

IN VIVO EVOLUTION OF *BURKHOLDERIA*
MULTIVORANS IN THE CYSTIC FIBROSIS LUNG

by

Sarah Harrison

A thesis submitted to the faculty of
The University of North Carolina at Charlotte in partial
fulfillment of the requirements
for the degree of Master of Science in

Biology

Charlotte

2021

Approved by:

Dr. Todd Steck

Dr. Way Sung

Dr. Matthew Parrow

ABSTRACT

SARAH HARRISON. The In-vivo Evolution of *Burkholderia multivorans* in a Cystic Fibrosis Patient. (Under the direction of DR. TODD R. STECK)

Burkholderia multivorans is an opportunistic pathogen that poses a health risk to patients with cystic fibrosis (CF). The use of broad-range antibiotics to combat chronic bacterial lung infections has led to the evolution of antibiotic-resistant and multi-drug resistant *Burkholderia multivorans* (ABR and MDR). Understanding how these bacteria respond genetically to antibiotic therapy would be useful in optimizing drug therapy as well as to provide insight into how an established, complex, and chronic bacterial community adapts in response to multiple drug treatments. To this extent, we analyzed the *in vivo* evolution of *B. multivorans* during multiple pulmonary exacerbations over three years in a CF patient who underwent extensive antibiotic treatment. We found the population diversified into at least two coexisting lineages with an average of 2.7 SNPs/year. Within these lineages, we found that 12 genes had a significant excess of mutations. Importantly, we observed an excess of mutations in the *dacB* gene and one in *ampD*, in strains isolated following IV administration of ceftazidime (CAZ) during a 2-week hospitalization. Both *dacB* and *ampD* have previously been reported to regulate β -lactamase expression, where deactivating mutants leads to hyper-inducible AmpC and/or PenA in other *Burkholderia* and gram-negative bacteria. These data suggest possible β -lactamase regulators' importance to resistance to β -lactam antibiotics in *B. multivorans*.

TABLE OF CONTENTS

<i>LIST OF TABLE</i>	vi
<i>LIST OF FIGURES</i>	vii
<i>LIST OF ABBREVIATIONS</i>	viii
<i>CHAPTER 1: INTRODUCTION</i>	1
<i>CHAPTER 2: MATERIALS AND METHODS</i>	4
2.1: <i>Sample Collection and DNA Extraction</i>	4
2.2: <i>Genome Sequencing, Assembly, and SNP/INDEL Calling</i>	6
2.3: <i>Detection of Structural Variants</i>	6
2.4: <i>Phylogeny Analysis</i>	7
2.5: <i>Antibiotic Susceptibility Testing</i>	8
<i>CHAPTER 3: RESULTS</i>	9
3.1: <i>All Isolates Descend from A Single Strain</i>	9
3.2: <i>Population Structure of Study Isolates</i>	10
3.3: <i>Mutation Frequency</i>	12
3.4 <i>Frequency of SNPS/INDELS and Functional Characterization</i>	13
3.5: <i>Multi-Mutated Gene</i>	17
3.6 <i>Frequency of Structural Variations</i>	19
3.7: <i>Antibiotic Susceptibility</i>	21
<i>CHAPTER 4: DISCUSSION</i>	25
<i>SUPPLEMENTAL INFORMATION</i>	28

1: IGV Image of dacB mutations28

2: Summary statistics of Isolates.....29

3: SNP distribution and overlaps between isolates30

REFERENCES.....31

LIST OF TABLES

TABLE 1: Summary of Isolate Information5

TABLE 2: Genes with Significant Number of Mutations18

TABLE 3: Regions with Structural Variation.....20

LIST OF FIGURES

FIGURE 1: Phylogenetic Analysis in context with other <i>Burkholderia multivorans</i> genomes	10
FIGURE 2: SNP-Phylogeny of Study Isolates AS218-AS240.....	11
FIGURE 3: SNP Accumulation Overtime.....	13
FIGURE 4: Sample Mutation Counts Grouped by COG.....	15
FIGURE 5: Overall COG distribution	16
FIGURE 6: Antibiotic Resistance Profiles Overtime	23
FIGURE 7: PIP and CAZ Resistant Grouped by Mutation	24

LIST OF ABBREVIATIONS

ABR: Antibiotic-resistant

MDR: Multidrug resistant

CF: Cystic Fibrosis

PE: Pulmonary exacerbation

Bcc: *Burkholderia cepacia* complex

SNP: Single-nucleotide polymorphism

INDEL: Insertion/deletion

SV: Structural Variation

COG: Clusters of orthologous groups

MIC: Minimum Inhibitory Concentration

ZOI: Zone of Inhibition

CHAPTER 1 INTRODUCTION

Chronic airway colonization with members of the *Burkholderia cepacia* complex (Bcc) is often associated with a severe decline in lung functionality and reduced survival rates in patients with cystic fibrosis (CF) (Jones et al., 2004). Less than 3% of CF patients become infected with *Burkholderia* and infection usually occurs during or after adolescence (Cystic Fibrosis Foundation Patient Registry 2019 Annual Data Report). *B. multivorans* and *B. cenocepacia* account for a majority of infections by the complex and are associated with lethal septicemia known as “cepacia syndrome” (Mahenthiralingam et al., 2008; Shafiq et al., 2011). However, in recent years *B. multivorans* have surpassed *B. cenocepacia* in new infections, becoming the most prevalent of the Bcc (Kenna et al., 2017). The use of antibiotics in CF patients is common, both prophylactic and in response to pulmonary exacerbations (PE). Understanding how these bacteria evolve ABR over time would contribute useful information for optimizing drug therapy. This is needed to combat bacterial pathogenicity and reduce health care costs.

Frequent therapy using multiple drugs creates a fluctuating environment that selects for resistant mutants by eradicating sensitive sub-populations (Lamrabet et al., 2019). Single-nucleotide polymorphisms (SNPs) that arise in these environments are subject to the primary evolutionary forces of selection and drift. Any mutation that becomes fixed is expected to be associated with replication speed, biofilm formation, virulence, or antibiotic resistance (ABR) (Podnecky et al., 2015; Rhodes & Schweizer, 2016). Indicating the importance of understanding the evolutionary dynamics of bacterial communities during strong selective pressure. To date there is limited data on the evolution and diversification of *B. multivorans* in vivo following extensive antibiotic therapy.

In recent years there have been multiple studies documenting the evolution of bacteria within Bcc in the CF lung (Diaz Caballero et al., 2018; Hassan et al., 2020; Silva et al., 2016). A comparative genomics study analyzed the whole genome sequence (WGS) of 22 isolates of *B. multivorans* recovered from a single CF patient over 20 years and found that the population diversified into 3 coexisting subpopulations. They identified parallel adaptive variations targeting adherence, metabolism, and changes in the cell envelope related to adaptation to the biofilm lifestyle (Silva et al., 2016). A genome-wide variation study of 111 *B. multivorans* isolates from a CF patient over a ten-year time also reported the incident isolate to diversify into multiple distinct lineages and found parallel patho-adaptation in genes associated with virulence and resistance. They found polymorphisms in the *araC* and *ampD* genes to be associated with increased resistance to β -lactam, aminoglycoside, and fluoroquinolone antibiotics (Diaz Caballero et al., 2018). Another comparative genomics study of 11 *B. cenocepacia* and 9 *B. multivorans* isolates during a co-infection over 4.4 years saw the majority of genes containing mutations to be associated with the oxidative stress response, transition metal metabolism, and antibiotic resistance (Hassan et al., 2020). These studies all demonstrated an original infecting strain to establish a population that diversified into multiple distinct lineages that differ considerably both genotypically and phenotypically. This shows the complexity of chronic and antibiotic-resistant *B. multivorans* infections where genetically diverse sub-lineages can co-exist during and after antibiotic treatment.

The genomics variability and stability within these sub-populations during exposures to antibiotic treatments remains widely unexplored. Understanding how changes in the underlying population structure drive the evolutionary trajectory of *B. multivorans* have important implications for the development of novel treatment options. To this extent, we monitored the

evolution of *B. multivorans* over six different pulmonary exacerbation periods that involved multiple applications of different antibiotics. We tracked the clonality, evolution, and genome-wide variation of these isolates, to provide insight into the micro-evolution and adaptation of *B. multivorans*, and highlight novel mechanisms enabling survival in the CF lung during extensive antibiotic therapy. My goal is to better understand these factors and identify genes under strong selection, which is a necessary step in tailoring therapeutics.

CHAPTER 2 MATERIALS AND METHODS

2.1: Sample Collection and DNA Extraction

Isolates from sputum samples of a single adult male CF patient, age 24 at the beginning of the sampling period. His treatment regimen during the study included oral enzymes for CF-related malabsorption, along with antibiotics administered as prophylactic agents and various antibiotics (Table 1.). The patient has a heterozygous deltaF508/unknown CF transmembrane conductance regulator (CFTR) genotype. Patient FEV1 values ranged from 33%-40% during the time of the study, which is consistent with severe obstruction. He experienced nine PEs that required antibiotic intervention (6 of which we sequenced samples). Serial expectorated sputum samples were obtained with informed consent, twice weekly for almost 3 years (IRB Protocol Approved #11-12-36). Samples were collected in the morning by the patient expectorating sputum into a sterile 15-ml Falcon tube, which was placed on ice during transport to the lab, and stored at -80C until use. Sputum samples of desired time points were allowed to thaw, and an aliquot was struck out for single colonies onto *Burkholderia cepacia* selective agar (BCSA) plates and incubated at 37C for 48h. Single colonies with morphology consistent with *B. multivorans* were picked and replated as described above. Single colonies were chosen and inoculated into 2 mL of LB broth and incubated at 37C for 24h. DNA extraction followed standard procedures for microbes.

TABLE 1: Isolate Information

The patient was sampled during six pulmonary exacerbation (PE) periods displayed below in different colors. The filled blue lines indicate that the patient was taking the antibiotic on the left side during that PE period, with the provided administration route (oral or intravenous). The volume exhaled at the end of the first second of forced expiration (FEV1 %) values were measured during each PE event, and provided below each PE event. The date that each of the 20 samples is matched (by color) with the PE event from which it was sampled. There were two samples not taken during a PE event (AS142, and AS154). For samples that were taken during a PE, it is indicated whether it was sampled at the beginning or end of the PE period.

