

122 Jovel, J. *et al.* (2016) Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics. *Frontiers in Microbiology* 7,

123 McArthur, A.G. *et al.* (2013) The Comprehensive Antibiotic Resistance Database. *Antimicrobial Agents and Chemotherapy* 57, 3348–3357

124 Chmiel, J.F. *et al.* (2014) Antibiotic Management of Lung Infections in Cystic Fibrosis. I. The Microbiome, Methicillin-ResistantStaphylococcus aureus, Gram-Negative Bacteria, and Multiple Infections. *Annals of the American Thoracic Society* 11, 1120–1129

125 Nielsen, S. *et al.* (2016) Achromobacter Species Isolated from Cystic Fibrosis Patients Reveal Distinctly Different Biofilm Morphotypes. *Microorganisms* 4, 33

126 Emerson, J.B. *et al.* (2017) Schrödinger's microbes: Tools for distinguishing the living from the dead in microbial ecosystems. *Microbiome* 5,

127 Soejima, T. *et al.* (2007) Photoactivated Ethidium Monoazide Directly Cleaves Bacterial DNA and Is Applied to PCR for Discrimination of Live and Dead Bacteria. *Microbiology and Immunology* 51, 763–775

128 Erkus, O. *et al.* (2016) Use of propidium monoazide for selective profiling of viable microbial cells during Gouda cheese ripening. *International Journal of Food Microbiology* 228, 1–9

129 Leo, S. *et al.* (2017) Detection of Bacterial Pathogens from Broncho-Alveolar Lavage by Next-Generation Sequencing. *International Journal of Molecular Sciences* 18, 2011

130 Zhao, J. *et al.* (2012) Impact of Enhanced Staphylococcus DNA Extraction on Microbial Community Measures in Cystic Fibrosis Sputum. *PLoS ONE* 7, e33127

131 Thompson, L.R. *et al.* (2017) A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 551, 457 EP-

132 Callahan, B.J. *et al.* (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods* 13, 581–583

133 Wang, Q. *et al.* (2007) Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Applied and Environmental Microbiology* 73, 5261–5267

134 Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag New York.

135 Oksanen, J. *et al.* (2017) *vegan: Community Ecology Package*,

136 Paradis, E. *et al.* (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290

137 Yu, G. *et al.* (2016) ggtree: anrpackage for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* 8, 28–36

138 Hunter, S.J. *et al.* (2011) Selective removal of human DNA from metagenomic DNA samples extracted from dental plaque. *Journal of Basic Microbiology* 51, 442–446

139 Nadkarni, M.A. *et al.* (2002) Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology* 148, 257–266

140 Handschur, M. *et al.* (2009) Preanalytic removal of human DNA eliminates false signals in general 16S rDNA PCR monitoring of bacterial pathogens in blood. *Comparative Immunology, Microbiology and Infectious Diseases* 32, 207–219

141 Burns, M. and Valdivia, H. (2007) Modelling the limit of detection in real-time quantitative PCR. *European Food Research and Technology* 226, 1513–1524

142 Bustin, S.A. *et al.* (2009) The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. *Clinical Chemistry* 55, 611–622

143 Bolger, A.M. *et al.* (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120

144 Kozich, J.J. *et al.* (2013) Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. *Applied and Environmental Microbiology* 79, 5112–5120

145 Quast, C. *et al.* (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* 41, D590–D596

146 Kircher, M. *et al.* (2012) Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research* 40, e3–e3

147 Sinclair, L. *et al.* (2015) Microbial Community Composition and Diversity via 16S rRNA Gene Amplicons: Evaluating the Illumina Platform. *PLoS ONE* 10, e0116955

148 Salter, S.J. *et al.* (2014) Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC biology* 12, 1

149 R Core Team (2017) *R: A Language and Environment for Statistical Computing*,

150 Li, D. *et al.* (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676

151 Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9, 357 EP-

152 Hyatt, D. *et al.* (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11, 119

153 Kim, D. *et al.* (2016) Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Research* at <http://genome.cshlp.org/content/early/2016/11/16/gr.210641.116.abstract N2 - Centrifuge is a novel microbial classification engine that enables rapid, accurate, and sensitive labeling of reads and quantification of species on desktop computers. The system uses an indexing scheme based on the Burrows-Wheeler transform (BWT) and the Ferragina-Manzini (FM) index, optimized specifically for the metagenomic classification problem. Centrifuge requires a relatively small index (4.2 GB for 4078 bacterial and 200 archaeal genomes) and classifies sequences at very high speed, allowing it to process the millions of reads from a typical high-throughput DNA sequencing run within a few minutes. Together, these advances enable timely and accurate analysis of large metagenomics data sets on conventional desktop computers. Because of its space-optimized indexing schemes, Centrifuge also makes it possible to index the entire NCBI nonredundant nucleotide sequence database (a total of 109 billion bases) with an index size of 69 GB, in contrast to k-mer-based indexing schemes, which require far more extensive space.>

154 Eren, A.M. *et al.* (2015) Anvi'o: an advanced analysis and visualization platform for `omics data. *PeerJ* 3, e1319

155 Cystic Fibrosis Foundation (2015) *Cystic Fibrosis Foundation Patient Registry 2015 Annual Data Report*, ©2016 Cystic Fibrosis Foundation.