		Patient antibiotic use for PE period (m/day/yr)					
Administration route	Antibiotic	3/15/09-4/5/09	1/7/10-1/23/10	7/9/10-7/15/10	3/3/11-3/8/11	7/27/11-8/1/11	12/22/11-1/4/12
Oral:	Ciprofloxacin	[Blue bar]				[Blue bar]	
	Bactrim	[Blue bar]					
	Minocycline		[Blue bar]			[Blue bar]	[Blue bar]
I.V.:	Ceftazidime		[Blue bar]			[Blue bar]	
FEV1 %:		38-40	22-39	38	37	32-36	34-35
Strain ID:	Date Isolation						
AS142	3/7/09						
AS218	4/3/09	end					
AS219	4/4/09	end					
AS222	1/7/10	onset					
AS223	1/7/10	onset					
AS224	7/9/10			onset			
AS225	7/13/10			end			
AS226	7/13/10			end			
AS227	7/14/10			end			
AS228	3/3/11				onset		
AS229	3/3/11				onset		
AS230	3/4/11				onset		
AS231	3/8/11				end		
AS232	3/8/11				end		
AS154	5/26/11						
AS233	7/27/11					onset	
AS236	8/1/11					end	
AS237	12/23/11						onset
AS240	1/2/12						end

2.2: Genome Sequencing, Assembly, and SNP/INDEL calling

Isolated single colonies were sent to Omega Bioservices, where whole-genome sequencing (WGS) was performed using 151 bp paired-end reads with an Illumina HiSeq 2500 platform. The quality of raw files was assessed using FastQC-0.11.8 (Andrews, 2010). Adapters and low-quality reads were removed using Trimmomatic-0.35 (Bolger et al., 2014). BWA mem was used for globally aligning each sample's reads to the reference genome of *Burkholderia multivorans* FDAARGOS_246 (NCBI Accession: PRJNA231221) (Li & Durbin, 2009). The alignment files were sorted using Samtools sort and converted to a tab-delimited pileup format using Samtools mpileup (Li et al., 2009). Custom python scripts were used to parse out variants from the mpileup files, which were then merged to display the variant calls of every strain. Several criteria were established to consider a SNP. There had to be at least two forward and two reverse reads. The variant must also meet a coverage threshold of three standard deviations from the mean coverage, and >70% allele frequency. An SNP call was excluded if all strains called the variant, as they were considered to be ancestral. Similar criteria were applied for the identification and filtering of INDEL calls.

2.3: Detection of Structural Variation

Structural variants (SVs) involve duplication, deletion, or translocation of large-scale fragments of the genome, ranging from >10bp to entire chromosomes. I wrote a custom pipeline to detect and verify these SV's. Full description of pipeline and code for merging/overlap of calls available at <https://github.com/skharrison/SV>. SVs were identified using a variety of software packages including PINDEL, BREAKDANCER, GRIDSS, MANTA, LUMPY, and DELLY (D. L. Cameron et al., 2017; Chen et al., 2016; Fan et al., 2014; Layer et al., 2014; Rausch et al.,

2012; Ye et al., 2009). Files produced by each software were converted to compatible formats and Structural Variant Annotation R package used to annotate GRIDSS and LUMPY output before merging (D. Cameron & Dong, 2020). Sample calls from each software were merged at a threshold of 75% and kept if at least 4 callers identified a variant in that region. Samples merged calls (from all callers) were then overlapped at a threshold of 75% to all other sample total calls to produce a final list to be verified by eye. Bedtools (-igv command) was used to produce a script to automate the generation of region images using igvTools. Sample bams were loaded into IGV and images were produced to verify each region (Quinlan & Hall, 2010; Thorvaldsdóttir et al., 2013). Many regions showed to either be reference variants in all samples or false positives which were then removed from the analysis.

2.4 Phylogenetic Analysis

To determine if the strains are clonal all samples were assembled using SPADES with default parameters (Bankevich et al., 2012). Assemblies were input into the TYGS (Type Strain Genome Server) with six other reference *B. multivorans* genomes (NCBI Accessions: PRAJNA231221, PRJNA279182, PRJNA434393, PRJNA279182, PRJNA279182, PRJNA475602, PRJNA600378). The job was restricted to only genomes of interest (Meier-Kolthoff & Göker, 2019). TYGS phylogenetic inference done by resulting intergenomic distances was used to infer a balanced minimum evolution tree with branch support via FASTME 2.1.4 including SPR postprocessing (Lefort et al., 2015). Branch support was inferred from 100 pseudo- bootstrap replicates each and the tree was rooted at the midpoint. The whole genome-based phylogenetic tree generated from TYGS was used and downloaded from the TYGS webserver. To distinguish the phylogenetic relationship among isolates, consensus

sequences were first constructed by taking SNP calls for each sample and placing them into the reference genome. For each sample all gene sequences were separately pulled out from the consensus sequence, creating separate gene sequence fastas for each sample for all genes (including those that did not contain a mutation). All sample genes and the reference were aligned separately using MUSCLE (Edgar, 2004). R was used to concatenate each gene alignment and create a partition file to run RAxML to create a maximum likelihood tree (Stamatakis, 2014). The GTRGAMMA model was chosen and ran with 100 bootstrap replicates within RAxML. The best scoring output tree was rooted to the reference strain and visualized. All trees were loaded into FIGTREE to be displayed and colored (<http://tree.bio.ed.ac.uk/software/figtree/>).

2.5 Antibiotic Susceptibility Testing

Antimicrobial susceptibility testing was carried out via disk diffusion on Mueller-Hinton agar, plates were grown at 37 C and after 24 hours zones of inhibition (ZOI) were measured to obtain the antibiogram. The HardyDisk AST cartridges were used for all except meropenem and minocycline where Thermo Scientific Oxoid Discs were used. All AST ZOI measurement tests were performed using Clinical and Laboratory Standards Institute (CLSI) procedures (CLSI, 2020).

CHAPTER 3: RESULTS

3.1 All Isolates Descend from A Single Incident Isolate

To be able to compare mutations longitudinally requires having isolates that are known to descend from a single strain. A CF patient may be colonized by one, or more than one, *Burkholderia multivorans* strain. These two possibilities can be distinguished by comparing the genomic sequences of the 20 strains with those of multiple other *B. multivorans* strains. A single index pathogen would result in the 20 strains having greater similarity to each other than to any other *B. multivorans* strain. That the 20 strains comprise members of two or more independent *B. multivorans* infections would be reflected by there being a corresponding number of clades in a phylogenetic tree. We find all 20 isolates when compared to six reference strains, cluster into a single clade, consistent with all being evolved from a single infection event (Figure 1). Isolates also all shared >40,000 SNPs when compared to FDAARGO_246 reference strain, which is consistent with their descendants from a single incident isolate.

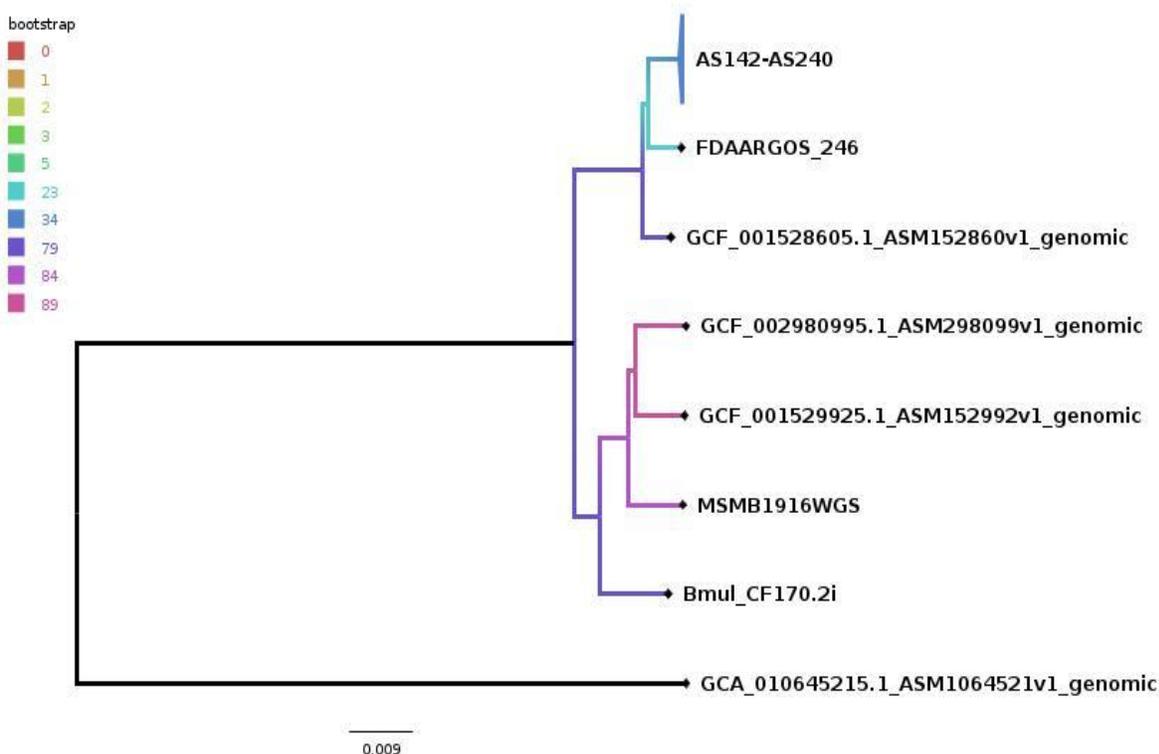


Figure 1. Tree inferred with FastME 2.1.6.1 from GBDP distances calculated from genome sequences. The branch lengths are scaled in terms of GBDP distance formula $d5$ (Meier-Kolthoff et al., 2013). The numbers above branches are GBDP pseudo-bootstrap support values > 60 % from 100 replications, with an average branch support of 17.0 %. The tree was rooted at the midpoint.

3.2 Population Structure of Study Isolates

To understand the adaptation of *B. multivorans* in the CF host during PE events the evolutionary relationship among the 20 isolates was determined. A maximum-likelihood tree was built using a total of 188 identified SNPs (Figure 2). The phylogeny indicates that at minimum the population diverged into 2 to 3 sub-populations (S1-S3). The differentiation between isolates sometimes did not match chronology. For example, AS222 and AS223 were isolated on the same day and are shown to belong to distinct lineages on the phylogeny. This indicates the long-term coexistence of multiple sub-populations that evolved from a single index strain.

One possible explanation for the coexistence of numerous sub-populations is the complex and dynamic nature of the CF lung could be promoting diversification into an array of specialist niches. Previous studies have shown that varying nutritional conditions, reduced dispersal, host immune response, and competing colonizing microbial species can all drive evolutionary diversification in the CF lung (Bernardy et al., 2020; Klockgether et al., 2018; Miller et al., 2015; Palmer et al., 2007) It has been hypothesized that this diversification which generates broad phenotypic and genotypic variability could be one of the first key steps for developing chronic infections (Schick & Kassen, 2018).