156 Lee, S. *et al.* (2009) Targeting a bacterial stress response to enhance antibiotic action. *Proceedings of the National Academy of Sciences* 106, 14570–14575

157 Ratjen, A. *et al.* (2015) In vitro efficacy of high-dose tobramycin against Burkholderia cepacia complex and Stenotrophomonas maltophilia isolates from cystic fibrosis patients. *Antimicrobial agents and chemotherapy* 59, 711–713

158 Horrevorts, A.M. and Mouton, J. (1991) Susceptibility to various antimicrobial agents and tolerance to methicillin of Staphylococcus aureus isolates from cystic fibrosis patients. *European Journal of Clinical Microbiology & Infectious Diseases: an international journal on pathogenesis, diagnosis, epidemiology, therapy, and prevention of infectious diseases* 10, 785–786

159 Enright, P.L. *et al.* (2004) Repeatability of Spirometry in 18,000 Adult Patients. *American Journal of Respiratory and Critical Care Medicine* 169, 235–238

160 Fuchs, H.J. *et al.* (1994) Effect of aerosolized recombinant human DNase on exacerbations of respiratory symptoms and on pulmonary function in patients with cystic fibrosis. *New England Journal of Medicine* 331, 637–642

161 Nelson, M.T. *et al.* (2019) Human and Extracellular DNA Depletion for Metagenomic Analysis of Complex Clinical Infection Samples Yields Optimized Viable Microbiome Profiles. *Cell Reports* 26, 2227-2240.e5

162 Li, L. *et al.* (2014) The importance of the viable but non-culturable state in human bacterial pathogens. *Frontiers in microbiology* 5, 258

163 Ayrapetyan, M. *et al.* (2015) Bridging the gap between viable but non-culturable and antibiotic persistent bacteria. *Trends in microbiology* 23, 7–13

164 Gloor, G.B. *et al.* (2017) Microbiome datasets are compositional: and this is not optional. *Frontiers in microbiology* 8, 2224

165 Aitchison, J. *et al.* (2000) Logratio analysis and compositional distance. *Mathematical Geology* 32, 271–275

166 Mahboubi, M.A. *et al.* (2016) Culture-Based and Culture-Independent Bacteriologic Analysis of Cystic Fibrosis Respiratory Specimens. *Journal of Clinical Microbiology* 54, 613–619

167 Pasquaroli, S. *et al.* (2013) Antibiotic pressure can induce the viable but non-culturable state in Staphylococcus aureus growing in biofilms. *Journal of Antimicrobial Chemotherapy* 68, 1812–1817

168 Brito, I.L. *et al.* (2016) Mobile genes in the human microbiome are structured from global to individual scales. *Nature* 535, 435–439

169 Turnbaugh, P.J. *et al.* (2009) A core gut microbiome in obese and lean twins. *Nature* 457, 480–484

170 Worlitzsch, D. *et al.* (2009) Antibiotic-resistant obligate anaerobes during exacerbations of cystic fibrosis patients. *Clinical Microbiology and Infection* 15, 454–460

171 Mirković, B. *et al.* (2015) The Role of Short-Chain Fatty Acids, Produced by Anaerobic Bacteria, in the Cystic Fibrosis Airway. *American Journal of Respiratory and Critical Care Medicine* 192, 1314–1324

172 Sibley, C.D. *et al.* (2008) A polymicrobial perspective of pulmonary infections exposes an enigmatic pathogen in cystic fibrosis patients. *Proceedings of the National Academy of Sciences* 105, 15070–15075

173 Finland, M. *et al.* (1976) Susceptibility of pneumococci and Haemophilus influenzae to antibacterial agents. *Antimicrobial agents and chemotherapy* 9, 274–287

174 Murray, M.P. *et al.* (2010) Do processing time and storage of sputum influence quantitative bacteriology in bronchiectasis? *Journal of medical microbiology* 59, 829–833

175 Zhao, J. *et al.* (2011) Effect of sample storage conditions on culture-independent bacterial community measures in cystic fibrosis sputum specimens. *Journal of clinical microbiology* 49, 3717–3718

176 Tunney, M.M. *et al.* (2011) Use of culture and molecular analysis to determine the effect of antibiotic treatment on microbial community diversity and abundance during exacerbation in patients with cystic fibrosis. *Thorax* 66, 579–584