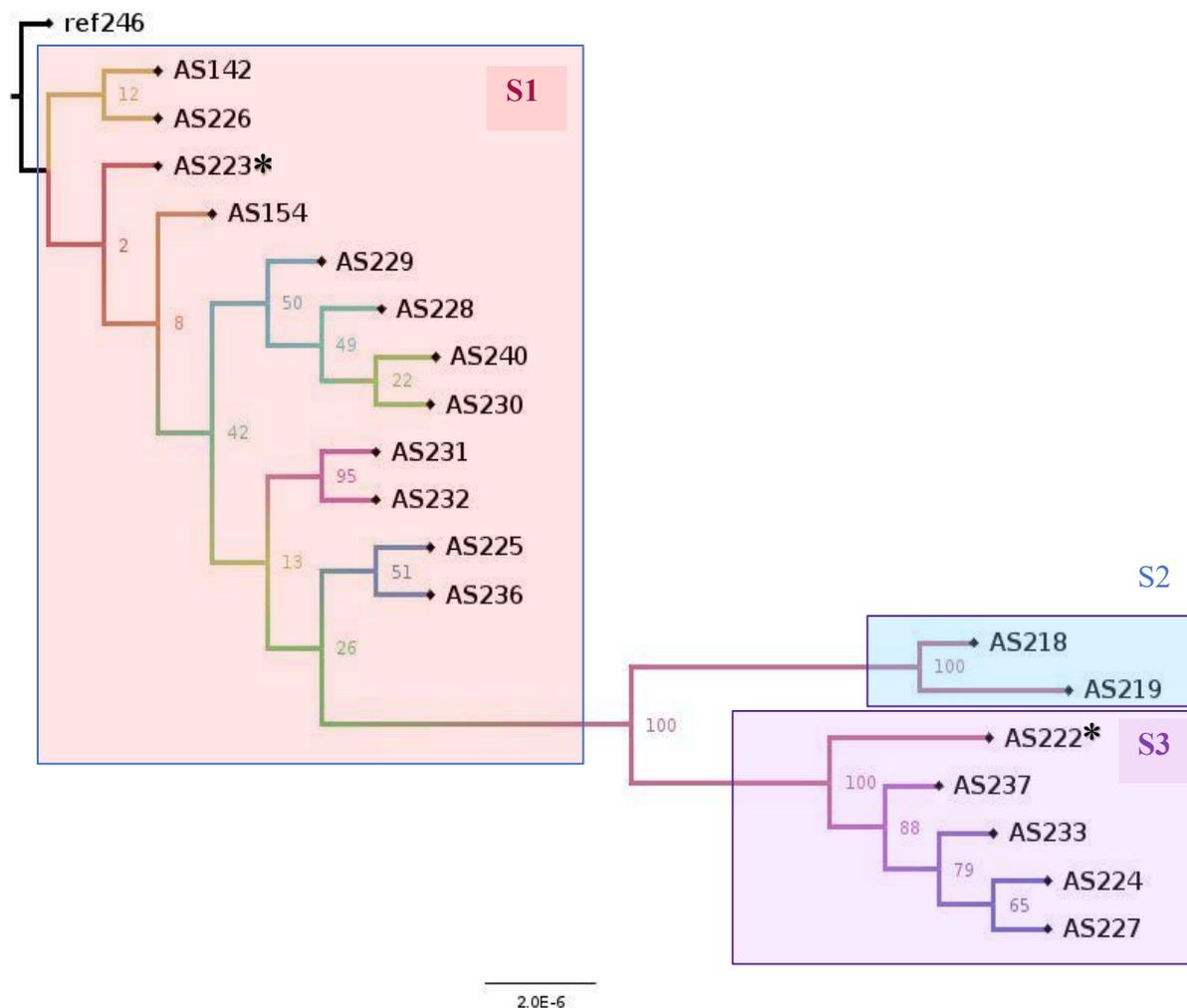


Figure 2. Maximum Likelihood phylogenetic tree inferred by RaxML showing coexisting clades within the population. The colors describe phylogenetic clades S1 to S3. The * denotes samples that were isolated from the patient on the same date and clustered on separate clades.

3.3: Mutation Rate within Lineages

To understand the overall mutation rate of the *B. multivorans* population during these three years, the number of cumulative SNPs for each sample was plotted over time of isolation. It was found that over the three-year study period there was an average accumulation of 2.7

SNPs/year (Figure 3.), regardless of sub-population (AS218 and AS219 excluded). This indicates that indifferent of extensive antibiotic pressure, the overall mutation rate remains consistent with a similar longitudinal study of *B. multivorans* in the CF lung, which has described a rate of 2.4 SNPs/year over 20 years (Silva et al., 2016). However, if split population into individual sub-population designated by the phylogeny, it is shown that isolates of the S3 lineage have a rate of 5.3 SNPs/year (orange in Figure 3.), while the S1 lineage appears to have a slower mutation rate with an average of 1.7 SNPs/year (green in Figure 3.). The S2 population was excluded from analysis due to the two samples being taken only a day apart, and the disappearance of this lineage. This data suggests that within a given patient, the rate of mutation accumulation can vary over the same study period between diverged sub-populations. This also indicates that extensive selective pressure likely does not impact the overall mutation rate of *B. multivorans*, hinting that after initial diversification the population remains relatively stable.

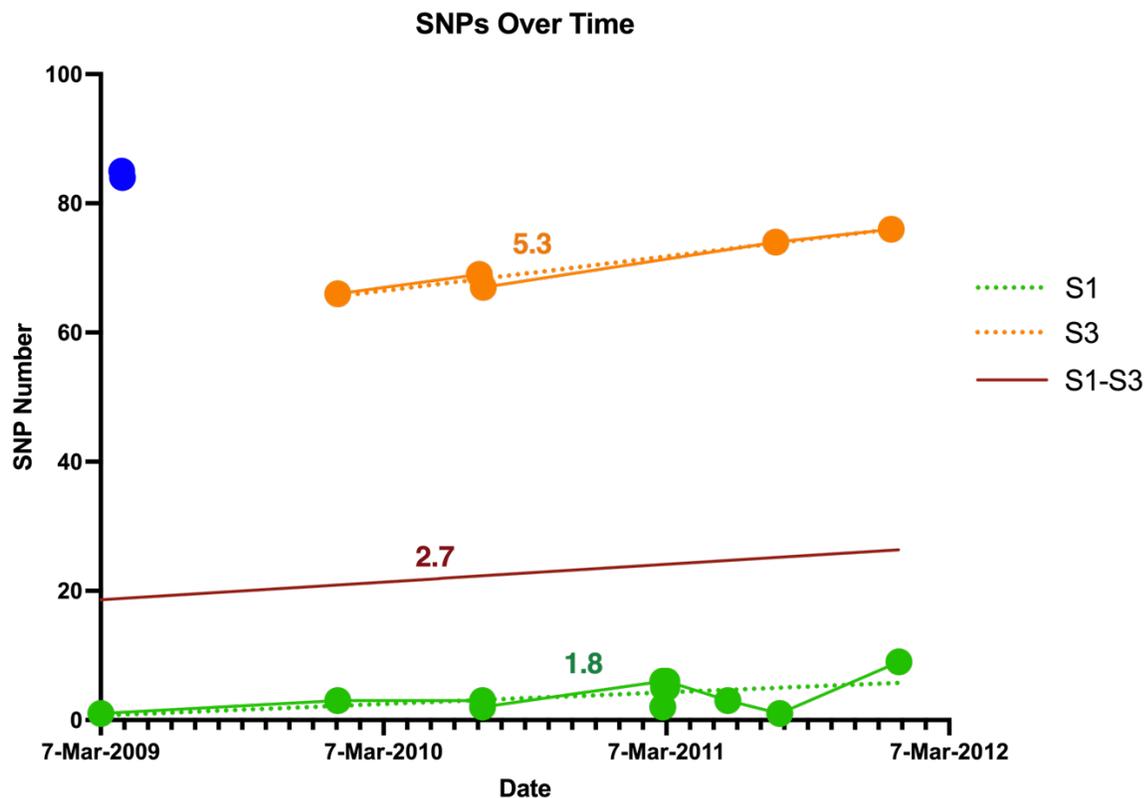


Figure 3. SNPs plotted over time of sample isolation. Colors denoted by phylogenetic subpopulation. Green = S1, Blue= S2, and Orange=S1. Sample SNP numbers. A linear fit with a slope was plotted over time (excluding S2), as well as for each sub-lineage.

3.4: Frequency and Functional Characterization of SNPs and INDELS

All SNPs and indels in the AS samples were identified (see methods). A total of 188 SNPs and 40 indels were discovered. Samples had a range of SNPs (1-85) and INDELS (8-17) which corresponded to which subpopulation the samples belonged. The S2 population (AS218, AS219) contained the highest number of mutations on average (96). Isolates of S3 had an average of 82.2 and those within the S1 group had the lowest with an average of 18.6. Of the 188 SNPs, 108 (57.4%), 29 (20.7%), and 41 (21.8%) were non-synonymous, synonymous, and intergenic polymorphisms. There were 12 intergenic mutations located in putative regulatory regions (100bp ahead of a start codon).

To describe the variety of mutations at any given time point, analyzed mutations shared and unique to each of the defined subgroups. A significant proportion (42%) of SNPs were singletons (only in one sample). There were also 29 mutations shared between all isolates of the S2 and S3 subpopulations, while those in the S1 did not share any (Supplementary Figure 2). Sought to identify SNPs in any genes fixed in all isolates over time, polymorphisms that arise in the population and remain fixed are candidates for those maintained by selection (fixed polymorphisms). However, it was found that no mutations arose and were fixed across all isolates.

COG analysis of all genes showed that the majority of genes were associated with membrane (15.88%), unknown (13.53%), transcription (13.53%), and amino acid transport/metabolism (12.94%) (Figure 5.). Only samples from S2 and S3 contained mutations in H, and L COG categories (Figure 4.). To further investigate the mutation accumulation differences between the lineages (see Section 3.3), sought to analyze mutated genes involved with replication, recombination, and repair (L), and cell cycle control and division (D) as defective repair or replication mechanisms could help to explain the increased mutation rate in the S2, and S3 sub-populations in comparison to S1. The only mutation classified as D that was shared among both subpopulations was in the A8H40_RS22860 locus, which encodes a peptidoglycan-binding protein LysM. No mutations involving DNA repair were shared among all isolates of S2 and S3, however, isolates in the S3 population were shown to all contain a mutation in *recB* which prepares dsDNA breaks for recombinational DNA repair. Also, worth noting is that both isolates in the S2 population share a mutation in a DNA repair exonuclease locus A8H40_RS29745.

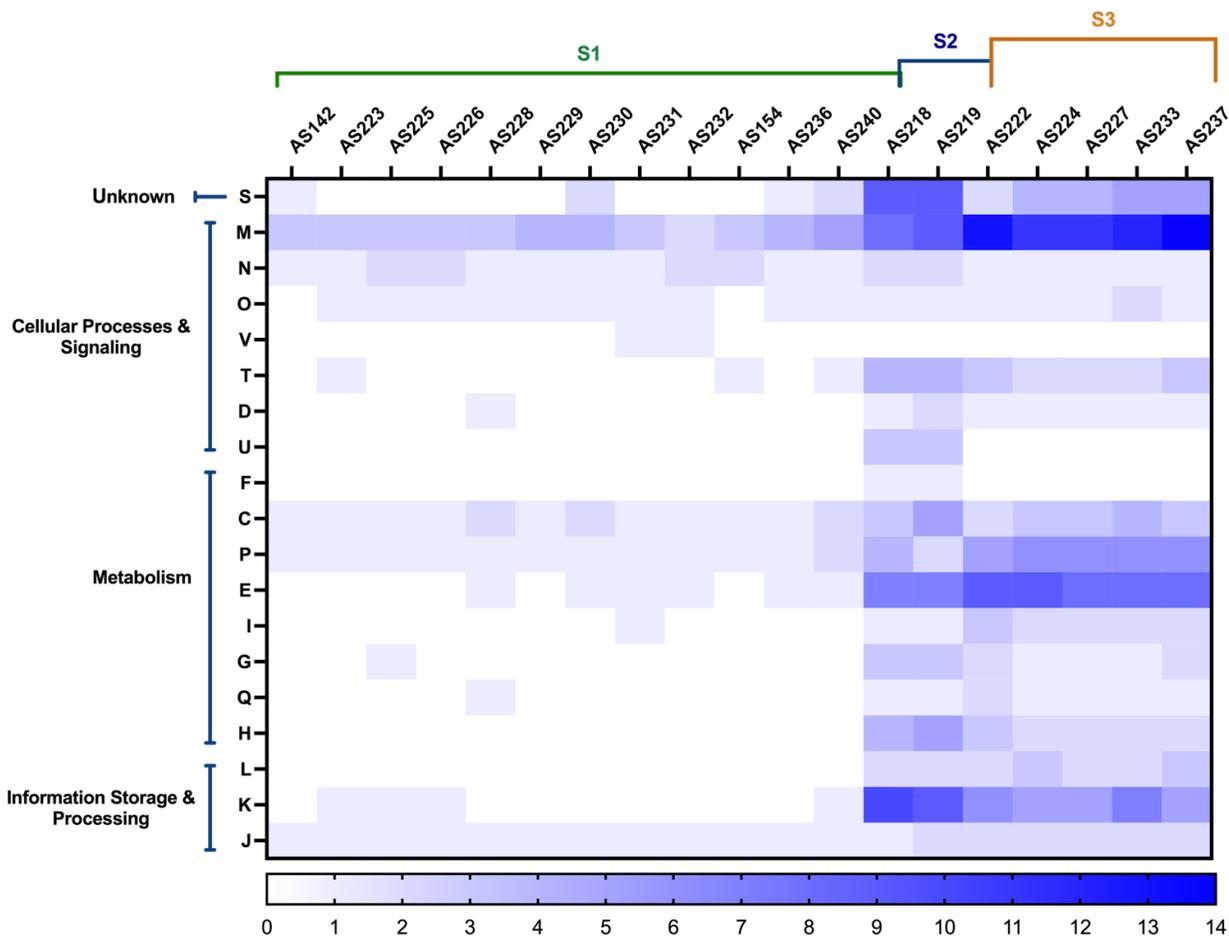


Figure 4. Breakdown of COG categories of mutated proteins. L: DNA replication, recombination, and repair; E: metabolism and transport of amino acids; H: metabolism and transport of coenzymes; G: metabolism and transport of carbohydrates; C: production and conversion of energy; J: transcription; I: lipidic metabolism; M: Cell wall structure, biogenesis, and outer membrane; O: posttranslational modifications; U: Intracellular trafficking, secretion, and vesicular transport; T: mechanisms of signal transduction; Q: metabolism and transport of nucleotides; P: transport of inorganic ions; N: cell motility; K: translation including ribosomes biogenesis; F: metabolism and transport of nucleotides; R: general prediction function only; S: function unknown; V: Defense mechanisms; D: Cell cycle control and replication; N: Cell motility

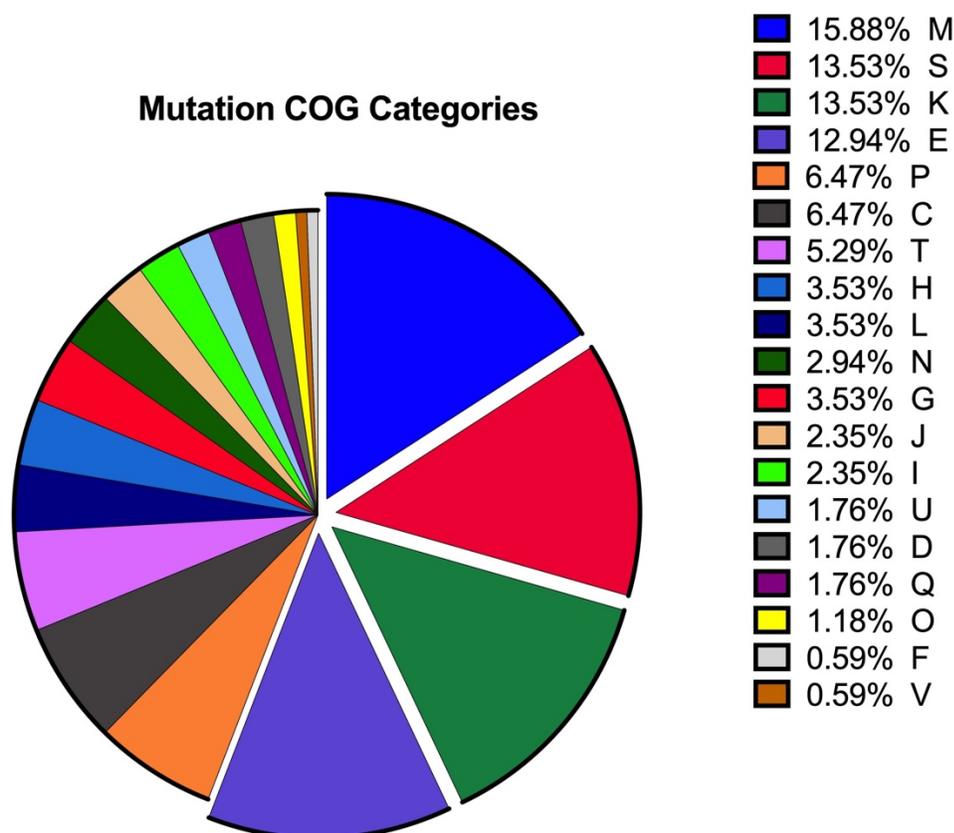


Figure 5. The number of mutations in each sample that product falls into the designated COG category. Breakdown of COG categories of mutated proteins. L: DNA replication, recombination, and repair; E: metabolism and transport of amino acids; H: metabolism and transport of coenzymes; G: metabolism and transport of carbohydrates; C: production and conversion of energy; J: transcription; I: lipidic metabolism; M: Cell wall structure, biogenesis, and outer membrane; O: posttranslational modifications; U: Intracellular trafficking, secretion, and vesicular transport; T: mechanisms of signal transduction; Q: metabolism and transport of nucleotides; P: transport of inorganic ions; N: cell motility; K: translation including ribosomes biogenesis; F: metabolism and transport of nucleotides; R: general prediction function only; S: function unknown; V: Defense mechanisms; D: Cell cycle control and replication; N: Cell motility

3.5: Genes with Multiple Independent Mutations

To determine the probability of observing a given number of mutations based on gene size, I used a two proportion Z-test in R with the `prop.test` command correcting for multiple tests using the Bonferroni correction. Numerous mutations in genes can indicate selective

enhancement of changes driven by adaptation to host pressures and other environmental variables. Mutations in genes that are involved in maintaining or increasing bacterial fitness should be present at a non-random frequency. There were 12 genes identified that contained mutations at a frequency higher than chance (Table 2.). No specific mutation was found to occur more than once. Two of the 12 multi-mutated genes have previously been shown to be involved with antibiotic resistance. One of the two, the gene A8H40_RS14655 encoding a LysR family transcription regulator, had 4 mutations ($p < .0001$) which were all nonsynonymous SNPs in four different isolates but none of the four mutations were found in the later isolates. This gene has shown to be a close homolog (99% identity) of BMUL_0631 which was found to contain 7 SNPs in another long-term study of *B. multivorans* in the CF lung, indicating the possible importance of survival during antibiotic treatment (Diaz Caballero et al., 2018). The second multi-mutated gene associated with antibiotic resistance is *dacB*, which was independently mutated four times in a total of 12 isolates; two were nonsynonymous SNPs, a single one bp deletion, and a 161 bp deletion. The *dacB* gene encodes a non-essential penicillin-binding protein (PBP4) and has previously been reported to be involved in the expression regulation of β -lactamases in *P. aeruginosa* (Moya et al., 2009). Another gene with multiple mutations worth noting is *fabF*, which contains two nonsynonymous SNPs and encodes a beta-ketoacyl-ACP synthase II involved in fatty acid metabolism. A study on the in-host CF evolution of *B. pseudomallei* found multiple convergent nonsynonymous mutations affecting the FabF protein as well as fatty acid elongation proteins (Viberg et al., 2017). Chronically infecting *B. cenocepacia* isolates have also been seen to contain a larger ratio of unsaturated fatty acids when compared to initial infecting isolates (Coutinho et al., 2011). A majority of SNPs in multi-mutated genes were found to be nonsynonymous (92%), which further provides evidence of strong selection at these 12 loci. In

summary, genes associated with cell wall/membrane biogenesis appeared to be under selection in multiple isolates at the highest frequency, as well as genes involving metabolic functions and transport (ions, lipids, and amino acids).

Table 2: Genes with Multiple Independent Mutations.