177 Fodor, A.A. *et al.* (2012) The Adult Cystic Fibrosis Airway Microbiota Is Stable over Time and Infection Type, and Highly Resilient to Antibiotic Treatment of Exacerbations. *PLoS ONE* 7, e45001

178 Quinn, R.A. *et al.* (2016) Ecological networking of cystic fibrosis lung infections. *NPJ biofilms and microbiomes* 2, 1–11

179 Coburn, B. *et al.* (2015) Lung microbiota across age and disease stage in cystic fibrosis. *Scientific Reports* 5,

180 Saiman, L. *et al.* (2014) Infection Prevention and Control Guideline for Cystic Fibrosis: 2013 Update. *Infection Control & Hospital Epidemiology* 35, S1–S67

181 Brakstad, O.G. *et al.* (1992) Detection of Staphylococcus aureus by polymerase chain reaction amplification of the nuc gene. *Journal of clinical microbiology* 30, 1654–1660

182 Szczesniak, R. *et al.* (2017) *Use of FEV1 in cystic fibrosis epidemiologic studies and clinical trials: A statistical perspective for the clinical researcher*, 16

183 Goss, C.H. *et al.* (2009) Patient-reported respiratory symptoms in cystic fibrosis. *Journal of Cystic Fibrosis* 8, 245–252

184 Johnson, E.J. *et al.* (2016) Molecular Identification of Staphylococcus aureus in Airway Samples from Children with Cystic Fibrosis. *PLOS ONE* 11, e0147643

185 Anuj, S.N. *et al.* (2009) Identification of Pseudomonas aeruginosa by a duplex real-time polymerase chain reaction assay targeting the ecfX and the gyrB genes. *Diagnostic Microbiology and Infectious Disease* 63, 127–131

186 Olson, A.B. *et al.* (2010) Development of Real-Time PCR Assays for Detection of the Streptococcus milleri Group from Cystic Fibrosis Clinical Specimens by Targeting the cpn60 and 16S rRNA Genes. *Journal of Clinical Microbiology* 48, 1150–1160

187 Martin, F.E. *et al.* (2002) Quantitative microbiological study of human carious dentine by culture and real-time PCR: association of anaerobes with histopathological changes in chronic pulpitis. *Journal of clinical microbiology* 40, 1698–1704

188 Chouvarine, P. *et al.* (2016) Filtration and normalization of sequencing read data in whole-metagenome shotgun samples. *PloS one* 11,

189 Simon, H.Y. *et al.* (2019) Benchmarking metagenomics tools for taxonomic classification. *Cell* 178, 779–794

190 Teo, Y.V. and Neretti, N. (2016) A comparative study of metagenomics analysis pipelines at the species level. *bioRxiv*

191 Tamames, J. *et al.* (2019) Assessing the performance of different approaches for functional and taxonomic annotation of metagenomes. *BMC genomics* 20, 1–16

192 Palarea-Albaladejo, J. and Martín-Fernández, J.A. (2015) zCompositions—R package for multivariate imputation of left-censored data under a compositional approach. *Chemometrics and Intelligent Laboratory Systems* 143, 85–96

193 van den Boogaart, K.G. and Tolosana-Delgado, R. (2008) “Compositions”: a unified R package to analyze compositional data. *Computers & Geosciences* 34, 320–338

194 Fernandes, A.D. *et al.* (2014) Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome* 2, 15

195 Fernandes, A.D. *et al.* (2013) ANOVA-like differential expression (ALDEx) analysis for mixed population RNA-Seq. *PLoS One* 8, e67019

196 Truong, D.T. *et al.* (2017) Microbial strain-level population structure and genetic diversity from metagenomes. *Genome research* 27, 626–638

197 Robinson, M.D. *et al.* (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)* 26, 139–40

198 Jorth, P. *et al.* (2015) Regional Isolation Drives Bacterial Diversification within Cystic Fibrosis Lungs. *Cell Host & Microbe* 18, 307–319

199 Carmody, L.A. *et al.* (2013) Changes in Cystic Fibrosis Airway Microbiota at Pulmonary Exacerbation. *Annals of the American Thoracic Society* 10, 179–187

200 Blair, J.M. *et al.* (2015) Molecular mechanisms of antibiotic resistance. *Nature reviews microbiology* 13, 42

201 Munita, J.M. and Arias, C.A. (2016) Mechanisms of antibiotic resistance. *Microbiology spectrum* 4,

202 Kanehisa, M. *et al.* (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research* 40, D109–D114

203 Manor, O. and Borenstein, E. (2017) Revised computational metagenomic processing uncovers hidden and biologically meaningful functional variation in the human microbiome. *Microbiome* 5,