Table displays all genes that had a statistically significant number of mutations. Probabilities were Bonferroni corrected for multiple comparisons. A total of 12 genes were found to contain a significant number based on size of gene and number of mutations.

Gene/Locus	Product	Total #	SNPs,Indels,CNV	Probability	COG
fabF	beta-ketoacyl-ACP synthase II	2	2,0,0	2.80E-06	I
A8H40_RS09725	hypothetical protein	4	1,2,1	2.90E-77	M
yafL/A8H40_RS11145	C40 family peptidase	3	3,0,0	6.20E-43	M
dacB	Beta-lactamase/transpeptidase-like	4	2,1,1	5.20E-36	M
A8H40_RS14655	LysR family transcriptional regulator	4	4,0,0	5.20E-59	K
A8H40_RS15235/kefB	potassium transporter	5	4,1,0	9.40E-46	P
secB	protein-export chaperone SecB	2	2,0,0	1.40E-20	U
plpD/A8H40_RS21210	BamA/TamA family outer membrane protein	3	2,1,0	5.60E-10	M
A8H40_RS24455	zinc-binding alcohol dehydrogenase protein	2	2,0,0	7.70E-09	C
A8H40_RS24565	hypothetical protein	2	2,0,0	6.30E-59	S
A8H40_RS05775	DUF4136 domain-containing protein	2	0,2,0	3.40E-13	S
pheA	Prephenate dehydratase	4	0,4,0	2.00E-52	E

3.6 Frequency of Structural Variation

Structural variations (SVs) are becoming increasingly recognized to represent a significant, yet often poorly understood, source of genetic variation. Structural variation is used to define a region of DNA that shows a change in copy number (deletions, insertions, and duplications), or orientation (like inversions), as well as chromosomal location (such as translocations, and fusions) between individuals. However, most current methods are poor at defining the breakpoints for SV detection which are sequence boundaries where an SV begins and ends, and algorithms often result in a lot of false-positive calls. In an attempt to overcome the performance limitations of existing SV-calling methods, I used multiple complementary algorithms to call SV loci then merged them, only keeping calls in which at least 4 out of the 6 used callers had detected a variant at that region.

In order to validate each call, pair orientation and insert size were manually inspected for each region of all sample alignments within the integrative genome visualizer (IGV), to confirm true positives. A total of nine SVs were identified (Table 3.), eight were deletions, and one an inversion (INV). There were three mutations of >35,000 bp in length, two of which were deletions (36,898 bp, and 60,026 bp) found in only AS240. Two of the other eight deletions fall within 12,000bp of one another, the first (1968944-1968970) a 27bp deletion in the *rimO* gene which encodes the methyl-thioltransferase RimO, the other a 122bp deletion (1980490-1980612) in the A8H40_RS09725 locus encoding a hypothetical protein. AS225 was found to contain a unique 32bp deletion in *ptsP* gene that encodes for a phosphotransferase. Towards the end of chromosome one, AS142, AS223, AS225, AS226, and AS154 all share a 183bp deletion in A8H40_RS16235 which also encodes a hypothetical protein. Other than the large deletion in AS240 mentioned above, there was one other SV identified on chromosome 2 which is a 955bp

deletion in only AS218 and AS219 that falls within an outer membrane encoding gene (A8H40_RS23195). Unexpectedly, no SVs were identified on the third chromosome.

Table 3. Table of Detected and Verified SVs

The table shows each region (chromosome, start position, stop position) for which a structural variation was called and verified. The IGV image of each region by sample is displayed to indicate the true presence of the variant. The length and type of SV are also given. The color of the lines represents sampling period.

CHROM	START	STOP	142	218	219	222	223	224	225	226	227	228	229	230	231	232	154	233	236	237	240	LEN	TYPE
NZ_CP020397.1	48863	86761																				37898	DEL
NZ_CP020397.1	1505982	1548069																				42087	INV
NZ_CP020397.1	1980490	1980612																				122	DEL
NZ_CP020397.1	2929122	2929282																				160	DEL
NZ_CP020397.1	3161880	3161912																				32	DEL
NZ_CP020397.1	3299493	3299676																				183	DEL
NZ_CP020398.1	1066240	1126266																				60026	DEL
NZ_CP020398.1	1385790	1386745																				955	DEL

3.7: Antibiotic Susceptibility

Antibiotic susceptibility testing (AST) via a disc diffusion assay was performed for all isolates to analyze the variability of resistance profiles between subpopulations, as well as, interpret changes over time resultant of extensive antibiotic treatment. A total of eight antibiotics were chosen for testing ceftazidime (CAZ), ceftazidime/avibactam (CZA), meropenem (MEM),

meropenem/vaborbactam (MEV), piperacillin (PIP), ciprofloxacin (CIP), trimethoprim/sulfamethoxazole (SXT), and minocycline (MH). The ZOI for each antibiotic were plotted over time by the date of isolation from each sample. It was seen that over the course of three years there was a general trend downward for CAZ and PIP (Figure 6.A, and 6.E). Interestingly all found a general trend upwards (more susceptible) for SXT over time, regardless of lineage splitting (Figure 6.G).

Assuming that extensive antibiotic exposure during the infection sampling period would result in strong selection for resistance-associated genotypes, sought to identify mutations that arose following treatment with CAZ during a 2-week hospitalization in January 2010 (shown as the green line in Figure. 6). It was found that the *dacB* gene appears to be an important selective target since it was independently mutated a total of four times within the collection across all sub-populations. No samples isolated prior to the January hospitalization were seen to have any mutations in this gene, suggesting that selective pressure imposed by IV administration of CAZ could have selected for an advantageous phenotype (increased β -lactam resistance) associated with mutations in this gene in the total of 12 samples that contained a mutation in this gene. Samples with any of the four were thought to have a lower ZOI (more resistant) to CAZ (Figure 7. A and B). However, after statistical testing of ZOIs grouped by mutations only found two mutations significantly different than the *dacB*⁺ (no mutation). The first mutation was the 161bp deletion (shown in green) and the substitution of C>A at position 820 (shown in pink).

One such possible explanation for the increased resistance to β -lactam is that this gene has been seen to be involved in the expression of AmpC, β -lactamases, and deactivation can result in hyper-production of AmpC, and in some cases when inactivated significantly increases minimum inhibitory concentration (MIC) to cephalosporin antibiotics in *P. aeruginosa* (Moyá et al., 2012).

It has also been found to play a role in the activation of the two-component regulatory system BlrAB (CreBC) which has been associated with resistance (Fisher & Mobashery, 2014; Moyá et al., 2012). It can be seen in Figure 7. B that no mutation appeared to have any impact on resistance to ceftazidime-avibactam (CZA). This shows that deactivation of β -lactamases restores the function of CAZ and overexpression of β -lactamases would not influence resistance. Interestingly, the only two samples following CAZ treatment that did not contain a mutation in the *dacB* gene were samples AS230 and AS231. These two isolates are more susceptible to CAZ than those containing a mutation in *dacB* (shown in purple in Figure 7. A), however, does show to have a large reduction in PIP susceptibility although was not found to be statistically significant (Figure 7.B). Both of these samples are shown to contain a nonsynonymous SNP in the *ampD* gene. Mutations in the *ampD* have been found to result in the induction of AmpC and PenB in *B. cenocepacia*, and also to be statistically associated with β -lactam resistance in a longitudinal study of *B. multivorans* (Diaz Caballero et al., 2018; Hwang & Kim, 2015). Together these results suggest positive selection for mutations in genes that regulate β -lactamases following administration of CAZ, where mutations that deactivate to a higher degree such as a large deletion have a larger impact on resistance.

Interestingly, AS218, AS219, and AS223 all show resistance to SXT and ciprofloxacin (CIP). There is only one gene in which AS223 shares a mutation with AS218 and AS219 which is A8H40_RS14655 which is predicted to encode a LysR family transcription regulator. This gene was independently mutated four times in samples AS218, AS219, AS222, and AS223 but none of the mutations appeared and in samples isolated downstream. All samples following the hospitalization and increased CAZ and PIP resistance show to be more susceptible to SXT regardless of which sub-population the sample belongs to.

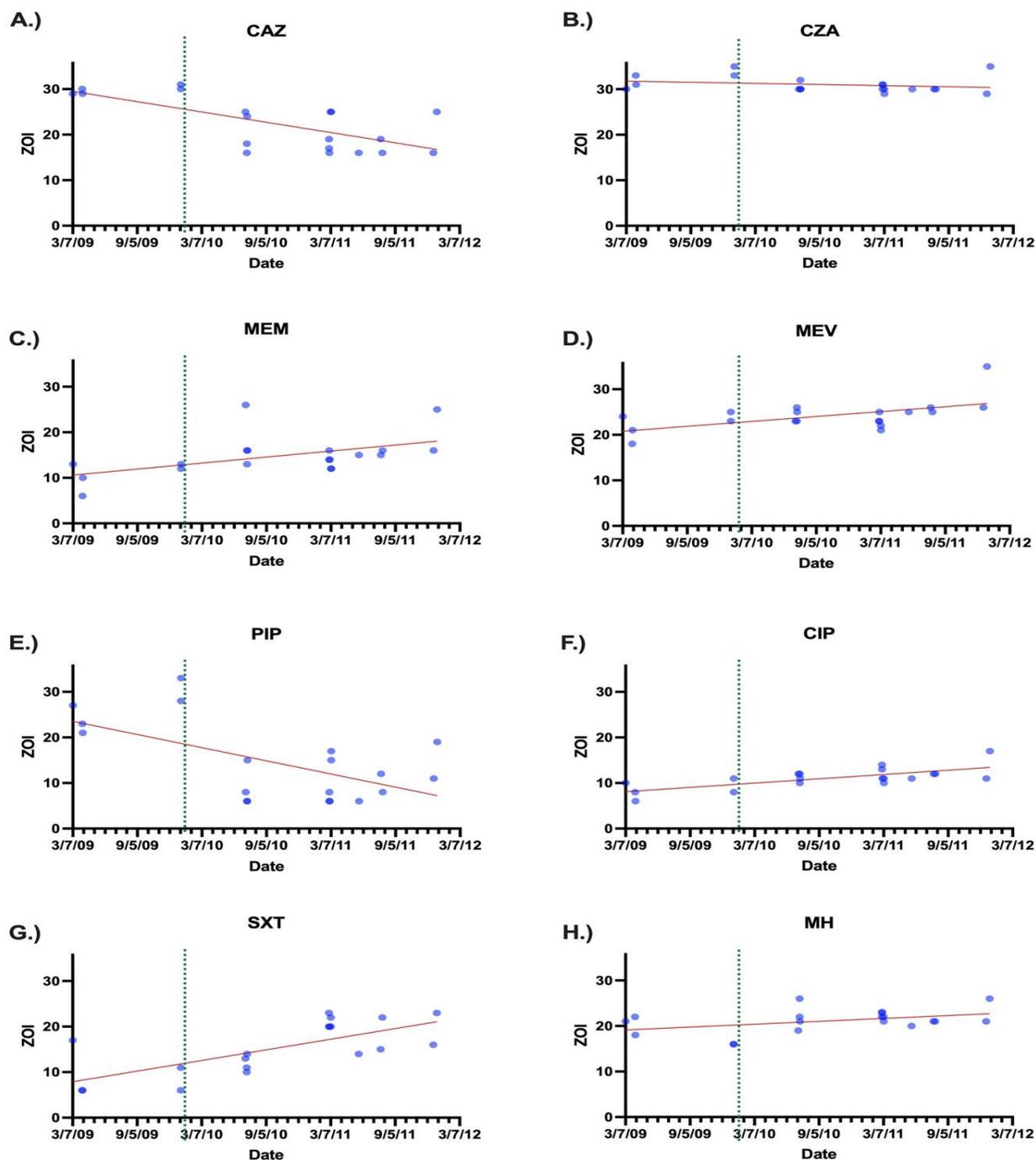


Figure 6. Antibiotic Susceptibility overtime for A. Ceftazidime (CAZ) B. Ceftazidime/Avibactam C. Meropenem (MEM) D. Meropenem/Vaborbactem E. Piperacillin (PIP) E. Ciprofloxacin (CIP) F. Trimethoprim / Sulfamethoxazole (SXT) G. Minocycline (MH)

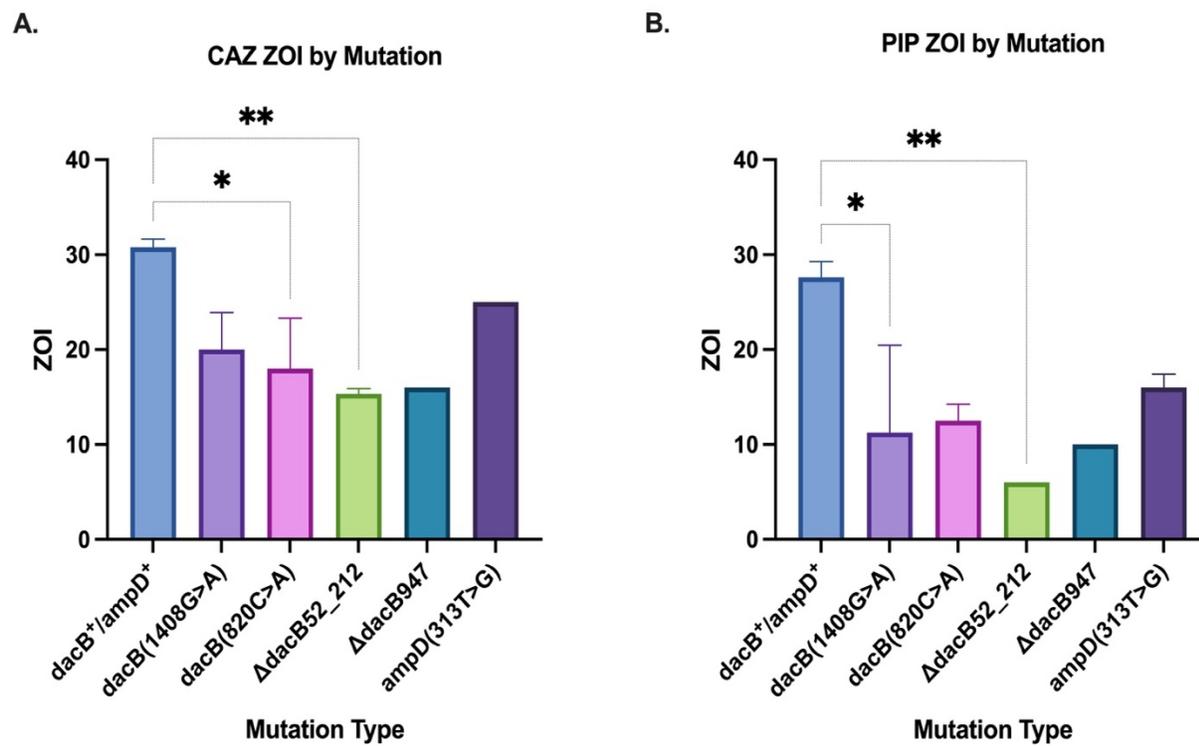


Figure 7. A) Ceftazidime (CAZ) ZOI by mutation B) Piperacillin (PIP) ZOI by mutation. The asterisks (*) indicate $P < .05$ for multiple comparison tests using Kruskal-Wallis and Dunn's Correction.

CHAPTER 5: DISCUSSION

Eradication of *Burkholderia* infections is difficult because members of the Bcc generally have a high level of intrinsic and acquired resistance mechanisms to commonly used antibiotics in clinical practice. Inside of the CF lung, bacteria face a changing and stressful environment. The heterogeneous ecosystem of the CF lung generates ecological micro-niches with differing characteristics, thus forming variable selective forces (Palmer et al., 2007). Divergent evolutionary patterns of colonially related isolates of Bcc bacteria have not been investigated to the extent as other more prevalent CF pathogens such as *P. aeruginosa*. Previous studies have found the prevalence of each lineage within a patient to be highly dynamic over the course of infection, which affected diversification processes of *P. aeruginosa* considerably (Bernardy et al., 2020; Marvig et al., 2015; Schick & Kassen, 2018). Within-host micro-evolution of diversified lineages employed by *B. multivorans* in the CF lung is not well understood, while it has important implications for long-term studies. This also shows possible limitations of sequencing only a single strain per time point as this may only provide a fraction of possible evolutionary avenues undertaken within the bacterial population. In the present study, it is clear that the population was diverse with two primary sub-populations coexisting within the same sputum samples. This study supported the presence of co-existing diversified lineages that have a considerable difference in the number and rate of mutations (Figure 2 and).

A subset of genes was found to have multiple mutations at independent positions, which could indicate loci under selection. Sought to compare the multi-mutated list of genes with similar studies, as overlaps would suggest the importance of these loci in the persistence of *B. multivorans* in the CF lung. Correlation of gene lists with Caballero et al (2018) showed only two genes in common, the first within a LysR transcriptional regulator (A8H40_RS14655), and

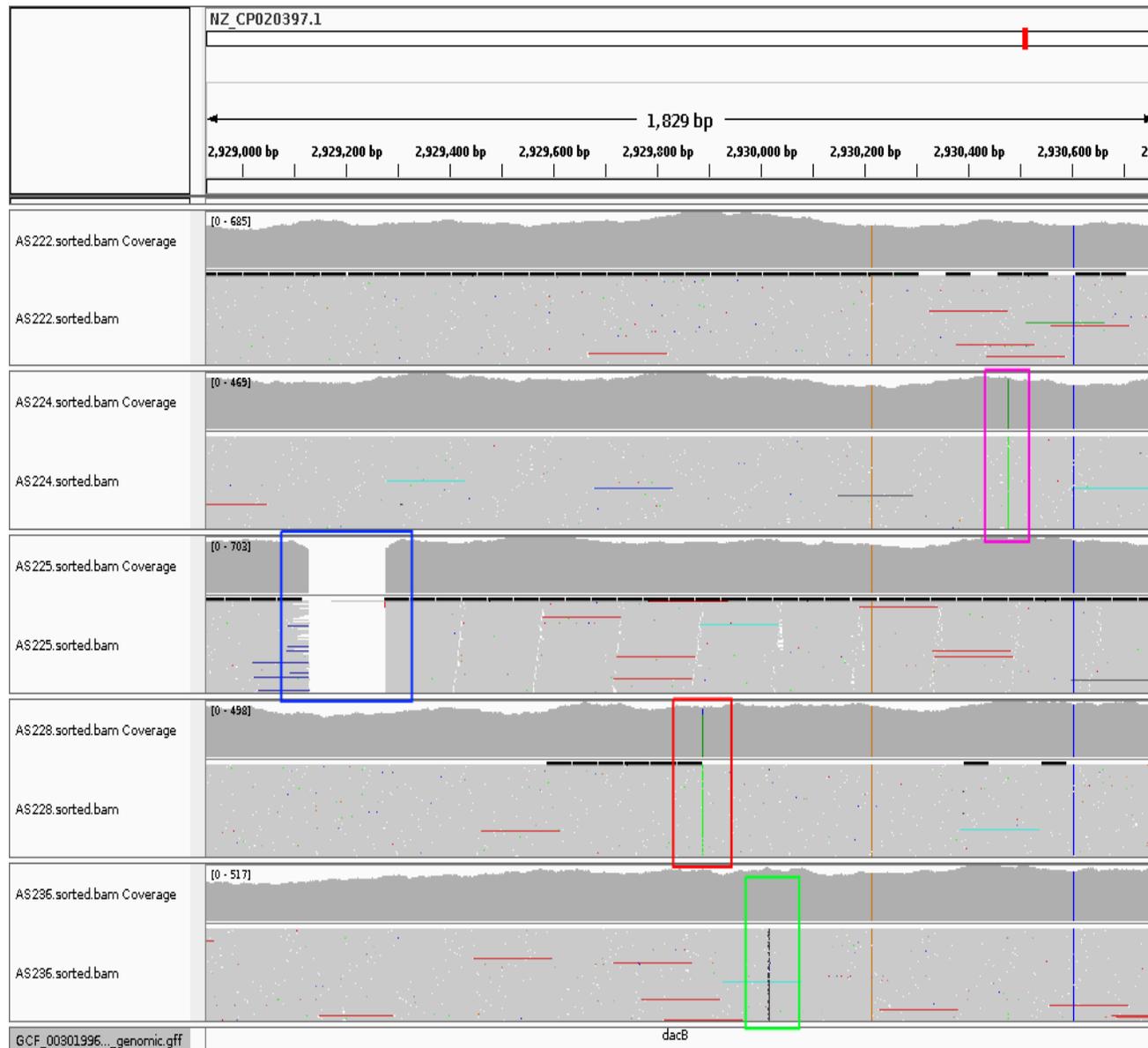
secB. We ran the Silva et al (2016) isolate data through our variant calling pipelines to directly compare all variants. Two genes were found to be mutated multiple times in both data sets which encoded a HAMP domain-containing histidine kinase (A8H40_RS29590) and a palatatin-like phospholipase family protein (A8H40_RS21210), however, statistical analysis did not find two mutations in either gene to be significant. Interestingly, one region within the SV analysis did overlap between data sets one large region on the first chromosome which contained 31 one genes, this appeared as a large deletion in AS240 of 37,898bp, and a duplication of 38,489bp in the same location which was called in four samples within the Silva data (BM4, BM6, BM7, and BM8). No overlaps between multi-mutated genes were found within Lood et. al (2021), or Hassan et. al (2020). Other genes with multiple independent mutations which were unique to our data set were in *fabF*, *pheA*, *dacB*, a potassium transporter, a zinc-binding alcohol dehydrogenase, a C40 family peptidase, and a *plpD* ortholog encoding a BamA/TamA outer membrane protein.