204 Khan, W. *et al.* (2010) Aminoglycoside resistance of Pseudomonas aeruginosa biofilms modulated by extracellular polysaccharide. *International microbiology: the official journal of the Spanish Society for Microbiology* 13, 207

205 Bryan, L. *et al.* (1984) Lipopolysaccharide changes in impermeability-type aminoglycoside resistance in Pseudomonas aeruginosa. *Antimicrobial agents and chemotherapy* 26, 250–255

206 Wright, G.D. (1999) Aminoglycoside-modifying enzymes. *Current opinion in microbiology* 2, 499–503

207 Parker, A.C. and Smith, C.J. (1993) Genetic and biochemical analysis of a novel Ambler class A beta-lactamase responsible for cefoxitin resistance in Bacteroides species. *Antimicrobial agents and chemotherapy* 37, 1028–1036

208 Roberts, M.C. (2005) Update on acquired tetracycline resistance genes. *FEMS microbiology letters* 245, 195–203

209 Allison, D. and Matthews, M. (1992) Effect of polysaccharide interactions on antibiotic susceptibility of Pseudomonas aeruginosa. *Journal of applied bacteriology* 73, 484–488

210 Tkhilaishvili, T. *et al.* (2018) Bacteriophage Sb-1 enhances antibiotic activity against biofilm, degrades exopolysaccharide matrix and targets persisters of Staphylococcus aureus. *International journal of antimicrobial agents* 52, 842–853

211 Venkatesan, N. *et al.* (2015) Bacterial resistance in biofilm-associated bacteria. *Future microbiology* 10, 1743–1750

212 Mah, T.-F.C. and O'Toole, G.A. (2001) Mechanisms of biofilm resistance to antimicrobial agents. *Trends in microbiology* 9, 34–39

213 Mah, T.-F. *et al.* (2003) A genetic basis for Pseudomonas aeruginosa biofilm antibiotic resistance. *Nature* 426, 306–310

214 Garo, E. *et al.* (2007) Asiatic acid and corosolic acid enhance the susceptibility of Pseudomonas aeruginosa biofilms to tobramycin. *Antimicrobial agents and chemotherapy* 51, 1813–1817

215 de Breij, A. *et al.* (2016) The licorice pentacyclic triterpenoid component 18β-glycyrrhetinic acid enhances the activity of antibiotics against strains of methicillin-resistant Staphylococcus aureus. *European Journal of Clinical Microbiology & Infectious Diseases* 35, 555–562

216 Resende, J.A. *et al.* (2014) Dynamics of antibiotic resistance genes and presence of putative pathogens during ambient temperature anaerobic digestion. *Journal of applied microbiology* 117, 1689–1699

217 Buelow, E. *et al.* (2017) Comparative gut microbiota and resistome profiling of intensive care patients receiving selective digestive tract decontamination and healthy subjects. *Microbiome* 5, 88

218 Banerjee, S. *et al.* (2018) Keystone taxa as drivers of microbiome structure and functioning. *Nature Reviews Microbiology* 16, 567–576

219 Manor, O. and Borenstein, E. (2015) MUSiCC: a marker genes based framework for metagenomic normalization and accurate profiling of gene abundances in the microbiome. *Genome Biology* 16,

220 Manor, O. *et al.* (2016) Metagenomic evidence for taxonomic dysbiosis and functional imbalance in the gastrointestinal tracts of children with cystic fibrosis. *Scientific reports* 6, 22493

221 Smith, D.J. *et al.* (2014) Pyrosequencing reveals transient cystic fibrosis lung microbiome changes with intravenous antibiotics. *European Respiratory Journal* 44, 922–930

222 Chiang, W.-C. *et al.* (2013) Extracellular DNA Shields against Aminoglycosides in Pseudomonas aeruginosa Biofilms. *Antimicrobial Agents and Chemotherapy* 57, 2352–2361

223 Alhede, M. *et al.* (2014) Pseudomonas aeruginosa biofilms: mechanisms of immune evasion. In *Advances in applied microbiology* 86pp. 1–40, Elsevier

224 Kavanaugh, J.S. *et al.* (2019) Identification of extracellular DNA-binding proteins in the biofilm matrix. *MBio* 10, e01137-19

225 Heirali, A.A. *et al.* (2017) The effects of inhaled aztreonam on the cystic fibrosis lung microbiome. *Microbiome* 5,

226 Heirali, A.A. *et al.* The effects of inhaled aztreonam on the cystic fibrosis lung microbiome. 5,

227 Clancy, J.P. *et al.* (2007) No detectable improvements in cystic fibrosis transmembrane conductance regulator by nasal aminoglycosides in patients with cystic fibrosis with stop mutations. *American journal of respiratory cell and molecular biology* 37, 57–66

228 Umeki, S. (1995) Anti-inflammatory action of gentamycin through inhibitory effect on neutrophil NADPH oxidase activity. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* 110, 817–821