The overall genomic stability of *B. multivorans* in response to strong selective pressure (i.e antibiotic therapy) is not well documented. Here it is demonstrated that within lineages *B. multivorans* accumulated on average 2.7 SNPs/year when S2 removed from the analysis, this is comparable for what has been found previously for *Burkholderia* during chronic infection, for *B. multivorans* (2.2–2.4 SNPs/year), *B. cenocepacia* (1.7–2.1 SNPs/year) and *B. dolosa* (2.1 SNPs/year) (Diaz Caballero et al., 2018; Lieberman et al., 2011; Morarty et al., 2007). When analyzing SNP accumulation separately for each sub-population found the S3 population is mutating at almost double the rate of the S1 (5.3 vs 2.7 SNPs/year).

The CF environment presents stressful and variable conditions, where multiple mutations may occur that increase fitness. *B. multivorans* adaptation to the CF lung could be established

through the selection of combinations of the genetic mutations that arise which leads to diversification of the initial infection into sub-populations over time. The present study illustrates that once diverse lineages are established, they are relatively stable and extensive antibiotic treatment did not appear to impact the overall mutation rate of the population. However, there was evidence detected of more mutable sub-populations (S2/S3) which contained multiple mutations in DNA repair mechanisms which could be contributing to the dramatic difference in cumulative mutations between lineages. It was seen that IV administration of ceftazidime could have resulted in a rapid expansion of β -lactam resistance in the population and potentially indicating the importance of *dacB* and *ampD* in acquired β -lactam resistance in *B. multivorans*.

SUPPLEMENTAL INFORMATION:

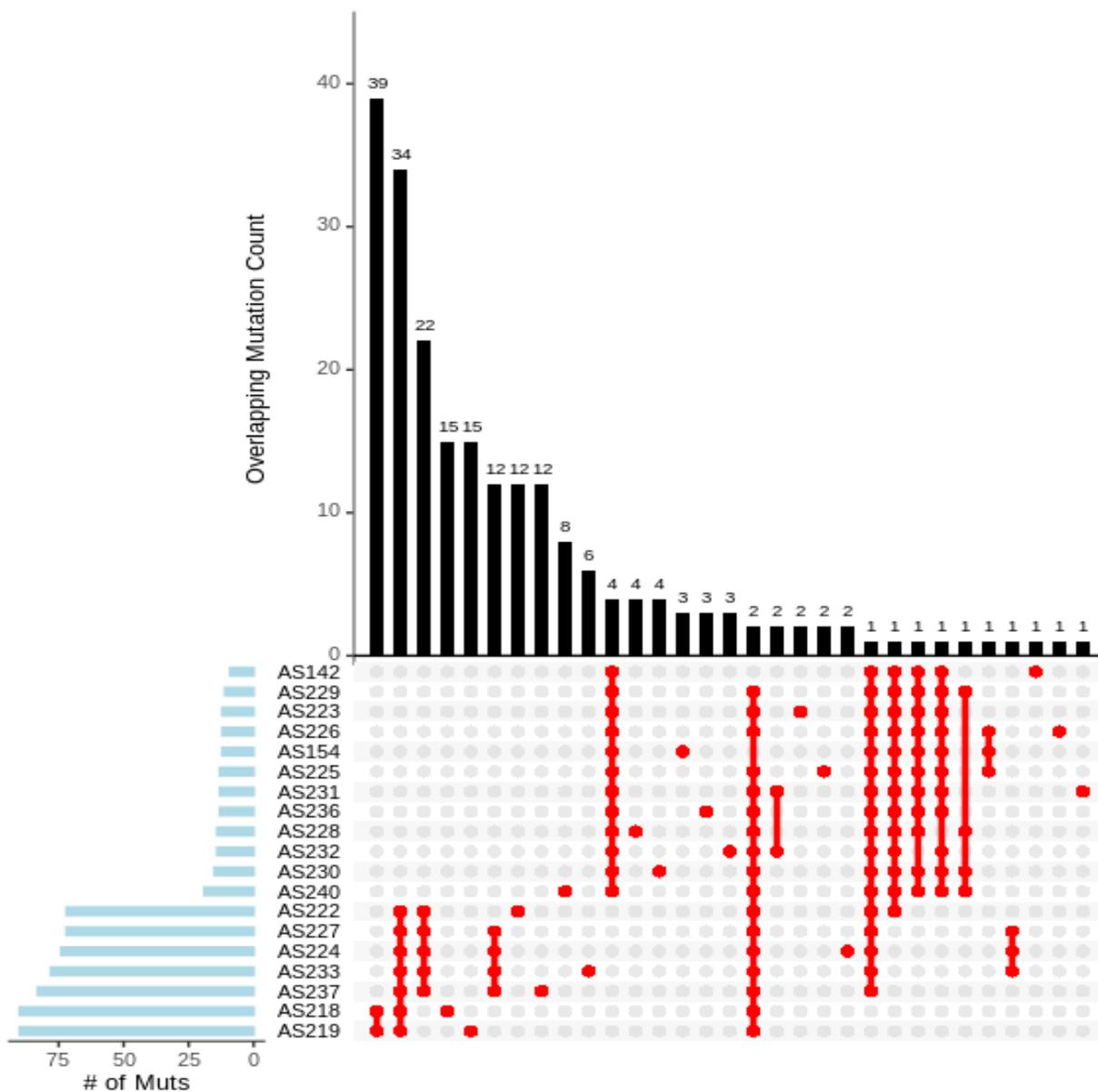


Supplemental Figure 1. IGV image of *dacB* mutations. Shows IGV image of *dacB* of AS222, AS224, AS225, AS228, and AS236 to display each type of mutation in *dacB*. The top shows no mutation in AS222 that was not shared among all samples, the pink and red boxes are showing the two nonsynonymous SNPs, the blue box is a 161bp deletion, and the green box a single bp deletion.

Supplemental Table 1. Average coverage/sample stats:

The table displays summary statistics by sample. Shows the average read coverage, average frequency of SNP and INDEL calls, and the number of SNP and Indels that each contained.

Sample	Avg Coverage	Avg SNP Freq	Avg INDEL Freq	SNPs	INDELs
AS142	51.558	0.938	0.920	1	8
AS218	343.868	0.943	0.739	85	12
AS219	279.830	0.900	0.697	84	11
AS222	364.674	0.884	0.762	66	12
AS223	520.953	0.938	0.799	3	17
AS224	287.186	0.922	0.807	69	11
AS225	442.956	0.941	0.789	3	17
AS226	445.957	0.931	0.792	2	16
AS227	311.476	0.848	0.754	67	11
AS228	260.896	0.845	0.741	6	14
AS229	393.978	0.942	0.806	2	15
AS230	377.383	0.940	0.820	5	17
AS231	267.596	0.934	0.793	5	15
AS232	318.375	0.946	0.780	6	14
AS154	48.500	0.976	0.859	3	9
AS233	270.5106	0.936	0.818	74	13
AS236	292.1818	0.911	0.776	1	17
AS237	291.420	0.883	0.806	76	12
AS240	339.280	0.975	0.797	9	15



Supplemental Figure 2. SNP Distribution/overlaps. The figure shows overlaps between sample SNP numbers to visually display the difference in SNP numbers between the lineages. Can see that there are 23 SNPs shared just between AS218/AS219 (S2), 20 shared among all isolates of S2/S3, and 12 among those in the S3. All other samples contain very few overlaps with those in S2 and S3, and none overlap between all in S1.

REFERENCES:

1. CLSI. Performance Standards for Antimicrobial Susceptibility Testing. 30th ed. CLSI supplement M100. Wayne, PA: Clinical and Laboratory Standards Institute; 2020.
2. *Cystic Fibrosis Foundation Patient Registry 2019 Annual Data Report*. (n.d.). <https://www.cff.org/Research/Researcher-Resources/Patient-Registry/2019-Patient-Registry-Annual-Data-Report.pdf>
3. Alexander, B. M., Petren, E. K., Rizvi, S., Fink, A., Ostrenga, J., Sewall, A., & Loeffler, D. (2016). *Cystic Fibrosis Foundation Patient Registry*. Annual Data Report. <https://www.cff.org/Our-Research/CF-Patient-Registry/2015-Patient-Registry-Annual-Data-Report.pdf>
4. Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data*. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
5. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*(5), 455–477. <https://doi.org/10.1089/cmb.2012.0021>
6. Bernardy, E. E., Petit, R. A., Raghuram, V., Alexander, A. M., Read, T. D., & Goldberg, J. B. (2020). Genotypic and Phenotypic Diversity of *Staphylococcus aureus* Isolates from Cystic Fibrosis Patient Lung Infections and Their Interactions with *MBio*, *11*(3), e00735-20. <https://doi.org/10.1128/mBio.00735-20>
7. Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
8. Cameron, D. L., Schröder, J., Penington, J. S., Do, H., Molania, R., Dobrovic, A., Speed, T. P., & Papenfuss, A. T. (2017). GRIDSS: Sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly. *Genome Research*, *27*(12), 2050–2060. <https://doi.org/10.1101/gr.222109.117>
9. Cameron, D., & Dong, R. (2020). StructuralVariantAnnotation: Variant annotations for structural variants. In *R package version 1.6.0* (R package version 1.6.0.).
10. Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., Cox, A. J., Kruglyak, S., & Saunders, C. T. (2016). Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*, *32*(8), 1220–1222. <https://doi.org/10.1093/bioinformatics/btv710>
11. Coutinho, C. P., de Carvalho, C. C. R., Madeira, A., Pinto-de-Oliveira, A., & Sá-Correia, I. (2011). *Burkholderia cenocepacia*

- Phenotypic Clonal Variation during a 3.5-Year Colonization in the Lungs of a Cystic Fibrosis Patient. *Infection and Immunity*, 79(7), 2950 LP – 2960.
<https://doi.org/10.1128/IAI.01366-10>
12. Dennis, J. J. (2017). Burkholderia cenocepacia virulence microevolution in the CF lung: Variations on a theme. *Virulence*, 8(6), 618–620.
<https://doi.org/10.1080/21505594.2016.1253660>
 13. Diaz Caballero, J., Clark, S. T., Wang, P. W., Donaldson, S. L., Coburn, B., Tullis, D. E., Yau, Y. C. W., Waters, V. J., Hwang, D. M., & Guttman, D. S. (2018). A genome-wide association analysis reveals a potential role for recombination in the evolution of antimicrobial resistance in Burkholderia multivorans. *PLoS Pathogens*, 14(12).
<https://doi.org/10.1371/journal.ppat.1007453>
 14. Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792–1797.
<https://doi.org/10.1093/nar/gkh340>
 15. Fan, X., Abbott, T. E., Larson, D., & Chen, K. (2014). BreakDancer: Identification of genomic structural variation from paired-end read mapping. *Current Protocols in Bioinformatics*, SUPPL.45. <https://doi.org/10.1002/0471250953.bi1506s45>
 16. Fisher, J. F., & Mobashery, S. (2014). The sentinel role of peptidoglycan recycling in the β -lactam resistance of the Gram-negative Enterobacteriaceae and Pseudomonas aeruginosa. *Bioorganic Chemistry*, 56, 41–48.
<https://doi.org/https://doi.org/10.1016/j.bioorg.2014.05.011>
 17. Hassan, A. A., dos Santos, S. C., Cooper, V. S., & Sá-Correia, I. (2020). Comparative Evolutionary Patterns of Burkholderia cenocepacia and B. multivorans During Chronic Co-infection of a Cystic Fibrosis Patient Lung. *Frontiers in Microbiology*, 11.
<https://doi.org/10.3389/fmicb.2020.574626>
 18. Hwang, J., & Kim, H. S. (2015). Cell Wall Recycling-Linked Coregulation of AmpC and PenB β -Lactamases through ampD Mutations in Burkholderia cenocepacia. *Antimicrobial Agents and Chemotherapy*, 59(12), 7602–7610.
<https://doi.org/10.1128/AAC.01068-15>
 19. Jones, A. M., Dodd, M. E., Govan, J. R. W., Barcus, V., Doherty, C. J., Morris, J., & Webb, A. K. (2004). Burkholderia cenocepacia and Burkholderia multivorans: influence on survival in cystic fibrosis. *Thorax*, 59(11), 948–951.
<https://doi.org/10.1136/thx.2003.017210>
 20. Kenna, D. T. D., Lilley, D., Coward, A., Martin, K., Perry, C., Pike, R., Hill, R., & Turton, J. F. (2017). Prevalence of burkholderia species, including members of burkholderia cepacia complex, among UK cystic and non-cystic fibrosis patients. *Journal of Medical Microbiology*, 66(4), 490–501. <https://doi.org/10.1099/jmm.0.000458>
 21. Klockgether, J., Cramer, N., Fischer, S., Wiehlmann, L., & Tümmler, B. (2018). Long-Term Microevolution of Pseudomonas aeruginosa Differs between Mildly and Severely Affected Cystic Fibrosis Lungs. *American Journal of Respiratory Cell and Molecular Biology*, 59(2), 246–256. <https://doi.org/10.1165/rcmb.2017-0356OC>
 22. Lamrabet, O., Martin, M., Lenski, R. E., & Schneider, D. (2019). Changes in intrinsic antibiotic susceptibility during a long-term evolution experiment with escherichia coli. *MBio*, 10(2). <https://doi.org/10.1128/mBio.00189-19>
 23. Layer, R. M., Chiang, C., Quinlan, A. R., & Hall, I. M. (2014). LUMPY: A probabilistic framework for structural variant discovery. *Genome Biology*, 15(6).

- <https://doi.org/10.1186/gb-2014-15-6-r84>
24. Lefort, V., Desper, R., & Gascuel, O. (2015). FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. *Molecular Biology and Evolution*, 32(10), 2798–2800. <https://doi.org/10.1093/molbev/msv150>
 25. Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
 26. Lieberman, T. D., Michel, J.-B., Aingaran, M., Potter-Bynoe, G., Roux, D., Davis, M. R., Skurnik, D., Leiby, N., LiPuma, J. J., Goldberg, J. B., McAdam, A. J., Priebe, G. P., & Kishony, R. (2011). Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nature Genetics*, 43(12), 1275–1280. <https://doi.org/10.1038/ng.997>
 27. Mahenthiralingam, E., Baldwin, A., & Dowson, C. G. (2008). Burkholderia cepacia complex bacteria: Opportunistic pathogens with important natural biology. In *Journal of Applied Microbiology* (Vol. 104, Issue 6, pp. 1539–1551). <https://doi.org/10.1111/j.1365-2672.2007.03706.x>
 28. Marvig, R. L., Dolce, D., Sommer, L. M., Petersen, B., Ciofu, O., Campana, S., Molin, S., Taccetti, G., & Johansen, H. K. (2015). Within-host microevolution of Pseudomonas aeruginosa in Italian cystic fibrosis patients. *BMC Microbiology*, 15(1), 218. <https://doi.org/10.1186/s12866-015-0563-9>
 29. Meier-Kolthoff, J. P., & Göker, M. (2019). TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nature Communications*, 10(1). <https://doi.org/10.1038/s41467-019-10210-3>
 30. Miller, R. R., Hird, T. J., Tang, P., & Zlosnik, J. E. A. (2015). Whole-Genome Sequencing of Three Clonal Clinical Isolates of B. cenocepacia from a Patient with Cystic Fibrosis. *PLOS ONE*, 10(11), e0143472. <https://doi.org/10.1371/journal.pone.0143472>
 31. Moriarty, T. F., McElnay, J. C., Elborn, J. S., & Tunney, M. M. (2007). Sputum antibiotic concentrations: Implications for treatment of cystic fibrosis lung infection. *Pediatric Pulmonology*, 42(11), 1008–1017. <https://doi.org/https://doi.org/10.1002/ppul.20671>
 32. Moyá, B., Beceiro, A., Cabot, G., Juan, C., Zamorano, L., Alberti, S., & Oliver, A. (2012). Pan-β-Lactam Resistance Development in *Pseudomonas aeruginosa* Clinical Strains: Molecular Mechanisms, Penicillin-Binding Protein Profiles, and Binding Aff. *Antimicrobial Agents and Chemotherapy*, 56(9), 4771 LP – 4778. <https://doi.org/10.1128/AAC.00680-12>
 33. Moya, B., Dötsch, A., Juan, C., Blázquez, J., Zamorano, L., Haussler, S., & Oliver, A. (2009). Beta-lactam resistance response triggered by inactivation of a nonessential penicillin-binding protein. *PLoS Pathogens*, 5(3), e1000353. <https://doi.org/10.1371/journal.ppat.1000353>
 34. Palmer, K. L., Aye, L. M., & Whiteley, M. (2007). Nutritional Cues Control *Pseudomonas aeruginosa* Multicellular Behavior in Cystic Fibrosis Sputum. *Journal of Bacteriology*, 189(22), 8079 LP – 8087. <https://doi.org/10.1128/JB.01138-07>
 35. Podnecky, N. L., Rhodes, K. A., & Schweizer, H. P. (2015). Efflux pump-mediated drug

- resistance in burkholderia. *Frontiers in Microbiology*, 6(MAR), 1–25.
<https://doi.org/10.3389/fmicb.2015.00305>
36. Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, 26(6), 841–842.
<https://doi.org/10.1093/bioinformatics/btq033>
 37. Rausch, T., Zichner, T., Schlattl, A., Stütz, A. M., Benes, V., & Korbel, J. O. (2012). DELLY: Structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*, 28(18). <https://doi.org/10.1093/bioinformatics/bts378>
 38. Rhodes, K. A., & Schweizer, H. P. (2016). Antibiotic resistance in Burkholderia species. *Drug Resistance Updates*, 28, 82–90. <https://doi.org/10.1016/j.drug.2016.07.003>
 39. Schick, A., & Kassen, R. (2018). Rapid diversification of *Pseudomonas aeruginosa* in cystic fibrosis lung-like conditions. *Proceedings of the National Academy of Sciences*, 115(42), 10714 LP – 10719.
<https://doi.org/10.1073/pnas.1721270115>
 40. Shafiq, I., Carroll, M. P., Nightingale, J. A., & Daniels, T. V. W. (2011). Cepacia syndrome in a cystic fibrosis patient colonised with Burkholderia multivorans. *BMJ Case Reports*. <https://doi.org/10.1136/bcr.08.2010.3296>
 41. Silva, I. N., Santos, P. M., Santos, M. R., Zlosnik, J. E. A., Speert, D. P., Buskirk, S. W., Bruger, E. L., Waters, C. M., Cooper, V. S., & Moreira, L. M. (2016). Long-Term Evolution of Burkholderia multivorans during a Chronic Cystic Fibrosis Infection Reveals Shifting Forces of Selection. *MSystems*, 1(3).
<https://doi.org/10.1128/msystems.00029-16>
 42. Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312–1313.
<https://doi.org/10.1093/bioinformatics/btu033>
 43. Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, 14(2), 178–192. <https://doi.org/10.1093/bib/bbs017>
 44. Viberg, L. T., Sarovich, D. S., Kidd, T. J., Geake, J. B., Bell, S. C., Currie, B. J., & Price, E. P. (2017). Within-Host Evolution of *Burkholderia pseudomallei* during Chronic Infection of Seven Australasian Cystic Fibrosis Patients. *MBio*, 8(2), e00356-17. <https://doi.org/10.1128/mBio.00356-17>
 45. Ye, K., Schulz, M. H., Long, Q., Apweiler, R., & Ning, Z. (2009). Pindel: A pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*, 25(21), 2865–2871.
<https://doi.org/10.1093/bioinformatics/btp394>
 46. Ye, K., Schulz, M. H., Long, Q., Apweiler, R., Ning, Z., Thorvaldsdóttir, H., Robinson, J. T., Mesirov, J. P., Stamatakis, A., Silva, I. N., Santos, P. M., Santos, M. R., Zlosnik, J. E. A., Speert, D. P., Buskirk, S. W., Bruger, E. L., Waters, C. M., Cooper, V. S., Moreira, L. M., ... Andrews, S. (2020). FastQC: a quality control tool for high throughput sequence data. *Bioinformatics*, 32(5), Article R package version 1.6.0.
<https://doi.org/10.1101/gr.222109.117>